



Case Formulation in Cognitive Behavioural Therapy for Psychosis

A thesis submitted to Newcastle University for the degree of
Doctor of Philosophy in the Faculty of Medical Sciences

December 2024

Helen M. Spencer

School of Psychology

Translational and Clinical Research Institute

Thesis abstract

This thesis explored case formulation (CF) in Cognitive Behavioural Therapy for psychosis (CBTp). CF is considered a key process in therapy. Chapter 1 introduces key definitions and outlines a range of psychological models used as the basis of a formulation. Chapter 2 provides a systematic review and narrative synthesis of the CF literature, focusing on the key question of whether formulations are reliable. Findings indicated a lack of consistency in how reliability is investigated, in part reflecting the wide range of models that may be used when formulating. To address this low rate of agreement, Chapter 3 demonstrates how a 3-stage Delphi method was used to establish international expert consensus for components of a CF in relation to auditory hallucinations, and persecutory delusions. Findings revealed that CFs should be parsimonious and focused on perpetuating (maintaining) factors to facilitate change. Chapter 4 uses methods of causal inference to examine whether CF modified treatment effects, for service users that received CBTp in a RCT. The findings provided a tentative signal that longitudinal CFs in CBTp may lead to poorer treatment effects, if explored too early on in therapy. In addition, a ‘dose-response’ was found, suggesting that length of therapy appears to be important. Using Reflexive Thematic Analysis, Chapter 5 explores the personal impact of CF from a service user perspective. Several themes were developed, including an overarching theme that considered the role of different levels of CF as a vehicle for change. Using mixed methods (systematic review, Delphi expert consensus, predictors of response to treatment, qualitative interviews with people with lived experience), the chapters in this thesis collectively highlight the importance of maintenance features in relation to the formulation of CBTp, but not in relation to outcome. Chapter 6 provides an overall discussion of the thesis findings.

Acknowledgements

Firstly, I would like to thank Dr Douglas Turkington for offering me the opportunity to undertake this part-time PhD. Also, for the many teaching opportunities afforded to me throughout my studies which helped to keep me afloat financially. Thank you for your valuable insights, Doug, and pearls of wisdom along the way. Secondly, I would like to thank Prof. Robert Dudley for more than a decade of regular high-quality supervision. Thank you for your patience, support, and encouragement Rob; the completion of this thesis would not have been possible without you. Thirdly, I would like to thank Prof. Mark Freeston for his kindness, academic advice, and guidance in overseeing this research.

In terms of external supervision, firstly I would like to thank Prof. Anthony Morrison in helping me think through some of the initial design ideas in relation to my research. Secondly, I would like to thank Dr Sarah Tully for her expertise and experience in guiding me through the qualitative reflexive thematic analysis, and to Dr Lynne Johnston, and Prof. Andrew Gumley for their methodological guidance in the earlier phases of the qualitative work. I would like to acknowledge and give thanks to the late Prof. Graham Dunn, and to Dr Martina McMenamin and Prof. Richard Emsley, for their expert statistical input regarding the ACTION trial secondary analysis.

I am grateful to the participants in the ACTION trial and the other studies presented here, for giving up their time to take part in my research. Special thanks are given to Ron (pseudonym) who co-developed and piloted the qualitative interview topic guide. I would also like to thank the Manchester Psychosis Research Unit (PRU) Service User Reference Group (SURG), who provided consultation and feedback on the qualitative study materials.

I am also grateful to many colleagues at Newcastle University including Beckie Headley, Dr Peter Gallagher, Dr Luke Gaughan, and Prof. Alison Tyson-Capper for supporting my (many) extensions/study interruptions. Thanks, are also owing to Linda Errington (Health Sciences Librarian) who taught me how to conduct systematic review searches for my literature review. Thanks also to Andrew Daniels in I.T who provided me with a newly refurbished university laptop (free of charge) which proved to be trustworthy and reliable, throughout my thesis write-up.

Special thanks are given to my PhD peers in our writing group who have motivated me and worked alongside me online in Discord every day for the past three

years. These incredible people have worked with me on their wide range of PhD's from different countries, cities, and time zones around the world: Dr Mel Escue (Florida, USA), Dr Dinesh Adepu (Mumbai, India), Dr Rachel Moss (Minnesota, USA), Dr Sophia Zhang (Texas, USA), Dr Hina Kanabar (Sheffield, UK), and especially Dr Sai (Geetha) Seri (Boston, USA) - I will miss working with you.

Heartfelt thanks to my closest friends who have spoon-fed me an abundance of kindness, compassion, and encouragement throughout the duration of my PhD; namely, Dr Lucy Stevens, Nicola McCreary, and Lynsey Harkness. Finally, thank you to my fiancé Richard, for his love and support.

Contents

Chapter 1: Introduction	1
1.1 Defining a case formulation	2
1.2 Defining (and making sense of) psychosis.....	3
1.3 A brief history of the origins of case formulation.....	5
1.4 Do we really need a case formulation? Isn't a psychiatric diagnosis sufficient?	9
1.4.1 The Power Threat Meaning Framework	13
1.4.2 The five Ps framework	14
1.5 Conceptual underpinnings of a CBT case formulation	17
1.6 Different psychological approaches, theoretical perspectives, and a multitude of models! The conundrum when developing a case formulation.....	20
1.6.1 What are the differences between a formulation and a model?	21
1.6.2 Psychodynamic formulation	21
1.6.3 Systemic formulation	23
1.6.4 Cognitive Analytic Therapy reformulation.....	24
1.7 Formulation across generic models of CBTp.....	26
1.7.1 A-B-C model.....	27
1.7.2 Maintenance model	28
1.7.3 Longitudinal model	29
1.8 Formulation across disorder-specific models of CBTp	31
1.8.1 Morrison's cognitive model of hallucinations and delusions	32
1.8.2 Freeman et al.'s cognitive model of persecutory delusions	34
1.9 Conclusion	37
1.10 Thesis aims.....	39
Chapter 2: Are case formulations reliable? A systematic literature review with a narrative synthesis.....	41
2.1 Abstract	41
2.2 Introduction	42
2.3 Aims	46
2.4 Method	46
2.4.1 Inclusion criteria.....	46

2.4.2 Exclusion criteria	47
2.4.3 Search strategy	47
2.5 Assessment of methodological quality	51
2.6 Results.....	52
2.6.1 Methodological quality findings	52
2.6.2 Data extraction	63
2.6.3 Methods used to assess CF reliability	63
2.6.4 Reliability measurement	125
2.6.5 Study location	127
2.6.6 Theoretical modality	127
2.6.7 Participant demographics	127
2.6.8 Training.....	128
2.6.9 Formulator/rater participant sample.....	129
2.6.10 Formulation data	130
2.6.11 Blinding.....	131
2.6.12 Overall key findings.....	132
2.6.13 Cognitive/cognitive behavioural formulations.....	141
2.6.14 Psychoanalytic/psychodynamic formulations.....	142
2.6.15 Integrative/combined formulations	143
2.6.16 Atheoretical formulations	144
2.7 Discussion	145
2.7.1 Methodological limitations of the studies	147
2.7.2 Clinical implications	148
2.7.3 Limitations, and recommendations for future research	150
2.8 Conclusion	151
Chapter 3: What are the essential ingredients of a CBT case formulation for auditory hallucinations and persecutory delusions in schizophrenia spectrum disorders? Using the Delphi method to establish an expert consensus.....	
3.1 Abstract	153
3.2 Introduction.....	153
3.2.1 The Delphi method.....	159
3.2.1.1 How to define an expert?	163
3.2.2 Aims	165

3.3 Methods.....	165
3.3.1 Participants.....	165
3.3.1.1 Pre-defined expert criteria.....	165
3.3.1.2 Expert identification and recruitment.....	166
3.3.1.3 Eligibility screening	167
3.3.2 Delphi procedure and analysis	167
3.3.2.1 Ethical approval	167
3.3.2.2 The 3-stage process	168
3.3.2.3 Stage 1.....	168
3.3.2.4 Stage 2.....	169
3.3.2.5 Stage 3.....	171
3.4 Results	173
3.5 Delphi secondary analysis.....	181
3.5.1 Secondary research question	181
3.5.2 Rationale	181
3.5.3 Expert groups	181
3.5.4 Secondary analysis.....	186
3.5.4.1 Perpetuating factors: What keeps the problems going?	186
3.5.4.2 Protective factors: What are the person's strengths?	187
3.5.5 Results.....	187
3.6 Discussion	188
3.6.1 Strengths and limitations.....	196
3.6.2 Future research	200
Chapter 4: Examining the treatment effects of case formulation in	
CBTp for service users in the absence of antipsychotics: A	
therapeutic process evaluation of the ACTION trial.....	202
4.1 Abstract	203
4.2 Introduction	203
4.2.2 Aims	208
4.2.3 Hypotheses	208
4.3 Method	209
4.3.1 Procedure.....	209
4.3.1.1 Ethical approval	209
4.3.1.2 ACTION trial design and participants	209

4.3.1.3 Intervention	210
4.3.1.4 The Positive and Negative Syndrome Scale (PANSS)	211
4.4 Data analysis	211
4.4.1 Statistical analysis 1: Principal stratification	211
4.4.2 Statistical analyses 2: Instrumental variable (IV) methods.....	214
4.5 Results.....	215
4.5.1 Process variables	215
4.5.2 Number of therapy sessions	218
4.6 Discussion	220
4.6.1 Strengths, limitations, and recommendations for future research..	221
Chapter 5: Case formulation – A vehicle for change? Exploring the impact of CBT formulation in early psychosis: A reflexive thematic analysis	225
5.1 Abstract	226
5.2 Introduction.....	226
5.2.1 Aims	229
5.2.2 Research questions.....	229
5.3 Methodology	230
5.3.1 Design	230
5.3.2 Procedure	230
5.3.2.1 Ethical approval	230
5.3.2.2 Participant sampling and recruitment	231
5.3.2.3 Interview topic guide	233
5.3.2.4 Interview pilot and SU consultation.....	233
5.3.2.5 Qualitative interviews	233
5.3.3 Thematic analysis.....	234
5.3.3.1 Reflexive, codebook, and coding reliability approaches to thematic analysis	234
5.3.3.2 The six-phase approach to reflexive TA	237
5.3.3.3 Ontology and epistemology	239
5.3.3.4 Reflexivity and other indicators of quality	240
5.4 Analysis.....	244
5.4.1 Contextualising the data.....	244
5.4.2 Overview of analysis.....	244

5.4.3 Overarching theme – CF - A vehicle for change?.....	249
5.4.4 Theme 1.....	249
5.4.4.1 Subtheme 1.....	251
5.4.5 Theme 2.....	253
5.4.5.1 Subtheme 2.....	255
5.4.6 Theme 3.....	257
5.4.7 Theme 4.....	263
5.5 Discussion	268
5.5.1 Strengths, limitations, and recommendations for future research..	277
Chapter 6: General discussion	281
6.1 Recap of the thesis aims and summary of key findings.....	281
6.2 Overarching themes and reflections.....	284
6.2.1 The CF versus diagnosis debate.....	284
6.2.2 Use of the four/five Ps frameworks to improve reliability and achieve consensus	285
6.2.3 Parsimony and simplicity.....	286
6.2.4 Different levels of CF: A word of caution surrounding the use of longitudinal formulations versus the use of maintenance formulations	287
6.3 Thesis strengths, limitations, and concluding remarks	288
7. References	292
8. Appendices	346

Word count: 88,296

List of Tables

Table 2.1 Assessment of methodological quality.....	53
Table 2.2 Study details: An update on Flinn et al. (2015) via database searching, citation searching, and reference list trawling (<i>n</i> = 9 studies)	65
Table 2.3 Study details: Papers identified by Easden & Kazantzis (2018), not included in Flinn et al. (2015; <i>n</i> = 8 studies)	68
Table 2.4 Study details: Papers identified in a review by Flinn et al. (2015; <i>n</i> = 15 studies)	94
Table 2.5 Methods used to assess CF reliability across all <i>n</i> = 32 studies included in the review	123
Table 2.6 The range of reliability scores (and method(s) used) across all studies included in the review	136
Table 2.7 The range of percentage agreement scores (and method(s) used) across all studies included in the review	139
Table 3.1 Participant characteristics at each stage of the study	174
Table 3.2 Components endorsed as ‘essential’ in the CF of auditory hallucinations, and persecutory delusions.....	178
Table 3.3 Components <i>not</i> endorsed in stage 3 as ‘essential’ in the CF of auditory hallucinations, and persecutory delusions	183
Table 3.4 Level(s) of expertise in stage 3, between a subgroup of experts with ‘higher esteem indicators’ versus the wider sample of experts with ‘applied indicators’	185
Table 4.1 Dichotomisation of each process variable.....	213
Table 4.2 Number of ‘compliers’ and ‘non-compliers’ in the treatment arm for each process variable.....	216
Table 4.3 Results from the principal stratification analysis for longitudinal formulation	217
Table 4.4 Descriptive statistics from the ACTION trial: Baseline and 18-month PANSS total scores, and number of sessions (organised by site, and randomisation allocation)	219
Table 5.1 Participant characteristics.....	245
Table 5.2 Clinical implications and recommendations for therapists in relation to the overarching theme, themes, and subthemes.....	273

List of Figures

Figure 1.1 An example of a five Ps formulation for early psychosis.....	16
Figure 1.2 An example of an A-B-C formulation for voice hearing.....	27
Figure 1.3 An example of an A-B-C formulation for a persecutory belief. .	28
Figure 1.4 An example of a maintenance formulation for voice hearing ...	29
Figure 1.5 An example of a longitudinal formulation in the pre-psychotic period.....	31
Figure 1.6 An example of a longitudinal formulation for auditory hallucinations, based on Morrison’s cognitive model of psychosis.....	34
Figure 1.7 An example of a vicious flower maintenance formulation, based on Freeman’s cognitive model of persecutory delusions.....	37
Figure 2.1 PRISMA flow diagram	50
Figure 3.1 Number of items included, rerated, and excluded at each stage of the study.....	177
Figure 3.2 Essential ingredients of a CF for auditory hallucinations.....	190
Figure 3.3 Essential ingredients of a CF for persecutory delusions.....	191
Figure 5.1 Thematic map of the overarching theme, themes, and subthemes.....	248

List of Appendices

Appendix A. Review of the reflexive TA study by ‘Psychology Tools’ in their August 2023 newsletter	346
Appendix B. Review of the Delphi study by ‘Psychology Tools’ in their November 2020 newsletter	347
Appendix C. PsycINFO search strategy	348
Appendix D. MEDLINE search strategy	349
Appendix E. CINAHL search strategy	350
Appendix F. Web of Science search strategy	351
Appendix G. Google Scholar search strategy	352
Appendix H. Example of an email invite to participate in stage 1 of the Delphi study	353
Appendix I. Example of an email invite to participate in stage 2 of the Delphi study	354
Appendix J. Example of an email invite to participate in stage 3 of the Delphi study	356
Appendix K. Flyer circulated at the 9 th World Congress of Behavioural and Cognitive Therapies (WCBCT), Berlin, 2019	358
Appendix L. Example of a recruitment invite circulated on Twitter	359
Appendix M. Qualtrics survey – Stage 1 of the Delphi study	360
Appendix N. Qualtrics survey – Stage 2 of the Delphi study	400
Appendix O. Letter of ethical approval from Newcastle University for the Delphi study	432
Appendix P. Components added or removed in stage 1, in the CF of auditory hallucinations, and persecutory delusions	433
Appendix Q. Components endorsed as ‘essential’ or ‘important’ in stage 2, in the CF of auditory hallucinations, and persecutory delusions..	439
Appendix R. Example of a survey sent out to an expert participant – Stage 3 of the Delphi survey	455
Appendix S. Components <i>not</i> endorsed as ‘essential’ in stages 2 and 3, in the CF of auditory hallucinations, and persecutory delusions	463
Appendix T. Components <i>not</i> endorsed as ‘essential’ or ‘important’ in	

stage 2, in the CF of auditory hallucinations, and persecutory delusions..	478
Appendix U. Letter of favourable ethical opinion from the Research Ethics Committee for the ACTION trial.....	482
Appendix V. Participant information sheet for the ACTION trial.....	485
Appendix W. Participant consent form for the ACTION trial.....	489
Appendix X. Therapist session record for the ACTION trial	491
Appendix Y. Letter of favourable ethical opinion from the Health Research Authority for the reflexive TA study	493
Appendix Z. Participant information sheet for the reflexive TA study	496
Appendix AA. Consent form for the reflexive TA study.....	499
Appendix AB. Interview topic guide	501
Appendix AC. Example of a coded transcript	508
Appendix AD. Nodes initially generated in NVivo	511
Appendix AE. Nodes developed and refined at a later phase of the analysis in NVivo.....	516
Appendix AF. Thematic map 1.....	517
Appendix AG. Thematic map 2	518
Appendix AH. Thematic map 3	519

Abbreviations

ACTION:	Assessment of Cognitive Therapy Instead of Neuroleptics
AP:	Antipsychotic medication
APA:	American Psychiatric Association
ARMS:	At-Risk Mental State
BABCP:	British Association for Behavioural and Cognitive Psychotherapies
BPD:	Borderline personality disorder
BPS:	British Psychological Society
CAARMS:	Comprehensive Assessment for At-Risk Mental State
CASP:	Critical Appraisal Skills Programme
CAT:	Cognitive Analytic Therapy
CBISs:	Cognitive-Behavioural-Interpersonal scenarios
CBISSAI:	Cognitive-Behavioural-Interpersonal Semi-Structured Assessment Interview
CBT:	Cognitive Behavioural Therapy
CBTp:	Cognitive Behavioural Therapy for psychosis
CCC-RS:	Collaborative Case Conceptualisation Rating Scale
CCDs:	Case Conceptualisation Diagrams

CCRT:	Core Conflictual Relationship Theme
CCRT-LU:	Core Conflictual Relationship Theme Leipzig/Ulm
CF:	Case formulation
CFCCM:	Case Formulation Content Coding Method
CI:	Central issue
CNTW:	Cumbria, Northumberland, Tyne and Wear NHS Foundation Trust
COMPARE:	Cognitive behaviour therapy Or Medication for Psychosis - A Randomised Evaluation
CORDS:	Comorbid, Often Re-hospitalized, Dysfunctional, and Suicidal
CREDES:	Conducting and REporting DELphi Studies
CRS:	Conceptualisation Rating Scale
CTS-R:	Cognitive Therapy Scale-Revised
DA:	Discourse Analysis
DCP:	Division of Clinical Psychology
DSM-IV:	Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition
DSM-5:	Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
DSM-5-TR:	Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision

EABCT:	European Association for Behavioural and Cognitive Therapies
EDIE-2:	Early Detection and Intervention Evaluation for people at high-risk of psychosis-2
EIP:	Early intervention in psychosis
FATR:	Formulation Assessment Tool and Rating
FEP:	First episode of psychosis
GAD:	Generalised anxiety disorder
GT:	Grounded Theory
HiTOP:	Hierarchical Taxonomy of Psychopathology
HRA:	Health Research Authority
HTA:	Health Technology Assessment
ICC:	Intraclass correlation coefficient
ICD-10:	International Classification of Diseases, Tenth Revision
ICD-11:	International Classification of Diseases, Eleventh Revision
ICF:	Ideographic Conflict Formulation
ICs:	Idiosyncratic cognitive schemas
ITT:	Intention to treat
IV	Instrumental variable

IoP:	Institute of Psychiatry
IPA:	Interpretative Phenomenological Analysis
IRT:	Interpersonal Reconstructive Therapy
JTC:	Jumping to conclusions
LoC:	Locus of Control
MAR:	Missing at random
MASQ:	Mood and Anxiety Symptoms Questionnaire
MDD:	Major depressive disorder
NAS2:	Second National Audit of Schizophrenia
NHS:	National Health Service
NICE:	National Institute for Health and Care Excellence
NOS:	Not Otherwise Specified
NPD:	Narcissistic personality disorder
NRES:	National Research Ethics Service
NTW:	Northumberland, Tyne and Wear NHS Foundation Trust
OCD:	Obsessive compulsive disorder
PANSS:	Positive and Negative Syndrome Scale

PD:	Personality disorder
PFM:	Plan Formulation Method
PIS:	Participant information sheets
PRISMA:	Preferred Reporting Items for Systematic reviews and Meta-Analyses
PRU:	Psychosis Research Unit
PTMF:	Power Threat Meaning Framework
QCTCF:	Quality of Cognitive Therapy Case Formulation
RAs:	Research Assistants
RAP:	Relationship Anecdotes Paradigm
RCT:	Randomised Controlled Trial
R&D:	Research and Development
RE:	Relationship episode
REC:	Research Ethics Committee
RO:	Responses of others
RQCFO:	Rating the Quality of Case Formulation for Obsessive-Compulsive Disorder
RRs:	Reciprocal Roles
RRM:	Role-Relationship Model

RRMC:	Role-Relationship Model Configuration
RRPs:	Reciprocal Role Procedures
RS:	Responses of self
SASB:	Structural Analysis of Social Behaviour
SDI:	Selective Dissemination of Information
SDR:	Sequential Diagrammatic Reformulation
SoAS:	Sense of Agency Scale
SU:	Service user
SURG:	Service User Reference Group
TA:	Thematic Analysis
TAU:	Treatment as usual
TEWV:	Tees, Esk and Wear Valleys NHS Foundation Trust
TLP:	Time Limited Psychotherapy
2SLS:	Two-stage least squares
UCSF:	University of California, San Francisco
W:	Wishes
WCBCT:	World Congress of Behavioural and Cognitive Therapies

WHO: World Health Organisation

Dissemination of the work

Key elements from Chapter 1 have been published as a book chapter by Cambridge University Press. In addition to this, the three empirical studies have been published in peer reviewed journals. Chapters 3 and 4 were both published by Schizophrenia Research, and Chapter 5 was published by Psychology and Psychotherapy: Theory, Research and Practice (PAPTRAP). These publications were expanded in more detail for the purposes of ensuring a coherent whole (i.e. this PhD is being examined ‘by thesis’ rather than ‘by publication’). The publications in relation to Chapters 3 and 5 have also received independent reviews and are featured online by ‘Psychology Tools’ (see Appendices A and B).

Funding

This PhD was part-funded by Dr Douglas Turkington’s research budget. I self-funded the remaining fees. I am also grateful to have been supported by an award for £2,070.00 from Funds for Women Graduates (FfWG) in the UK (grant number: 17136). This is the trading name of the British Federation of Women Graduates (BFWG) Charitable Foundation.

About the author

I completed my B.A. undergraduate degree in Social Sciences with Honours in Psychology from Glasgow Caledonian University in 2005. I started work in the National Health Service (NHS) in 2005 as a support worker in an Early Intervention in Psychosis (EIP) service. I was then promoted to the role of Higher Assistant Psychologist in the same service in 2006, where I worked with Prof. Robert Dudley.

In 2009, I accepted a position with Dr Douglas Turkington as a Research Assistant (RA). In this role I worked on several psychosis studies, including the first pragmatic open study, and then (a few years later) a randomised controlled trial (RCT) known as ACTION (Assessment of Cognitive Therapy Instead of Neuroleptics), that offered CBT to people experiencing psychosis, that had chosen not to take antipsychotic medication. It was in this full-time RA role that I registered for my part-time PhD in Psychology, in December of 2011.

In 2017, I was employed part-time (0.6wte) as a Clinical Studies Officer (CSO) in the Dementias and Neurodegeneration (DeNDRoN) specialty. Alongside this role, I also worked part-time post (0.2wte) as a Higher Assistant Psychologist on a Primary Care weight management programme called ‘Food for Thought’ to help fund my PhD.

In 2019, I accepted a part-time (0.6wte) position as a Clinical Studies Officer in the Mental Health specialty. In 2020, I began work as a RA on the Evaluation of a Parenting Intervention for Children with Challenging Behaviours and Intellectual Disabilities (EPICC-ID) trial.

When the EPICC-ID trial finished in 2022, I was able to take a sabbatical from the NHS to finish the write-up of my PhD thesis. I completed my PhD whilst on this sabbatical.

Chapter 1. An introduction to case formulation in psychosis

“To know thyself is the beginning of wisdom”.

– Inscribed on the frontispiece of the Temple of Delphi (n.d.).

Elements of section 1.2, and subsections 1.7.1 to 1.7.3 have been published as a book chapter for service users and carers, and the reference for this publication is:

Spencer, H. M. (2019). Making sense of psychosis. In D. Turkington & H. M. Spencer (Eds.), *Back to life, back to normality: CBT informed recovery for families with relatives with schizophrenia and other psychoses* (volume 2, pp. 27-32). Cambridge University Press. <https://doi.org/10.1017/9781316410523.007>

This thesis explores case formulation in relation to the understanding and treatment of psychosis. Case formulation is considered a key process in cognitive behavioural therapy, helping to develop a shared understanding of what has led to and what is maintaining the problems a person faces; it is used to guide treatment. Given its vital role, case formulation seems particularly well suited to psychotic symptoms which were long considered ‘difficult to understand’ and ‘hard to treat’.

This chapter will introduce key issues in the definition of formulation and in its application to psychosis.

1.1 Defining a case formulation

The term ‘case formulation’ (CF) (otherwise referred to as a ‘case conceptualisation’, or ‘psychological formulation’) was first defined in the behavioural psychology literature in 1979 as an explanatory hypothesis that: *“(1) relates all of the patient’s complaints to one another, (2) explains why the individual developed these difficulties, and (3) provides predictions regarding the patient’s behavior given any stimulus conditions”* (Meyer & Turkat, 1979, p. 261).

Despite absence of clear scientific data to determine how best to define a CF (Turkat, 2014) over time, definitions of CF have evolved. In 2010, the British Psychological Society’s (BPS) Division of Clinical Psychology (DCP) broadly defined CF as:

the summation and integration of the knowledge that is acquired by the assessment process that may involve psychological, biological and systemic factors and procedures. The formulation will draw on psychological theory and research to provide a framework for describing a client’s problems or needs, how it developed and is being maintained (...). This provides the foundation from which actions may derive (DCP, 2010, pp. 5-6).

In Cognitive Behavioural Therapy (CBT), CF has been described as the ‘cornerstone’ of CBT (Beck, 2020), a blend of both ‘science and art’ (Eells, 2022), and as the ‘lynchpin’ of therapy (Butler, 1998). In CBT, CF has been defined as: *“a process whereby [the] therapist and client work collaboratively first to describe and then to explain the issues a client presents in therapy. Its primary function is to guide therapy in order to relieve client distress and build client resilience”* (Kuyken et al., 2009, p. 3). In relation to CBT for psychosis (CBTp), the author of this thesis (H.M.S) has stated that: *“a key function of formulation is to help create understanding of these experiences, and direct treatment, with CFs attempting to make sense of the meaning and mechanisms of psychosis”* (Spencer et al., 2023, p. 329).

However, for the purpose of this thesis, the following definition By Tracy Eells is preferred: *“a hypothesis about the causes, precipitants, and maintaining influences of a person’s psychological, interpersonal, and behavioural problems...[that] serves as a blueprint to guide treatment and as a marker for change”* (Eells, 2022, p. 2). This

definition succinctly captures the levels of formulation, the range of contributory factors, and its utility in terms of facilitating treatment and change.

1.2 Defining (and making sense of) psychosis

German psychiatrist Emil Kraepelin was the first to devise the diagnostic classification for ‘dementia praecox’ (Kraepelin, 1919). This was viewed as an ‘organic brain disease’ – a chronic, progressive illness, leading to deterioration (Kraepelin, 1919). Despite changes to its boundaries and definitions (and stigmatised attitudes towards the label; Pyle & Morrison, 2014), ‘schizophrenia’ (as it is nowadays referred to; Bleuler, 1911/1950), has survived for over a century (Tandon et al., 2013). However, Bleuler was the first to describe ‘a group of schizophrenias’ (Bleuler, 1911/1950), and since the turn of the 21st century, different ‘subgroups’ of schizophrenia have been proposed (i.e. sensitivity, traumatic, drug-related, and anxiety psychosis), with new ways of understanding aetiology, prognosis, and treatment (Kingdon et al., 2008; Kingdon & Turkington, 2008).

A diagnosis of schizophrenia falls under the umbrella term ‘psychosis’ and is used to infer that the person has ‘lost touch with reality’ (Morrison et al., 2008). Positive symptoms of psychosis are primarily characterised by auditory hallucinations, and persecutory delusions (Wright et al., 2014). Typically, these core symptoms are the focus of treatment (Spencer et al., 2020). However, hallucinations may also be experienced via other sensory modalities (e.g. visual, olfactory, tactile). Delusions may also include grandiose beliefs, delusions of thought insertion, thought broadcast, delusions of reference etc. (American Psychiatric Association [APA], 2022).

The Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5; APA, 2013), defined hallucinations as: “*A perception-like experience with the clarity and impact of a true perception but without the external stimulation of the relevant sensory organ*” (APA, 2013, p. 822). The DSM-5 also defined delusions as:

A false belief based on incorrect inference about external reality that is firmly held despite what almost everyone else believes and despite what constitutes incontrovertible and obvious proof or evidence to the contrary. The belief is not ordinarily accepted by other members of the person’s culture... (APA, 2013, p. 819).

In addition, more than half of those diagnosed with chronic schizophrenia also experience at least one negative symptom (Bobes et al., 2010). Primary negative symptoms include volitional (motivational) impairment (e.g. avolition, anhedonia, social withdrawal), and emotional difficulties (e.g. alogia, and flattened affect; Mosolov & Yaltonskaya, 2022). Individuals may also experience secondary negative symptoms owing to the distressing nature of positive symptoms (e.g. comorbid depression; Mosolov & Yaltonskaya, 2022). As such, Wright et al. (2014) explained that: *“psychotic disorders are associated with significant morbidity and mortality and extensive disability in occupational, social, and day-to-day functioning as well as overall quality of life”* (p. 4).

Indeed, the disability associated with psychosis often requires family members to be involved in the person’s care which can have a negative impact on their own quality of life, and mental health (Wright et al., 2014). It can be very confusing and baffling for carers, family, and friends to understand what is happening to their loved one when they start to experience psychotic symptoms for the very first time (Spencer, 2019). It can also, of course, be a very strange and frightening experience for the individual themselves, who may ask family and friends *‘why is this happening to me?’* and *‘why now?’*. In schizophrenia, SUs and carers very often become paralysed by this question because they just can’t make sense of what is happening (Spencer, 2019). This is where formulation comes into play (see section 1.3 onwards, below).

The traditional view of psychosis was that its symptoms were ‘un-understandable’ (Jaspers, 1963/1997). Furthermore, it was believed that people with schizophrenia would not benefit from talking therapies (Freud, 1957). Psychoanalysis was seen as contraindicated, with the assumption that psychotic transference would lead to a risk of ill-health in the therapist (Freud, 1957). Owing to this, there was (and to a large extent, still is) a heavy reliance on pharmacological interventions for the treatment of psychosis (Wright et al., 2014), despite the multitude of side effects (e.g. see Alvarez-Jimenez et al., 2008; Correll & De Hert, 2013; Ho et al., 2011; Ray et al., 2009), and evidence to suggest the effectiveness of antipsychotics has been overestimated (Morrison, Hutton, Shiers et al., 2012).

The medical/biological model is the major model for understanding psychosis. However, recovery via the use of psychological approaches has led to an increase in the investment of psychologically informed interventions over recent years (NHS England,

2019, 2023). In Britain in the early 90s, CBT for schizophrenia first developed (Kingdon & Turkington, 1994). This brought about a change in the way that psychosis was understood. The focus of the work was to help SUs understand the *meaning* of psychosis - directly related to the individual's upbringing and life events. The aim was for symptoms such as delusions to become more understandable, in relation to a person's life history (Roberts, 1991). Similarly, the content of voices was understood to often mirror previous trauma(s) (Kingdon & Turkington, 1999). Normalising symptoms via a basic formulation that linked psychotic symptoms to common and understandable factors such as sleep deprivation, sensory deprivation, stress, and substance use (as opposed to being inexplicable signs of madness), helped to reduce their affective correlates (e.g. shame, anxiety, anger and/or sadness; Morrison, 1998). Indeed, the use of formulation brought about an assumption "*...that this process will render even the most unusual or disturbing behaviour and experiences understandable*" (DCP, 2011, p. 6). As Butler explained "*...at some level it all makes sense*" (Butler, 1998, p. 2).

Contemporary understanding of psychotic experiences, regard them as existing on a continuum, along with 'normal' experiences (Linscott & van Os, 2010; Wright et al., 2014), with approximately one in ten people reporting experiences such as hearing voices (Johns et al., 2014). Epidemiological studies suggest that psychotic symptoms are experienced by perhaps ten times as many people than those that actively seek help via mental health services (Tien, 1991; van Os et al., 2000). A symptom-based approach to treating psychosis, also helps to identify common mechanisms that underly different symptoms (Morrison et al., 2004). Furthermore, therapists need to be flexible by combining/adapting several different cognitive models, to capture the array of difficulties that may be experienced by an individual (Morrison et al., 2004; see also sections 1.7 and 1.8. below). It is these limitations that necessitate a formulation-based approach to treatment (Morrison et al., 2004).

1.3 A brief history of the origins of case formulation

There are variations in the literature regarding the origins of CF (Bruch, 2015). In 1931, Adolf Meyer, a prominent Swiss-born American Psychiatrist, delivered a series of lectures which argued that mental health difficulties could be understood as a 'psychobiological reaction' to life stressors (Christiansen, 2007; Makkena, 2024; Owen,

2023). This way of thinking was revolutionary at the time (Makkena, 2024). A. Meyer's psychobiological model (Meyer, 1931) challenged traditional biomedical views of psychiatry, which historically had focused solely on brain pathology (Makkena, 2024). Moving towards a pluralistic model of causation laid the foundations for a new approach to psychiatry (Owen, 2023). A. Meyer emphasised the need to collect detailed case histories via a 'life chart' (Meyer, 1951) – a diagram that showed inter-connections between a person's life history and mental/physical ill-health (Christiansen, 2007; Makkena, 2024; Owen, 2023). A. Meyer argued that 'formulating' the person in the context of their environment should take precedence over diagnostic labels (Makkena, 2024).

The eminent psychiatrist George Engel later developed A. Meyer's "*psychobiological integration*" (Meyer, 1931, p. 49), to what is now widely known as the biopsychosocial model (Engel, 1977; Owen, 2023). Nowadays, a version of the biopsychosocial model may also be understood in the context of the four (Bolton, 2014), or five Ps framework (Dudley & Kuyken, 2014; see section 1.4.2 below). Treatment then targets any of the four or five Ps, with protective factors seen as components of the formulation that should be strengthened or enhanced (Owen, 2023).

Despite the history outlined above, it is asserted that CF officially evolved out of the Maudsley Hospital in London in the early 1950s hence CF is credited to four 'influential clinicians' - Hans Eysenck, Victor Meyer, Monte Shapiro, and Ira Turkat - for their innovative work in pioneering the development of CF in psychotherapy (DCP, 2011). This coincided with the establishment of clinical psychology in England (also led by Eysenck; Bruch, 2015).

The notion of formulating individual cases, was consistent with the 'scientist-practitioner' method outlined by Eysenck and Shapiro (Turkat, 2014, 2015). This was part of the initial focus of clinical psychology - to understand (and treat) psychopathology at an idiosyncratic level, and to ground clinical actions in the knowledge produced by scientific research (Turkat, 2014, 2015). Indeed, there began a growing discontent for psychiatric diagnoses (Bruch, 2015; see section 1.3 below) and the ubiquitous approach of 'symptom-technique matching' – selecting standard methods which are 'prescribed' to match diagnostic categories and evaluated in randomised controlled trials (RCTs; Bruch, 2015).

CF represented a paradigm shift that moved away from taxonomic classification to an alternative means of conceptualising mental health difficulties (i.e. based on empirically supported theories; Tarrier, 2006a). Moreover, CF played a crucial role in establishing the status and independence of clinical psychology as a profession, which, up until this point, had been overshadowed by psychiatry.

Eysenck strongly encouraged experimental investigation of single cases, based on learning principles (Bruch, 2015). However, as explained above (in section 1.1), the definition of the term ‘case formulation’ first appeared in 1979 provided by V. Meyer and Turkat (Meyer & Turkat, 1979; Turkat, 2014). Turkat (a clinical psychologist and academic from the US who did his clinical internship with V. Meyer in the late 70s) embraced the model, refined it, and made it more accessible (Bruch, 2015). However, Turkat credits V. Meyer as the founding father of CF (Turkat, 2015) given that V. Meyer was the first to demonstrate a formulation-based approach successfully - devising a viable approach for the treatment of compulsive motor rituals (e.g. handwashing; Meyer, 1966; Turkat, 2014).

Indeed, V. Meyer (who had been one of Eysenck’s students) was well known for his teaching around the world, in which he carried out live demonstrations of the CF process with his patients, in front of audiences of mental health professionals (Turkat, 2012). Unlike psychoanalytic psychotherapy (in which the therapist would never, or minimally, share their formulation in relation to the patient’s psychopathology), V. Meyer openly revealed his preliminary formulation directly to the patient. He then provided a ‘simple’ outline of potential treatment procedures, and the rationale underlying them. V. Meyer then adapted his formulation accordingly, whilst explaining to the person that their treatment response(s) may lead to further treatment modifications (Bruch, 2015; Meyer, 1975). This may have required the production of novel interventions because “*in their day, they did not have the luxury of evidence-based treatments; they created them*” (Turkat, 2014, p. 182).

In 1969, the term ‘formulation’ first emerged in the BPS’s clinical psychology regulations (BPS, 1969). Thirty years later, it was seen as *the* ‘first principle’ in CBT (Beck, 1996), and today it continues to be recognised as a core skill/competency within the profession (Beck, 2020; DCP, 2010, 2011; Eells, 2022), particularly in the treatment of complex cases such as psychosis (Bucci & Tarrier, 2016). The DCP’s (2011) publication of ‘*Good Practice Guidance on the use of psychological formulation*’

represented a critical juncture within the profession of clinical psychology (Turkat, 2014). Indeed, the DCP continues to assert that: “*this activity [is] unique to clinical psychologists*” (DCP, 2010, p. 6). Similarly, Bieling and Kuyken have stated: “*...to the scientist-practitioner cognitive therapist, individualised case formulation is the heart of evidence-based practice. It occupies a fundamental place in clinical psychology, like the role of diagnosis in psychiatry*” (Bieling & Kuyken, 2003, p. 53).

Despite CF being strongly aligned with the profession of clinical psychology, CF also features in the curriculum for psychiatrists’ training in the UK (Royal College of Psychiatrists, 2022), and today, all mental health professionals are familiar with CF and its use in everyday clinical practice (Skills for Health, Health Education England, and Skill for Care, 2016; Turkat, 2014). Nevertheless, it is argued that psychiatrists may formulate in a distinctly different way - as an extension of the biomedical model, and as an addition to, not a replacement for, a psychiatric diagnosis (DCP, 2011; Johnstone, 2014a, 2018). For instance, a psychiatrist’s formulation might be summarised as: “*schizophrenia/psychosis triggered by familial aggregation, and environmental risk factors*”. Whereas, a psychological formulation for the same case might be summarised as: “*hearing critical voices as a response to (and internalisation of) the criticism and emotional abuse you experienced*”.

Despite the emphasis placed on CF, the process has remained under-researched (Kuyken et al., 2009). This may be because historically, researchers have been preoccupied with the development of treatment protocols/manuals, and the theoretical content and psychological principles on which CF is based; as well as ‘symptom-technique matching’ (Bruch, 2015). This type of research was (and still is) considered more scientific and powerful than single-case methodology, and so researchers have become less interested in individually tailored treatment programmes more suited to CF research (Bruch, 2015). As Beck once explained: “*as clinicians and researchers, we have [historically] been too focused on therapy techniques and outcome, at the expense of focusing on key processes such as formulation*” (A. T. Beck, personal communication, May 5, 2014). Nevertheless, a recent literature search found more than 2,700 articles published on CF since 1980, and more than 30 new books published on the subject since 2006. This indicates a surge of interest in the field over recent decades (Eells, 2022).

1.4 Do we really need a case formulation? Isn't a psychiatric diagnosis sufficient?

The *International Classification of Diseases, Eleventh Revision* (ICD-11; World Health Organisation [WHO], 2019), and the DSM-5 (APA, 2013) utilise principles of nosology (i.e. the science and classification of medical diseases; Jablensky, 2016; Surís et al., 2016) to organise and structure mental health 'disorders' (Guze, 1992). Owing to this, psychiatrists *primarily* 'diagnose' and 'treat' SUs via the use of psychopharmacological interventions. This led to a psychiatrist from the Institute of Psychiatry (IoP) in London recently stating that: "*formulation in psychiatry needs resuscitating (...) In recent times psychiatry has moved away from it, let it fade or has delegated it to psychotherapy*" (Owen, 2023, p. 1700).

Nevertheless, mental health diagnoses are useful clinically, especially when it comes to the organisation and funding of clinical services, awarding government benefits in the UK, and medical insurance payments in the US (Johnstone, 2014a). Diagnoses also have research utility in terms of sorting patients into homogenous samples, and for epidemiological purposes (Carey & Pilgrim, 2010; Guze, 1992; Kendell & Jablensky, 2003; Macneil et al., 2012). However, the Hierarchical Taxonomy of Psychopathology (HiTOP; Kotov et al., 2017) have recently proposed psychological symptom 'dimensions' as a more reliable and valid alternative to diagnostic categorisation, for the purposes of clinical research and practice (Conway et al., 2021; Markon et al., 2011).

Historically, it was reported that the ICD and DSM had high levels of reliability (i.e. the likelihood that different psychiatrists would arrive at the same diagnosis; Cooper et al., 1972; Grove et al., 1981; Kendell et al., 1968; Robins et al., 1981), and fair levels of validity (i.e. whether each diagnosis is a cohesive, discrete entity; Jablensky, 2016; Tandon et al., 2009), supported by a variety of antecedent (e.g. familial aggregation, environmental risk factors), and predictive (e.g. diagnostic stability, course of illness, treatment response) validators (Bromet et al., 2011; Korver-Nieberg et al., 2011). However, concurrent validation (e.g. biomarkers) has been less robust (Cortese et al., 2023), with researchers only recently making landmark discoveries in their ongoing quest for *the* 'schizophrenic gene(s)' (see Singh et al., 2022; Trubetskoy et al., 2022; and for a discussion of the findings, see Iyegbe & O'Reilly, 2022).

As the number of psychiatric diagnoses has expanded, the reliability and validity of newer editions of these classification systems has dwindled (Black, 2005; Chmielewski et al., 2015), and nowadays their reliability and validity are vehemently contested (Frances & Nardo, 2013; Insel, 2013; Johnstone, 2014a, 2018; Johnstone et al., 2018; Jones, 2012; Kinderman & Allsopp, 2018; Kraemer et al., 2012; McCarthy, 2013). Indeed, the number of psychiatric diagnoses has increased considerably since the classification systems were first devised (Surís et al., 2016). For example, the first edition of the DSM in 1952 produced 106 diagnoses (APA, 1952; Surís et al., 2016), which almost tripled to 298 diagnoses with the publication of the DSM-5 in 2013 (APA, 2013; Surís et al., 2016). Similarly, the ICD-11 introduced new diagnostic categories such as ‘gaming disorder’, ‘prolonged grief disorder’, and ‘premenstrual dysphoric disorder’ (Gaebel et al., 2020; WHO, 2019). Consequently, the classification systems have been criticised for over-medicalising and over-pathologising vast proportions of the general population (Aarseth et al., 2017; Sutton, 2022). Indeed, some psychiatric ‘symptoms’ overlap with experiences considered to be within the ‘normal’ range (Frances & Widiger, 2012; Sutton, 2022). As such, one of the main criticisms of the DSM and ICD is that their diagnostic categories are imprecisely demarcated (Aveline, 1999; Frances, 2013a; Kendell & Jablensky, 2003; Proglor, 2009; Tandon et al., 2013).

Another important criterion to define a valid diagnosis is aetiological specificity (Aveline, 1999; Carey & Pilgrim, 2010). However, aetiological absence in relation to functional psychiatric disorders (i.e. disorders without evidence of organic brain findings), cast doubt on the validity of diagnoses such as psychosis (Carey & Pilgrim, 2010). Moreover, it has led to questions concerning whether psychiatry is a legitimate branch of medicine (Boyle & Johnstone, 2014; Johnstone, 2018; Seery et al., 2021).

ICD and DSM classification systems have received many other criticisms with critics becoming increasingly vocal over recent decades (e.g. see Bentall, 2009; Johnstone 2014a, 2018, 2022). The origins of these criticisms stem not only from the profession of clinical psychology, but from within psychiatry itself, and even from those that have played a part in devising the classification systems themselves (Frances, 2013a). For example, Allen Frances, chairman of the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) Task Force, stated that: *“The new edition of the DSM “bible” is so flawed that the US National Institute of Mental Health*

is right to abandon it...There is no reason to believe that DSM-5 is safe or scientifically sound” (Frances, 2013b, para. 1 & 3).

Johnstone (2018) has argued that the most damaging impact of a psychiatric diagnosis is: *“loss of meaning...[as] stories of trauma, abuse, discrimination, and deprivation are sealed off behind a label”* (p. 31). Solely focusing on pharmacological interventions that address a diagnosis, ignore the underlying psychological mechanisms that are likely to be perpetuating the problem (Bruch, 2015; Dudley et al., 2011). Moreover, some presentations (whilst meeting ‘general’ diagnostic guidelines to be classified as a ‘disorder’), are not considered part of the usual presentation for any one psychiatric diagnosis. As such, the individual may be classified as having a diagnosis ‘Not Otherwise Specified’ (NOS; Dudley et al., 2011). Hence, CF aims to fill the conceptual void that is often left behind by a diagnostic explanation (Eells, 2022).

A mental health diagnosis is also said to convey messages of hopelessness, blame, shame, loss of self-agency, and stigma (Brooke, 2004; Colizzi et al., 2020; Bentall, 2009; Boyle, 2002; Johnstone, 2014a, 2021; Kirk & Kutchins, 1997; Matsunaga & Kitamura 2016; Mittal et al., 2014; Seery et al., 2021); with schizophrenia once described as: *“arguably the worst disease affecting mankind”* (Editorial, 1988, p. 95). Moreover, only 13% of those living with schizophrenia in the UK are said to endorse a biological explanation of illness (McCabe & Priebe, 2004). Those experiencing a first episode of psychosis (FEP) prefer to endorse causal explanations such as drug usage, traumatic experiences in childhood, personal sensitivity, and developmental vulnerabilities (Dudley et al., 2009); which may all feature within an idiosyncratic CF).

Indeed, formulation may be seen as an antidote to diagnosis (and its potentially damaging effects), by restoring self-agency, offering normalising explanations (see Spencer et al., 2023, and Chapter 5), and offering validation, meaning, and hope (Johnstone, 2018). The overarching message conveyed by a CF for trauma psychosis is: *“you are having a normal response to an abnormal situation. Anyone else who had been through the same experiences might well end up feeling the same. You too can recover”* (Johnstone, 2018, pp. 33-34).

A contentious and long-standing debate is whether CFs should be used as an adjunct, or as an alternative, to psychiatric diagnoses (Aftab, 2020; Carey & Pilgrim,

2010; Johnstone, 2014a, 2018; Macneil et al., 2012). Some have argued that a diagnosis may compliment and inform a CF; seeing the two as synergistic (Aftab, 2020; Aveline, 1999; Eells, 2022; Mellso & Benzato, 2006; Persons, 2005; Scott & Sembi, 2006; Tarrier & Calam, 2002; Turkat & Maisto, 1983). As Craddock and Mynors-Wallis (2014) stated: “*diagnosis should be part of a formulation that brings together aetiology, severity and functioning...*” (p. 93). The Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision (DSM-5-TR) also stated that: “*The primary purpose of DSM-5 is to assist trained clinicians in the diagnosis of mental disorders as part of a case formulation...*” (APA, 2022, p. 21). Nevertheless, as previously discussed in section 1.2, a psychiatric formulation may look quite different to a psychological formulation, with diagnosis drawing attention to the similarities and differences between diagnostic case presentations, whereas a CF “*aims to understand and guide intervention [within an individual presentation] at the level of the case*” (Dudley et al., 2011, p. 215). Moreover (as referred to in section 1.2 above), for one person, a diagnosis of ‘schizophrenia’ may present itself as a wide range of (positive) hallucinatory experiences, whereas for another person it may present as a wide range of (negative) symptoms such as avolition, anhedonia, social withdrawal, and alogia (Kingdon & Turkington, 2008; Mosolov & Yaltonskaya, 2022). Nevertheless, these two people will both be offered antipsychotic medication from a psychiatrist. Moreover, recent research suggests that in CBTp two people experiencing psychosis are likely to identify very different problem lists, with 60% of service user-identified priority problems often reflecting goals that are non-psychosis-related (e.g. wanting to feel happier, to have greater self-confidence; Freeman et al., 2019; Morrison et al., 2023). Consequently, CBTp CFs (and the interventions guided by these CFs) are likely to reflect SU priorities for treatment (Morrison et al., 2023). This helps to explain why early behaviour therapists promoted the use of CF as an *alternative* to diagnosis (Bruch, 2015).

Indeed, others have argued that a psychiatric diagnosis and a psychological CF do not marry up (Aftab, 2020; Johnstone, 2014a, 2018; Johnstone et al., 2018). They are seen as reflecting fundamental conceptual differences – competing paradigms that are incompatible with one another (i.e. medical versus psychological models; Aftab, 2020; Johnstone, 2014a, 2018; Johnstone et al., 2018). As Johnstone explained, it is the difference between the message: “*you are experiencing an understandable reaction to your life circumstances*” and “*your problems are the symptoms of a medical illness*”

(Aftab, 2020, para. 7). The *Good Practice Guidelines* (DCP, 2011) also stated that best practice formulations are: “*not premised on functional psychiatric diagnoses such as schizophrenia or personality disorder. Rather, the experiences that may have led to a psychiatric diagnosis...are themselves formulated*” (DCP, 2011, p. 12). Indeed, this view has also been shared by the Critical Psychiatry Network (2013), whereby 200 UK psychiatrists issued the following statement in relation to the DSM-5: “*The view of the Critical Psychiatry Network is that a DSM diagnosis is incapable of capturing the full range of experiences of distress in the way that a more comprehensive formulation can*” (para. 4).

As such, it has been argued that psychological formulation may be viewed as an alternative to psychiatric diagnosis (Aftab, 2020; DCP, 2013; Johnstone, 2018; Pilgrim, 2000) with an ongoing campaign to ‘drop the disorder’ (Kinderman et al., 2013; Watson, 2019). This campaign argues that with a good psychological formulation, a psychiatric diagnosis is unnecessary (e.g. Aftab, 2020; Bentall, 2009; Johnstone 2014a, 2018); and we have seen an example of this in the recent shift towards trauma-informed services (Sweeney et al., 2018).

1.4.1 The Power Threat Meaning Framework

The Power Threat Meaning Framework (PTMF) has been proposed as an alternative, non-medical paradigm of conceptualising mental distress and emotional suffering (Boyle & Johnstone, 2020; Johnstone, 2021; Johnstone et al., 2018). It makes the shift from asking ‘*what is wrong with you?*’ to ‘*what has happened to you?*’ (Johnstone, 2021). The PTMF is said to go a step *beyond* psychological formulation, helping people create hopeful stories/narratives about their lives and difficulties, instead of perceiving themselves as ‘mentally ill’ (Aftab, 2020). Indeed, the framework has been found to reduce stigmatising attitudes of lay people in relation to a description of psychosis, when compared to a description of the DSM-5 schizophrenia diagnosis. Therefore, it is seen as supporting the potential benefits of (non-diagnostic) formulation-based approaches in society (Seery et al., 2021).

1.4.2 *The five Ps framework*

Within mental health services in the National Health Service (NHS) in the UK, the five Ps framework (Dudley & Kuyken, 2014) is also increasingly used as an adjunct (or alternative) to psychiatric diagnosis. Five Ps facilitates the construction of CFs by eliciting and organising information in terms of: 1. Presenting issues (the key problems or difficulties the person is experiencing at that particular moment in time), 2. Predisposing factors (factors that make the individual vulnerable to the problem, and therefore at greater risk for developing psychosis), 3. Precipitating factors (events that happened in the person's life that preceded/triggered the onset of psychosis), 4. Perpetuating factors (the things that keep a problem going to maintain the psychosis), and 5. Protective factors (skills, strengths, resources, supports, or coping strategies, that have helped the person deal effectively with stressful life events, to help mitigate the impact of psychosis; Dudley & Kuyken, 2014). Indeed, the value of the five Ps framework, is that it incorporates both maintenance and longitudinal factors, as well as highlighting areas of strength and resilience (Dudley & Kuyken, 2014).

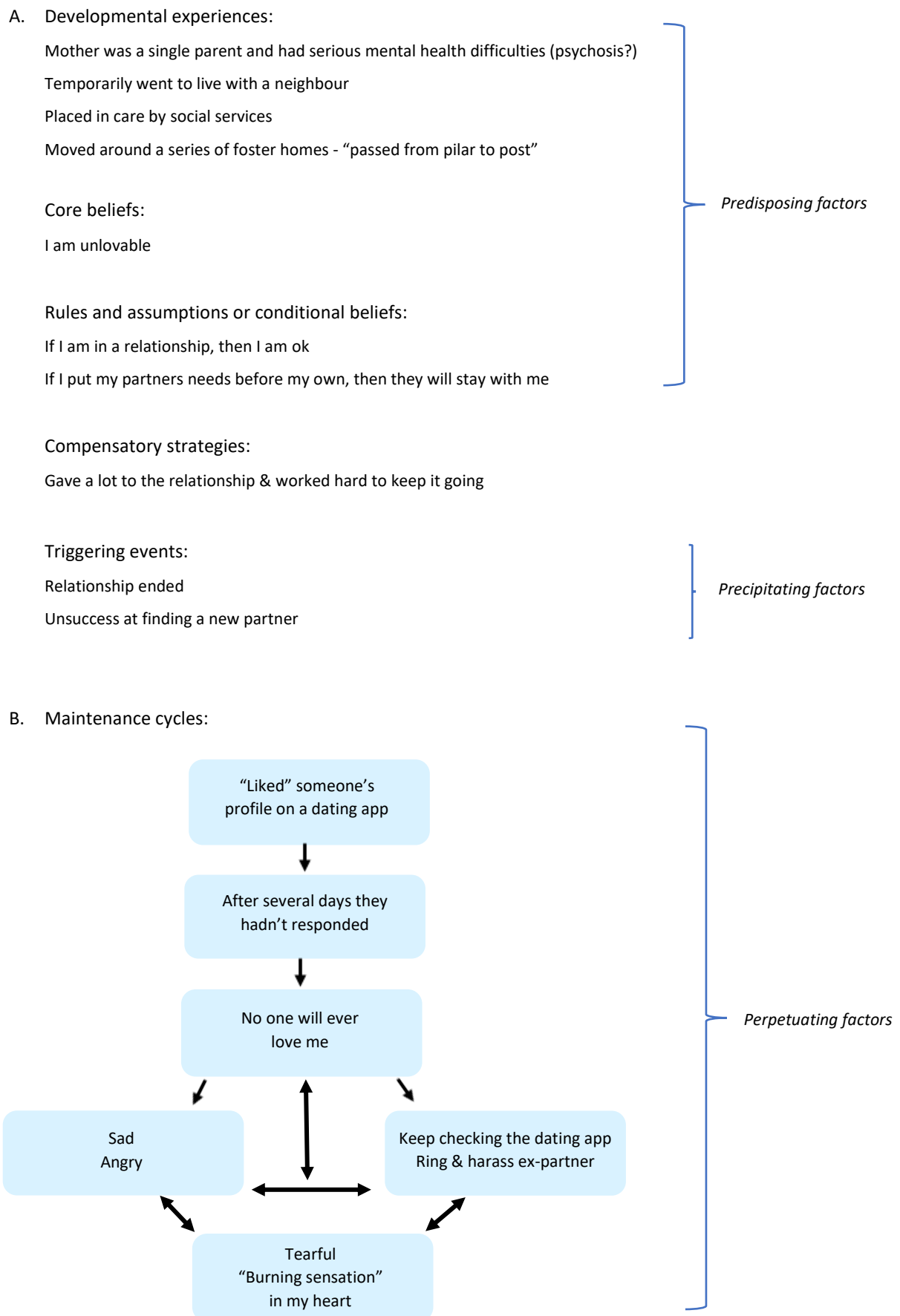
Furthermore, as mentioned above, the five Ps framework is based on the biopsychosocial model, and so it is largely generic and atheoretical (Owen, 2023). For example, predisposing factors associated with an atheoretical five Ps framework may include: exposure to maltreatment growing up, poor socio-economic status, and a family history of mental health difficulties. Precipitating factors associated with an atheoretical five Ps framework may include: loss of job, recent bereavement, and a breakdown of a relationship. These factors are known to increase the risk of mental health difficulties, but do not explain the development of specific disorders such as psychosis.

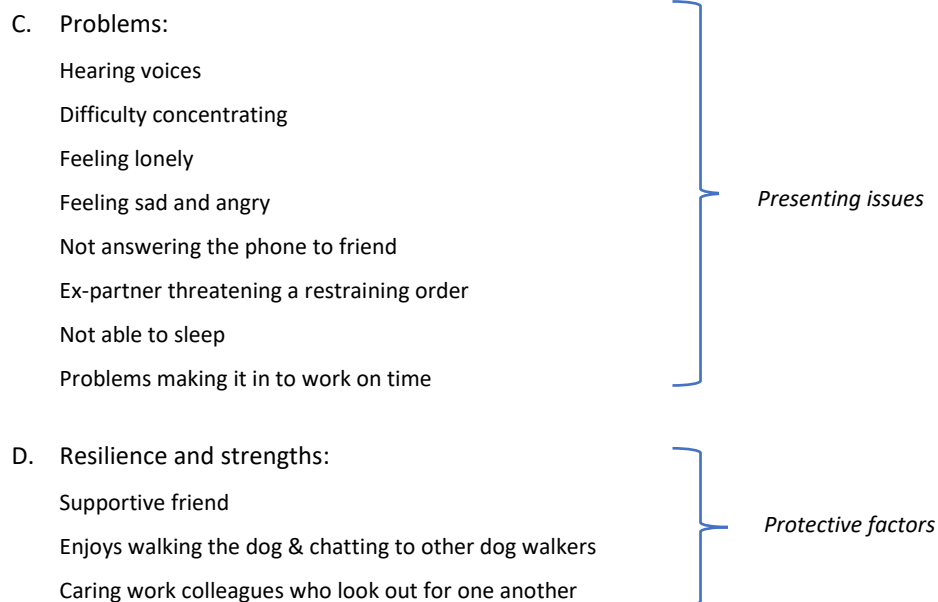
Nevertheless, the five Ps framework is also compatible with theory based psychological models, such as CBT. For example, Figure 1.1 below (adapted from Dudley & Kuyken, 2014) provides a brief description and illustration of how a person's recent onset of voice hearing may be formulated using the five Ps framework. The figure includes elements that are consistent with those found within a CBT model (e.g. core beliefs, rules and assumptions, and compensatory strategies). However, as noted, the five Ps approach is atheoretical. Therefore, it does not *need* to draw on a specific model, or empirical basis, to support elements of its framework.

In this example, in Figure 1.1, a person viewed themselves as unlovable (core belief) owing to developmental experiences of parental neglect and abandonment. As with all core beliefs, their belief about ‘unlovability’ was deep rooted and emotionally salient (Morrison, 2007). Yet, prior to the ‘triggering event’ (e.g. the person’s relationship ending, and subsequent difficulties in finding a new partner), they had (temporarily) succeeded in overriding their core belief by adopting rules, assumptions, or conditional beliefs to ensure it lay dormant (e.g. “*If I am in relationship, then I am OK*”). These rules, assumptions, or conditional beliefs may also take the form of: “I should” or “I must” statements (e.g. “*I must always be in a relationship*”; Dudley & Kuyken, 2014). Rules, assumptions, or conditional beliefs link to compensatory strategies that keep the person tied to their core belief system (e.g. giving their all to a relationship, to prevent re-abandonment). When the relationship ended, their rule was broken, which exposed/re-awakened their core belief. This ‘triggering event’ acted as a precipitant to the presenting issues (e.g. loneliness), which subsequently became maintained by perpetuating factors (e.g. ringing and harassing their ex-partner). Finally, protective factors (e.g. having a supportive friend), indicate areas of strength and resilience which can be drawn upon at times of need (Dudley & Kuyken, 2014).

The five Ps is also investigated in more detail in Chapter 3, in relation to expert consensus regarding the essential ingredients of a CF for auditory hallucinations and persecutory delusions in schizophrenia spectrum disorders (Spencer et al., 2020).

Figure 1.1 An example of a five Ps formulation for early psychosis





1.5 Conceptual underpinnings of a CBT case formulation

Underpinned by the scientist-practitioner model, CF synthesises the subjective and unique experience of an individual, with an empirically supported psychological theory or model (Butler, 1998; Kuyken et al., 2009; Tarrier & Calam, 2002). To some extent, as humans, we are all trying to formulate an understanding of our difficulties day-to-day with one other. However, it is the use of this empirical evidence that differentiates CF from the usual process of making sense of one's difficulties (for example, via having a chat with a friend; Frankl, 1960; Johnstone, 2018). Done well, it provides an explanation of the origins, development, and maintenance of the problem (Tarrier & Calam, 2002). Perhaps most importantly, "*it informs the therapist about what to do next*" (Eells, 2022, p. 2).

As such, a CF is viewed as *the* overarching process, scaffolding, or framework, from which cognitive therapists undertake their whole treatment approach (Kuyken et al., 2009; Routledge Mental Health's Podcast, 2010). The flexibility imbedded within this framework, enables both the SU and therapist to shape, mould, and refine the formulation as therapy progresses (Tarrier, 2006b). Indeed, in CBT, service users (SUs) and therapists work collaboratively to co-construct (and develop) a shared understanding of the presenting difficulties (Johnstone & Dallos, 2014). SUs contribute their understanding of the problem, their current perceptions (in terms of thoughts,

feelings, and behaviours), and life history. Whereas therapists contribute their ideas based on empirical research, psychological models, and past experiences in working with SUs with similar difficulties (Kuyken et al., 2009).

As Eells stated, the main goal of a CF is: *“to help the patient as quickly and efficiently as one can (...) to facilitate a positive outcome”* (Psychotherapy Expert Talks, 2016a, 8:16). Therefore, assessing whether a CF is functional (or not), largely depends on a) whether it closely resonates with the SU’s experience to provide a meaningful account of their idiosyncratic problems and difficulties (Kuyken et al., 2009; e.g. it should ‘ring true’ for the individual; Beck 2020) and b) whether it contributes to a good clinical outcome (Tarrier, 2006a). As Tarrier and Johnson stated, the function of a CF *“must have both clinical utility and subjective utility”* (Tarrier & Johnson, 2016, p.6).

However, developing a formulation has been described by Persons as: *“one of the hardest parts of clinical work”* (Psychotherapy Expert Talks, 2016b, 15:57). This may be because the process of CF is not static (Psychotherapy Expert Talks, 2016a), it unfolds over time as the formulation is revised and refined. Owing to this, a CF is never ‘complete’; it is an evolving hypothesis that both informs and is informed by therapy (Kuyken et al., 2009). Moreover, feedback is sought from the SU along the way, and the CF is modified in-line with new information (Beck, 2020; Eells, 2022; Johnstone & Dallos, 2014; Psychotherapy Expert Talks, 2016b); such as via additional assessment, in-depth discussion, observation, behavioural experiments, or via the implementation of intervention strategies; Bennett-Levy et al., 2004; Quinlan & Deane, 2021). Therapists then confirm, disconfirm, or modify their hypotheses accordingly (Beck, 2020). Indeed, a formulation is a working hypothesis to be tested, not a statement of fact (Butler, 1998). Tolerating this uncertainty is undoubtedly challenging for therapists, particularly for those in-training (Quinlan & Deane, 2021). As such, higher levels of ‘intolerance of uncertainty’ have been associated with reduced confidence in CF skill (Quinlan et al. 2022).

Developing a formulation is also viewed as a process that occurs ‘within the therapist’s head’, either during, or in-between therapy sessions (e.g. in supervision; Kuyken et al., 2009). Indeed, a complex formulation ‘inside the therapists head’, may rupture therapeutic alliance if fully shared with the SU (e.g. a SU would not be expected to understand complex psychological theory; Kuyken et al., 2009). Moreover, it has also

been argued that, for some SUs with complex mental health difficulties (e.g. systematised delusions), it would be unadvisable for the therapist to be fully explicit in sharing their CF (e.g. the CF may be rejected by the SU, particularly if shared too early on in therapy). As Eells explained: *“sometimes a formulation may be too strong a medicine (...) part of the task of the therapist is to present the formulation in a way that is responsive to where the patient is coming from at a given point in time”* (Psychotherapy Expert Talks, 2016a, 11:58). Indeed, a CF should enhance the therapeutic alliance, not damage it, so therapists need to use their clinical judgment to present a CF in a sensitive and considered way (Psychotherapy Expert Talks, 2016a). On the flip side, creating and sharing a formulation collaboratively, decreases the likelihood of misinterpretation (Kuyken et al., 2009). Therefore, the notion of being openly transparent, is a careful balancing act.

CFs are also often represented in the form of a visual diagram to highlight interrelationships between different aspects of a SU’s difficulties (Kuyken et al., 2009). As once stated by Butler: *“formulations, just like maps, provide an overall view (often in diagrammatic, conventional form) of something that is not possible to see directly all at once – the wood as well as the trees”* (Butler, 1998, p. 9). Nevertheless, SUs’ have argued that formulation diagrams should be as parsimonious and as simple as possible *“in case it becomes like spaghetti”* (Spencer et al., 2023, p. 338), i.e. too elaborate/complex to understand, and/or too difficult to remember (see Chapter 5, and Spencer et al., 2023).

The basis of a cognitive behavioural formulation is the cognitive model of emotional disorders. This suggests that the way individuals perceive events, affects how they feel and behave (Beck, 1964; Beck et al., 1979; Beck, 2020; Ellis, 1962). Therefore, a CBT CF places emphasis on understanding the appraisal - how the individual has interpreted key events/experiences in their lives, and the *personal meaning* they have ascribed to those events (Beck, 1964; Ellis, 1962). The appraisal then makes sense of any distress and maladaptive behaviours that serve to maintain the disorder (Butler, 2006; Greenberger & Padesky, 1995). As Beck stated: *“people’s reactions always make sense once we know what they are thinking”* (Beck, 2020, p. 29). Therefore, the aim of a CBT maintenance formulation, is to describe and to ‘map out’ thoughts, feelings, and behaviours, to generate an understanding of what has led to, and what is perpetuating, the distress and difficulty (or particular set of difficulties) that a

person brings with them to therapy (Butler, 2006; Greenberger & Padesky, 1995). In the context of a longitudinal formulation, this may also include identifying underlying psychological mechanisms, related to past experiences (Persons, 1989). This ‘sense making’ process, may be particularly helpful in understanding complex presentations such as psychosis, which at first might appear to be bizarre, nonsensical, or perplexing to both the therapist and SU (Spencer, 2019).

There are many key features and functions of a CBT CF. For example, it is said to: 1) normalise and validate the SUs problems, 2) provide a hypothesis that is open to revision and re-formulation, 3) draw on a psychological theory or model, 4) provide an evidence-based explanation for the development and maintenance of the SU’s difficulties, 5) provide a framework for the selection of the most simplest and cost-effective interventions, 6) guide the intervention towards the therapy goal(s), 7) alleviate SU distress, 8) promote SU engagement, 9) strengthen the therapeutic alliance, 10) anticipate therapeutic impasses, 11) help the SU and therapist to move past areas of stuckness, 12) identify SU strengths, resources, and ways to build resilience, and 13) enable high quality supervision and consultation (Beck, 2020; DCP, 2011; Butler, 1998; Johnstone & Dallos, 2014; Kuyken et al., 2009; Tarrier & Calam, 2002).

1.6 Different psychological approaches, theoretical perspectives, and a multitude of models! The conundrum when developing a case formulation

Formulation is a widely utilised process across all evidence-based psychological approaches (Jenkins, 2020). However, there are marked differences in the ways in which the *content* and *process* of a formulation is developed, shared, and applied (depending upon the psychological approach, theoretical perspective, and/or model that is used; Johnstone & Dallos, 2014). For example, even within the same psychological approach (e.g. CBT), there may be different theoretical perspectives, and different models that could be used when conceptualising a particular disorder (e.g. generalised anxiety disorder; GAD). This leaves scope for a diverse range of models to draw upon when formulating, each with their own unique focus/emphasis (Freeston, 2023).

Sections 1.6.2 to 1.6.4 below, provide a brief discussion of several other psychological approaches, to highlight some of the different ways in which psychosis might be conceptualised.

1.6.1 What are the differences between a formulation and a model?

Models are ways of conceptualising disorders or difficulties based on a particular psychological approach, and/or theoretical perspective. The model is then translated into a formulation (Butler, 1998), and structured templates with headings to guide formulation content are often utilised to assist with this process (e.g. see Beck et al., 1979, and Morrison, 2017).

A model provides guidelines for a formulation. However, it operates at a certain level of generality (Butler, 1998), whereas the content of a formulation is bespoke to an individual. *“Thus, the formulation illustrates, in ways that are clinically relevant, how the model applies, and does not apply, to the case”* (Butler, 1998, p. 7).

Formulations can be developed at the problem or case level. The problem level focuses on a specific issue, whereas the case level takes account of all the person’s difficulties (Flinn et al., 2015). If an individual is experiencing several comorbid difficulties (e.g. psychosis, alongside panic disorder, and agoraphobia), then it may be useful to a) draw on a trans-diagnostic approach to CF i.e. formulate the psychological processes that appear to co-occur across the range of disorders (such as intolerance of uncertainty; Bottesi et al., 2019; thought suppression; Markowitz & Purdon, 2008; *“which if targeted may produce change in several presenting issues”* Dudley et al., 2011, p. 215), b) draw on several models to construct a single formulation, or c) utilise a model with a higher order of generality (Butler, 1998; for example, Beck et al.’s 1979 cognitive model of emotional disorders). Consequently, a certain amount of trial and error (or artistry) may be needed in constructing a formulation, to fully explain the case. This is because: *“the absence of a meta-model to explain all data makes this trial and error unavoidable”* (Perry et al., 1987, p. 546).

In psychosis, for example, different psychological approaches (and models) may be used by different therapists to conceptualise the same presentation. Those models will all differ in their understanding of psychosis and their approach to treatment.

1.6.2 Psychodynamic formulation

Psychodynamic psychotherapy draws on the principles of psychoanalytic theory of which there are several main theoretical perspectives (e.g. ego-psychological, self-

psychological, and object relations; Perry et al., 1987; Ruffalo, 2020). The field has also recently been expanding in new directions (e.g. neuropsychanalysis; Cieri & Esposito, 2019; Mosri, 2021; Ramus, 2013); and each of these schools of thought have their own models. There are also shortened variations of psychodynamic psychotherapy (e.g. brief/time-limited psychodynamic therapy; Crits-Christoph, 1992; Levenson & Strupp, 2010) with some brief psychodynamic models focusing mainly on symptom reduction (Horowitz, 1991a), whilst other models target resolution of the oedipal conflict (Laikin et al., 1991).

Each psychodynamic theoretical perspective brings a core set of ideas and assumptions that shape the way CFs are constructed (Leiper, 2014). For example, in object relations theory the focus of a CF relates to an individual's internalised relationship with 'objects' (i.e. early attachments), usually the mother (Fritscher, 2020). Object relations is concerned with how individuals replay these relationships throughout adult life in an attempt to master them (Messer & Warren, 1995; Ruffalo, 2020). An example of object relations theory applied to psychosis, provided by Summers and Martindale (2013), referred to Jane (pseudonym) who developed a delusional belief that her newborn baby was actually her mother. This led to a formulation that the baby who needed so much care and attention was felt to be starving Jane, herself, of care. This unconsciously mirrored the way Jane felt when she had been neglected by her own mother as a child. Consequently, the baby (misidentified as her mother), became the 'object' of resentment and rage (Summers & Martindale, 2013).

Despite the different focuses, there are key characteristics and shared values within psychodynamic psychotherapy that loosely tie these theoretical perspectives together (Leiper & Maltby, 2004; Ruffalo, 2020; Wallerstein, 2002). Owing to this, psychodynamic CFs generally include a focus on a) conscious and unconscious processes, b) failing defence mechanisms that have given way to psychological disturbance, c) underlying core conflicts (e.g. 'I hate you, *but* I depend on you for my survival'), and d) repeated relationship patterns that are observed over time, across different relational contexts (Leiper, 2014; Ruffalo, 2020; Summers & Martindale, 2013). Psychodynamic formulations are then validated by observing the transference, countertransference, and the SU's response(s) to the therapists' interpretations (Leiper, 2014).

In contrast to CBT, psychodynamic CFs are developed as an individualised statement or narrative, and often, they are not explicitly shared with the SU (or only selected parts of the formulation are shared; Friedman & Lister, 1987; Summers & Martindale, 2013). The formulation may be developed away from the SU via group discussion with other professionals involved in the case, allowing time to reflect on their different countertransference responses (Summers & Martindale, 2013).

1.6.3 *Systemic formulation*

There are several approaches to family therapy including narrative, solution-focused, and social constructionist (Heatherington & Johnson, 2019). Socially constructed approaches to formulation emphasise culture and use of language (e.g. what it means to be ‘a good mother’, ‘the man of the house’, ‘a close family’; Dallos & Stedmon, 2014). These cultural ideas shape people’s beliefs and expectations about how families *should* be (e.g. the picture postcard of families coming together to be happy at Christmas; Dallos & Stedmon, 2014). A social constructionist approach also recognises multiple realities in the minds of individual family members (and hence multiple formulations; de Paula-Ravagnani et al., 2017), with individual family members encouraged to express their own perception of the formulation, whether collectively agreed or not (Dallos & Stedmon, 2014).

Systemic formulations also aim to understand patterns of family interaction and relational dynamics, with ‘symptoms’ sometimes viewed as representing ‘problems in communication between family members’, rather than residing within an individual (Dallos & Stedmon, 2014). For example, a young person’s psychosis may be understood as *functional* in the sense that it helps stabilise the family system - temporarily bringing both parents together, diffusing their marital conflict as they focus on the care of their child (Byng-Hall, 1980).

Like CBT, systemic formulations are collaborative (Dallos & Stedmon, 2014). However, in family therapy, the formulation may include others’ perceptions of the family’s difficulties (e.g. social services, the child’s school, the observation team¹;

¹ In ‘formal’ family therapy, an observation team observe the interactions between the individual family members, and the therapist via an observation screen, or by observing live in the therapy room. The observation team then share their formulation(s) with the therapist, in the presence of the family (Dallos & Stedmon, 2014).

Dallos & Stedmon, 2014). The observation team may share their thoughts in relation to the therapist-family interactions, and ask whether these interactions mirror the communication style within the home? As Dallos and Stedmon explained: *“by sharing our ideas with them [the family] we move towards a co-constructed formulation”* (2014, p. 74). Therapist-family relations are therefore viewed as essential to the formulation (Dallos & Stedmon, 2014).

In systemic formulation, there is less of a distinction between the stages of assessment-formulation-intervention, as in other therapies. Instead, these processes are much more intertwined (Dallos & Stedmon, 2014). For example, systemic formulations may start with a genogram of the immediate family and their relationships to external sources of support (Dallos & Stedmon, 2014).

The ‘progressive hypothesising’ approach to systemic formulation (whereby it is viewed as fluid, and recursive; Selvini et al., 1980) means that it is not judged in terms of offering an ‘objective truth’ but instead, by its usefulness in providing a route to change (Dallos & Stedmon, 2014).

1.6.4 Cognitive Analytic Therapy reformulation

Cognitive analytic therapy (CAT) integrates cognitive and psychoanalytic theory and practice, as well as Vygotskian and Bakhtinian ideas (Ryle & Kerr, 2020; Taplin et al., 2018). In CAT, CF is referred to as ‘reformulation’ (Ryle, 1994, 2003, 2005; Ryle & Kerr, 2020), and like CBT, the process is collaborative, exploratory, and open to revision (Jefferis et al., 2021). Reformulation is also time sensitive - often taking place within the first three to four sessions of therapy (Shine & Westacott, 2010). Reformulation ‘tools’ describe the development and maintenance of a SU's problems in both a written (reformulation letter), and diagram (‘mapping’) format, known as a Sequential Diagrammatic Reformulation (SDR; Taplin et al., 2018).

Firstly, a provisional reformulation letter is written from the therapists’ viewpoint, and ‘presented’ (read aloud) to the SU (Ryle & Kerr, 2020; Shine & Westacott, 2010; Tyrer & Masterson, 2019). The reformulation letter aims to encourage self-reflection in relation to childhood experiences - linking the past with the present (Dallos, 2006; Ryle & Kerr, 2020; Tyrer & Masterson, 2019). In particular, the reformulation letter aims to elucidate how problematic ‘reciprocal role procedures’

(RRPs - repetitive sequential patterns of thoughts, feelings, behaviours, and sensations; Jefferis et al., 2021), typically reinforce (formative) ‘reciprocal roles’ (RRs) that underly them (Ryle & Kerr, 2020). Unhelpful RRs are repetitive ways of relating to oneself (self–self), or others (self–others, and others–self; Ryle, 1985; Tyrer & Masterson, 2019), and tend to be the ‘default’ position that individuals adopt (Jenaway, 2019).

The CAT model of voice hearing in psychosis (Kerr et al., 2003, 2006), which centres on RRs, is therefore seen as an ideal approach to improve a person’s relationship with their voices (Perry, 2012). *“RRs help us to establish with our clients— in a simple, collaborative fashion— an understanding of voice- hearing experiences as mirrors or echoes of, or responses to, previous life experiences and internalised relationships”* (Luce & Barclay, 2022, p. 192). For example, Isabella (pseudonym) was shy, submissive, and a ‘people pleaser’. She had learnt to adopt this way of relating to others, to survive a domineering and controlling relationship with her mother (self–others). Isabella’s voices were also authoritarian and commanding, and she responded fearfully towards them too (self–others). Continuing to adopt a submissive role throughout adulthood, led others to take advantage of her, and allowed her voices to ‘win’, as she succumbed to their demands (others–self). This led to self-directed anger/self-loathing in how she related to herself (self–self); i.e. Isabella was deeply unhappy with herself for being a ‘push over’.

In the reformulation letter, therapists’ may also share their experience of ‘projective identification’ and what it feels like to be in a (therapeutic) relationship with the SU (e.g. feeling the need to control the SU, or feeling inadequate/apologetic around them; Dallos, 2006), and this can be powerfully revealing (Jenaway, 2019). Reformulation also helps to point out (or anticipate) transference - possible re-enactments of unhelpful reciprocal roles within the therapeutic relationship (Jefferis et al., 2021; Potter, 2010; Tyrer & Masterson, 2019), for example, with the SU acting (unconsciously) towards the therapist as if they were a parent or caregiver (Dallos, 2006). Nevertheless, it is argued that narrative descriptions are not always clear in demonstrating the links between RRP, and underlying RRs (Ryle & Kerr, 2020). Hence, co-constructing a CAT map/SDR compliments the reformulation letter by illustrating these relational patterns (Jefferis et al., 2021; Shine & Westacott, 2010), with the SDR suggesting ways in which to ‘exit’ from them (Jenaway, 2019; Tyrer & Masterson, 2019).

There are many ways to construct a reformulation diagram (Jefferis et al., 2021). However, it is recommended that therapists ‘map out’ provisional ideas on paper (Jefferis et al., 2021; Ryle and Kerr, 2020), and share the activity of mapping with the SU to enable joint ownership of the process (Potter, 2010; Taplin et al., 2018). Like CBT, the SU is encouraged to sit with the diagram (both within and between therapy sessions), to reflect on it, and bring forward their own ideas (Jefferis et al., 2021). SUs have reported that reformulation (in both a letter and SDR format) offers something ‘tangible’ (Taplin et al., 2018; Shine & Westacott, 2010) for the assimilation of a new perspective about themselves and their relationships, that describes a way forward for treatment (Ryle & Kerr, 2020).

1.7 Formulation across generic models of CBTp

Whilst several models can be applied to the formulation of psychosis (and newer models, such as CAT, demonstrate interest/hold promise in its treatment; see Taylor et al., 2019), CBTp currently represents one of the best evidenced based (and recommended) psychological treatments for psychosis (National Institute for Health and Care Excellence; NICE; 2014), hence this is the focus of sections 1.7 and 1.8.

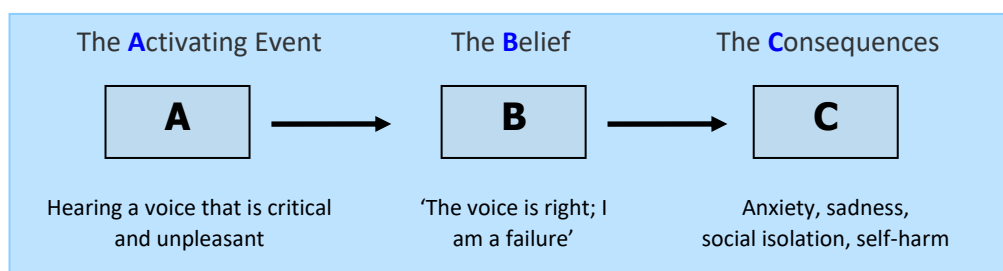
Generic CBT models propose that perceptions of an event, rather than the event itself, directly influences our emotions, physiology, and behaviour (Beck, 1963, 1964). Sections 1.7.1 to 1.7.3 below², outline some of the ways in which psychosis may be conceptualised across evolving levels of formulation - moving from descriptive, to maintenance, then to longitudinal levels of explanation (if appropriate; Kinderman & Lobban, 2000; Kuyken et al., 2009). Different levels of formulation are explored in more detail Chapter 5, in relation to SU experiences (also, see Spencer et al., 2023). However, brief descriptions are provided here with illustrative examples.

² As noted at the start of this chapter, elements of sections 1.7.1 to 1.7.3 were published as a book chapter for SUs and carers (see Spencer, 2019). The purpose of this book chapter was to explain to SUs and carers how they may be able make sense of psychosis, using simple formulation diagrams.

1.7.1 A-B-C model

An A-B-C model, or formulation (Ellis, 1957), shows how thoughts, feelings and behaviours may be linked together, for example in relation to voice hearing (see Figure 1.2 below).

Figure 1.2 An example of an A-B-C formulation for voice hearing

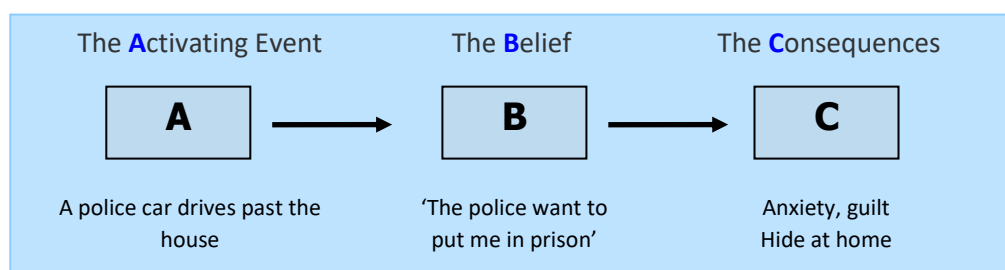


In other words, the **A** (Activating event) triggers the **B** (Belief) that the person believes to be true, which leads to the **C** (Consequences). In this example, the **A** refers to a person hearing a critical voice. Following this experience, the emotional **C**, are feelings of anxiety and sadness. The behavioural **C** might then be social isolation, or even self-harm.

A therapist or carer, can help the person to make sense of the **B** - what are the thoughts or beliefs that the person has about these voices? In this example it would be *‘the voices are telling the truth. I am a failure’*. This, of course, can be worked with in CBT by looking for evidence of this presumed sense of ‘failure’ and coming up with some alternative, rational responses to what appears to be a global unrealistic belief.

Another A-B-C formulation, for example in relation to a persecutory belief may look something like Figure 1.3 below:

Figure 1.3 An example of an A-B-C formulation for a persecutory belief



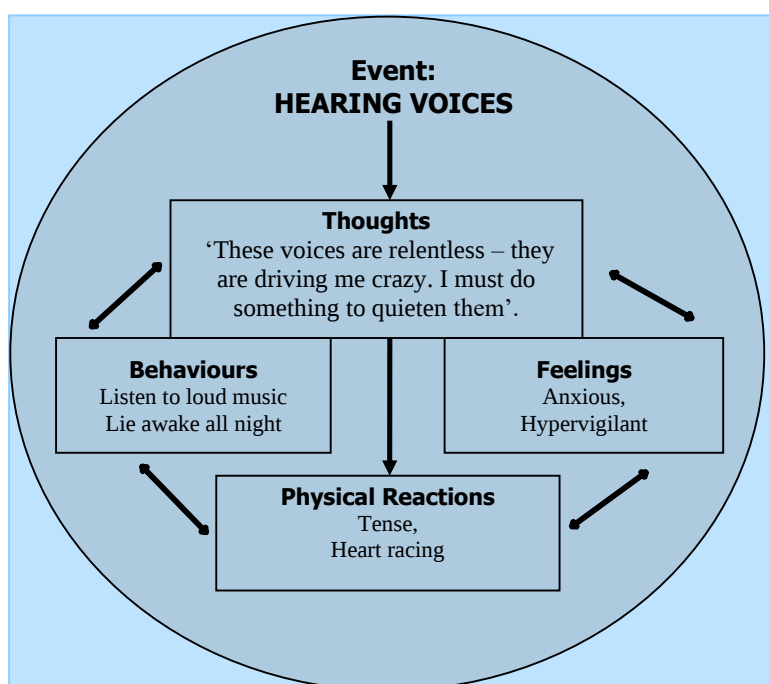
The **A** would be the activating event - seeing a police car drive past the house. The **B** would be the delusional belief itself - that the police are looking to arrest and imprison the person for no real reason at all. The emotional **C** would be severe anxiety and guilt, and the behavioural **C** would be to find a safe place to hide at home. Here the A-B-C helps to make sense of symptoms that appear to be puzzling and complex by mapping them out in a simple, linear diagram. Once we have made sense of the persecutory belief in this way, the next step might then be to ask the SU to make a diary record of how often police cars drive past his house - is this excessive, or the same as in other nearby streets? Are police cars really driving by as often as he thinks and, if they are, could there be an alternative explanation (i.e. is there a police station nearby that they are driving to and from?).

1.7.2 *Maintenance model*

The symptoms of psychosis are often made worse by feelings of panic, shame, and self-stigma. These feelings develop alongside 'safety behaviours' as people try to keep themselves safe and manage the frightening symptoms as best as they can. For example, listening to loud music may be used as a safety behaviour to 'drown out' the voices. The loud music prevents the person from hearing the voices' instructions (i.e. commands) to do things they do not wish to do. For someone with persecutory delusions, the person might close all the curtains in their house to prevent government surveillance. Although these safety behaviours may be seen as understandable responses to certain appraisals, unfortunately they are not always the best means of coping. Indeed, they can often make things worse or exacerbate symptoms by perpetuating the problem in an ongoing vicious cycle. Hence, maintenance formulations focus on perpetuating factors and

attempt to address questions such as: ‘what is keeping the problem(s) going?’ (Dudley & Kuyken, 2014). This process can be represented in a diagram which CBT therapists often refer to as a maintenance formulation (see Figure 1.4 below).

Figure 1.4 An example of a maintenance formulation for voice hearing



In psychosis, a generic maintenance formulation (e.g. Greenberger & Padesky, 1995) like the one shown above, is often presented in a circular diagram. This helps to show how certain thoughts, feelings, physical reactions, and behaviours (or ways of reacting to an event) can actually maintain the problem and keep it going in a way that is difficult (but not impossible) to break. If we refer back to the person with critical voices, we can see how ‘safety behaviors’ such as listening to loud music can make the person feel on edge and hypervigilant, and lying awake all night can lead to sleep deprivation. Both serve to maintain the experience of hearing voices.

1.7.3 Longitudinal model

One of the most crucial questions in making sense of psychotic symptoms is to ask: ‘how did this all start?’ and ‘why me?’. Longitudinal formulations help to address these questions by focusing on predisposing factors and precipitating factors (Dudley &

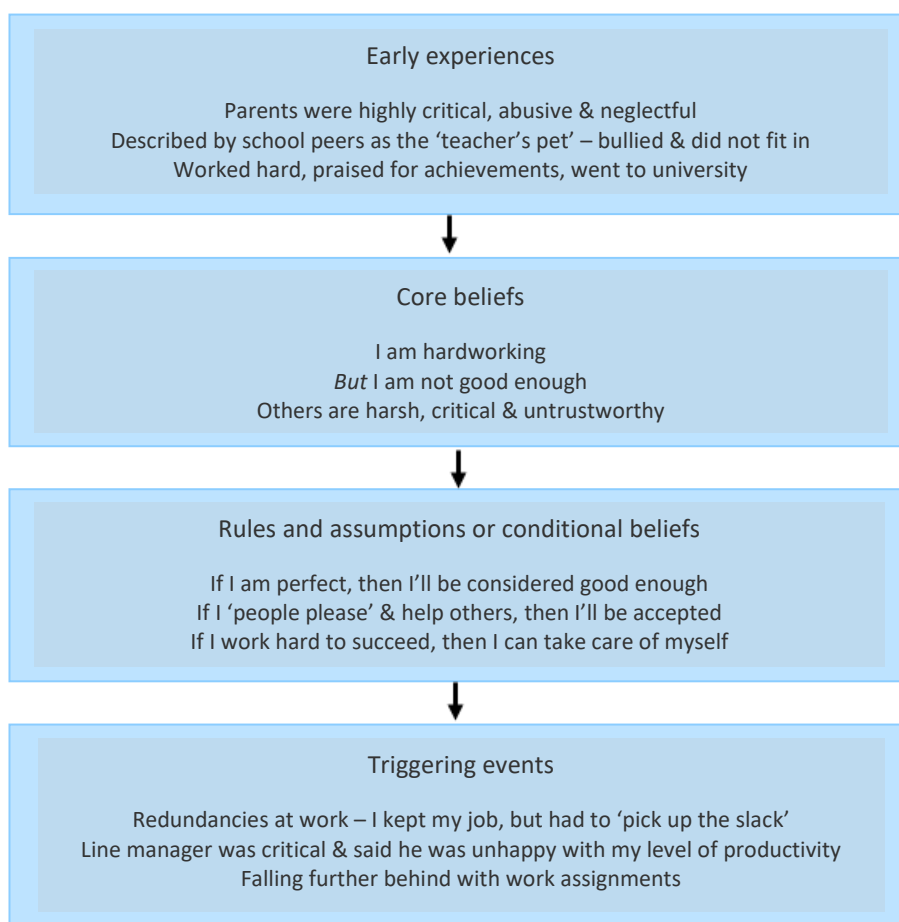
Kuyken, 2014). Longitudinal formulations (e.g. Beck et al., 1979) can help the person discover that there was a trigger (or multiple triggers) in the person's life that can explain why their symptoms developed in the first instance. Timelines may facilitate this process by acting as a detailed method for eliciting this information, to help inform a longitudinal formulation.

Longitudinal formulations can help to normalise psychotic symptoms based on triggers or past life events, such as a family history of schizophrenia, a traumatic life event, childhood abuse, and many other factors that may have resulted in a 'tipping point'. In CBT, the person can learn that each and every one of us could develop psychosis if we were pushed to our limits. Longitudinal formulations can also incorporate positive elements if they help the person recognise their areas of strength and resilience, instead of just highlighting all that is wrong with them (Kuyken et al., 2009). For example, how have they managed to cope with the stresses and strains of life so far? What are they good at? And what areas of strength can they draw upon to help them cope better with this particularly difficult time in their lives?

Referring back to the person who heard critical voices, Figure 1.5 below, provides an example of the pre-psychotic period, and the person's early childhood experiences (e.g. criticism and neglect from parents, being bullied at school, but also experiencing successes in life for working hard). These childhood experiences led to the development of core beliefs in relation to themselves and others (e.g. the belief that they are hardworking, but also the belief that they are not good enough). As well as compensatory rules for living, which they adopted to cope with the core beliefs they had internalised (e.g. believing that if they strive to do things perfectly, then they will avoid criticism and hopefully be considered 'good enough'). Finally, events describe what triggered the voice hearing in the first instance (e.g. an increase in workload meant the person could not keep up with their work whilst adhering to their own 'perfect' standards, resulting in criticism from their line manager - which they had worked so hard to avoid).

The complex deeper layers of longitudinal formulation can expose powerful and raw messages about the person when revisiting key, traumatic, or difficult issues in their upbringing.

Figure 1.5 An example of a longitudinal formulation in the pre-psychotic period



1.8 Formulation across disorder-specific models of CBTp

Disorder-specific CBT models propose key processes, beliefs and assumptions that are thought to account for a particular disorder (e.g. psychosis; Dudley & Kuyken, 2014). These highly specialised models have evolved over time, with the development of theory, empirical evidence, and clinical practice. As referred to in Chapter 3, within CBTp there are a variety of specific models for voice hearing (e.g. Kinderman, 2011; Morrison, 2001), and persecutory delusions (e.g. Freeman 2016; Turkington et al., 2011) which help guide the conceptualisation and treatment of individual psychotic symptoms. Disorder-specific models also emphasise different processes. For example, in relation to paranoia and persecutory beliefs, roles have been identified for 'jumping to conclusions' (JTC; Dudley et al., 2011), worry (Startup et al., 2016), rumination (Lebert et al., 2021), sleep (Waite et al., 2016), and safety behaviours (Freeman et al.,

2007). “Of course, this then begs the question of which models clinicians should draw on” (Dudley et al., 2011, p. 213).

Sections 1.8.1 and 1.8.2 below, outline some of the ways in which positive symptoms of psychosis may be conceptualised across disorder-specific models in CBT. Section 1.8.1 provides a brief description and illustration of Morrison’s cognitive model of hallucinations and delusions (Morrison, 2001, 2017). This model was employed on the ACTION (Assessment of Cognitive Therapy Instead of Neuroleptics) trial (Morrison et al., 2014), and is referred to in a secondary analysis of ACTION (see Chapter 4, and Spencer et al., 2018). The model was also utilised by some of the therapists in a qualitative study exploring the impact of different levels of formulation (see Chapter 5, and Spencer et al., 2023). Section 1.8.2 provides a brief description and illustration of Freeman’s cognitive model of persecutory delusions (Freeman, 2016). This model is currently being rolled out across NHS psychosis services in the UK as part of the ‘Feeling Safe Programme’, owing to its recent success in treating persecutory delusions within a randomised controlled trial (RCT; see Freeman et al., 2021).

1.8.1 *Morrison’s cognitive model of hallucinations and delusions*

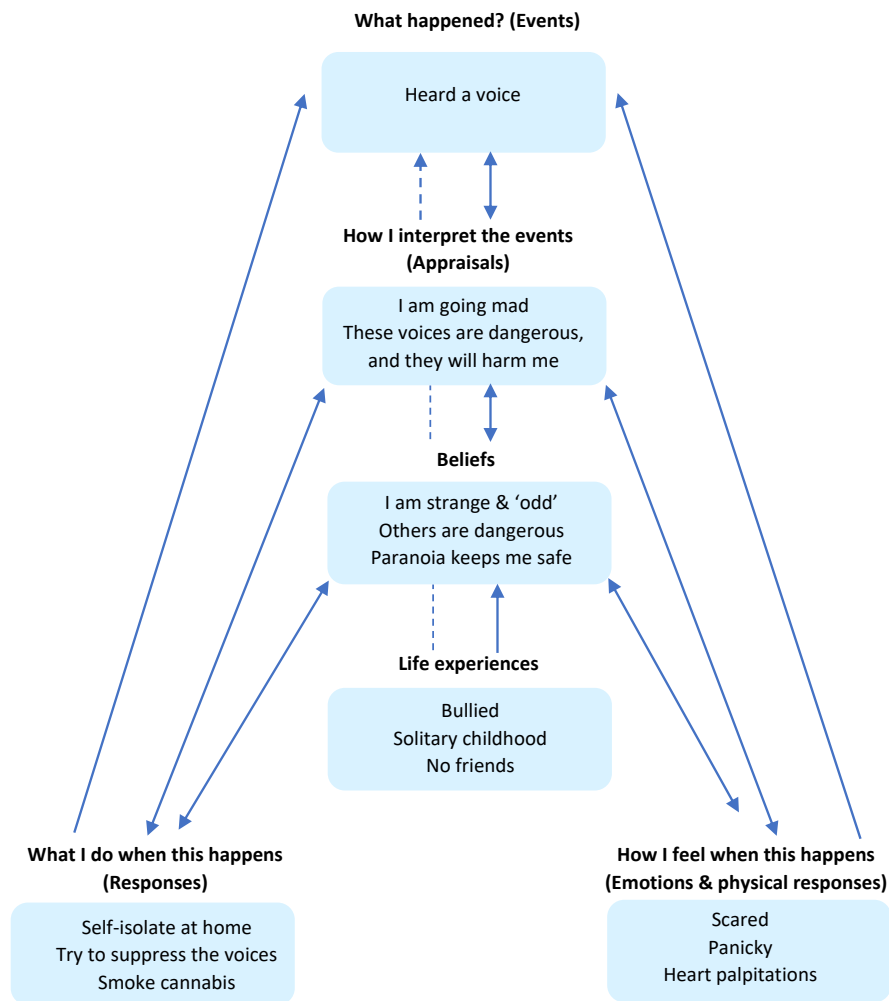
Morrison’s (2001, 2017) longitudinal model (which incorporates a maintenance cycle) argues that it is the unusual *appraisal* (i.e. *interpretation*) of an event, that leads to distress. The event may be an anomalous internal or external experience (e.g. ‘*someone looks at me*’), which is then misinterpreted as physically or psychologically threatening to the individual (e.g. ‘*That person looks suspicious, he could be a Russian spy*’; Morrison, 2001, 2017). The nature of the misinterpretation is influenced by the person’s life experience, core beliefs, and “*faulty self and social knowledge*” (Morrison, 2001, p. 257). In this example, the misinterpretation becomes more likely if the person has had experiences of serving in the military, has become hypervigilant to threat, and has formed a core belief about the dangerousness of others. In turn, the external anomalous experience (and appraisal in relation to that experience), continue to be maintained and exacerbated by mood, physiological arousal, and cognitive and behavioural responses (e.g. safety behaviours, including selective attention and counterproductive control strategies), that result in a vicious cycle (Morrison, 2001). Interestingly, Morrison explains how his cognitive model shares similarities with models for anxiety disorders,

the difference being, that a misinterpretation is deemed to be ‘psychotic’ if it is perceived as culturally unacceptable (Morrison, 2001).

For instance, if someone misinterprets their racing thoughts or palpitations as a sign of alien control or persecution via telekinesis, they will be classified as delusional, whereas misinterpretation of the same sensations as a sign of impending madness or heat attack would be regarded as indicative of panic disorder (Morrison, 2001, p. 260).

Figure 1.6 below, provides an example of a CF based on Morrison’s cognitive model of psychosis. This example illustrates how an event such as internal anomalous experience (e.g. hearing a voice), can lead to a ‘psychotic’ misinterpretation (e.g. *‘I am going mad; these voices are dangerous, and they will harm me’*), particularly if a person is not aware of how common the experience of voice hearing is, within the general population (Volpato et al., 2022). Indeed, in this example, the misinterpretation was more likely given the person’s life experiences (e.g. bullied at school, and treated ‘differently’ by their peers), and the subsequent core beliefs they had formed about themselves (e.g. *‘I am strange, ‘odd’ and different in some way’*). In this case example, the voice is then maintained by how the person feels (e.g. scared), their physiological responses (e.g. heart palpitations), and their behavioural responses (e.g. self-isolating at home to stay safe).

Figure 1.6 An example of a longitudinal formulation for auditory hallucinations, based on Morrison's cognitive model of psychosis



1.8.2 *Freeman et al.'s cognitive model of persecutory delusions*

Freeman's (2016) cognitive model of persecutory delusions highlights six key psychological processes (in the context of genetic and environmental risk factors; Bentall et al., 2012; Freeman et al., 2011; Shakoore et al., 2015; Sieradzka et al., 2014; Zavos et al., 2014), that are thought to be implicated in the initial development and maintenance of such beliefs (Freeman, Waite et al., 2016; Freeman et al., 2021). These processes include: 1) excessive worry, 2) low self-confidence, 3) anomalous

experiences, 4) sleep dysfunction, 5) reasoning biases, and 6) safety-seeking behaviours (Freeman 2016; Freeman, Waite et al., 2016; Freeman et al., 2021).

Evidence supports the benefits of targeting each of these causal mechanisms individually (see for example Freeman, Bradley, Antley et al., 2016; Freeman, Dunn, Murray et al., 2015; Freeman, Dunn, Startup et al., 2015; Freeman et al., 2014; Freeman, Waite et al., 2015; and Waller et al., 2015); and as a byproduct of targeting each mechanism, the paranoia comes down.

More recently, these individual processes have been combined into a multifactorial model, which has led to promising results from the ‘Feeling Safe Programme’ (see Freeman, 2016; and Freeman et al., 2021). Freeman’s model proposes that “*safety has to be relearned*” (Freeman, 2016, p. 685). Owing to this, SUs are supported to enter feared situations after their maintenance factors (from the list of six) have been identified, and reduced (Freeman, 2016; Freeman et al., 2021).

Figure 1.7 below, provides an example of a vicious flower formulation based on Freeman’s cognitive model of persecutory delusions (Freeman, 2016). This disorder-specific conceptualisation identifies each psychological process (e.g. worry, insomnia, and safety-seeking behaviours), which can then be individually formulated (Freeman et al., 2021) using a generic CBT maintenance formulation (e.g. Greenberger & Padesky, 1995).

At the centre of the vicious flower, is the strong belief that the person is under threat, with others deliberately intending to cause the person harm (Freeman, 2016). In the example shown in Figure 1.7, the person believed they were going to be abducted by the Freemasons. Each petal on the vicious flower represents one of the six key processes within the model, thus proposing that these factors – or at least one of them – will almost always play a role in the maintenance of persecutory delusions³ (Freeman, 2016). The vicious flower formulation also highlights bidirectional relationships between each process and the persecutory belief (Freeman, 2016). These bidirectional relationships are represented as arrows on the six individual petals of the vicious flower formulation diagram as seen in Figure 1.7, below. As Freeman explained: “*worry brings paranoid thoughts to mind, which, in turn, leads to greater levels of worry*” (Freeman, 2016, p. 688).

³ For illustrative purposes the formulation diagram in Figure 1.7, indicates that the person is experiencing difficulties in all six of these domains.

The first petal on the vicious flower, identifies anxiety-laden cognitions associated with anticipatory threat (Freeman, 2016). For example, a person may have worrying thoughts such as *'the Freemasons are coming to take me'*, which may be further reinforced by worrying images of threat (Schulze et al., 2013; e.g. the person brings to mind an image of big, burly men abducting them - bundling them into the boot of their car and driving off with them in broad daylight).

The second petal identifies negative self-beliefs. Indeed, adverse interpersonal experiences may reduce a person's self-confidence, leading to negative self-beliefs (Freeman, 2016; e.g. *'I am an easy target'*. *'I am vulnerable'*) and as Freeman explained: *"paranoia feeds on vulnerability"* (Freeman, 2016, p. 687).

The third petal identifies the presence of internal or external anomalous experiences (e.g. voice hearing) in the maintenance of paranoid ideation (Freeman, 2016). Odd sensations and/or unusual perceptions associated with cannabis misuse for example, may be implicated in enabling paranoia to thrive (Freeman, Dunn, Murray et al., 2015).

The fourth petal relates to poor sleep (e.g. insomnia, nightmares), with sleep dysfunction often increasing negative emotions and emotional dysregulation, which helps to maintain paranoia (Lincoln et al., 2015; Marwaha et al., 2014; Westermann & Lincoln, 2011). Disturbed sleep also limits a person's cognitive ability to be able to revise any misinterpretations of ambiguous situations (Freeman, 2016).

The fifth petal - reasoning biases (e.g. JTC reasoning bias) reduces data-gathering and prevents consideration of alternative explanations (Dudley et al., 2016). This then *"locks patients into the delusional explanation of events"* (Freeman, 2016, p. 688).

Finally, the sixth petal on the vicious flower formulation, suggests that almost all SUs attempt to reduce perceived threat by employing the use of safety behaviours (e.g. avoidance; Freeman et al., 2007). Subtle within-situation safety behaviours may also be present (e.g. the person might scan the room to look for potential escape routes; Freeman, 2016). Mistakenly, SUs often attribute the absence of harm to their use of safety behaviours, rather than the inaccuracy of the perceived threat (Freeman, 2016). Owing to each of these six reasons, persecutory delusions continue to be maintained (Freeman, 2016).

Figure 1.7 An example of a vicious flower maintenance formulation, based on Freeman's cognitive model of persecutory delusions



1.9 Conclusion

Definitions of the term CF have evolved since it was first defined by V. Meyer and Turkat in 1979. Whilst these definitions vary, there appear to be some commonality between the definitions, with CF seen as a process that: a) utilises an evidence-based psychological theory, framework, or model, b) attempts to understand a person's psychological, interpersonal, and/or behavioural difficulties, c) generates a hypothesis, and d) directs treatment for change.

V. Meyer is credited as the founding father of CF (Turkat, 2015), and the process appears to have originated in the early 1950's from the Maudsley Hospital in London (Bruch, 2015). This coincided with the establishment of the profession of clinical psychology, setting forth the birth of the 'scientist-practitioner' method. Therein began a growing quest to understand and treat psychological difficulties at a scientific and idiosyncratic level (Turkat, 2014, 2015).

CF represented a paradigm shift, that moved away from a taxonomic classification of mental health difficulties owing to a growing discontent with psychiatric diagnoses (Bruch, 2015). Indeed, critics have become increasingly vocal over recent decades (Bentall, 2009; Johnstone, 2014a, 2018, 2022). A contentious and long-standing debate is whether CFs should be used as an adjunct, or as an alternative, to psychiatric diagnoses (Aftab, 2020; Carey & Pilgrim, 2010; Johnstone, 2014a, 2018; Macneil et al., 2012). The PTMF for example, has been proposed as an alternative non-medicalised approach to conceptualising mental distress (Boyle & Johnstone, 2020; Johnstone, 2021; Johnstone et al., 2018). Similarly, the five Ps framework (Dudley & Kuyken, 2014) is increasingly used in the NHS as an adjunct (or alternative) to psychiatric diagnosis.

Formulation is a widely utilised process across all evidence-based psychological approaches (Jenkins, 2020). However, a review of different psychological approaches within this chapter, has emphasised marked differences in the ways in which the *content* and *process* of a CF is developed, shared, and applied (depending upon the psychological approach, theoretical perspective, and/or model that is used; Johnstone & Dallos, 2014). For example, CBT and psychodynamic approaches differ, regarding the extent to which they make the CF process transparent and explicit, and the extent to which they directly involve the SU as a collaborative co-pilot. Moreover, even within the same psychological approach (e.g. CBT), formulation doesn't necessarily mean one thing. There are multiple models/frameworks that could be drawn upon when conceptualising psychosis, that might all work, but perhaps to varying degrees. In CBTp, the range and complexity of models can be challenging for trainees to get to grips with and may cause confusion in terms of practice (Newman-Taylor et al., 2022).

Indeed, there has been an evolution of models and frameworks, from the ABC model (which was the earliest generic model; Ellis, 1957), to the five Ps framework (a generic framework, not based on established empirical evidence; Dudley & Kuyken, 2014), through to something much more sophisticated, such as the 'feeling safe' model which focuses on specific causal mechanisms in the 'here and now' (Freeman, 2016). In CBTp, a therapist may build upon different levels of formulation - moving up the levels to a generic longitudinal model (if clinically indicated, e.g. Beck et al., 1979). Alternatively, they may use a highly specialist disorder-specific model, that is much more robust in terms of evidence-base. However, for therapists to draw upon an

evidenced-based disorder-specific model, from which to base their CF, then an understanding of psychiatric diagnosis seems imperative.

In sum, even within a widely utilised, evidence-based approach such as CBTp there is no consistent agreement on the model to be used as the basis for the formulation. This leaves scope for a diverse range of formulations to draw upon, each with their own unique focus.

1.10 Thesis aims

The overarching aim of this thesis is to employ the use of mixed methods to further our understanding of CF, particularly in relation to CBT for the treatment of psychosis. The following specific aims for each chapter in this thesis are as follows:

- 1) The aim of Chapter 2 is to undertake a systematic review of the literature on CF. The focus of the review will be to investigate the reliability of CFs in relation to various psychological disorders, across a wide range of theoretical modalities.
- 2) The aim of the first empirical study in Chapter 3, is to employ the use of the Delphi method to investigate whether a panel of international CBTp experts can reach consensus on the essential ingredients of a CBT CF for auditory hallucinations, and persecutory delusions, in schizophrenia spectrum disorders.
- 3) The aim of the second empirical study in Chapter 4, is to employ the use of causal inference methods to examine whether a range of therapeutic process variables (e.g. maintenance CF, longitudinal CF) modified treatment effects, as measured by the primary outcome measure (the Positive and Negative Syndrome Scale; PANSS). This aim relates to a secondary analysis of a RCT that offered CBT to people with schizophrenia spectrum disorders, that chose not to take antipsychotic medication. A secondary aim of this study is to estimate the treatment effects of each additional session of therapy, on the PANSS total score.
- 4) The aim of the third empirical study in Chapter 5, is to employ the use of reflexive thematic analysis to explore the personal impact of CBT CF for SUs experiencing psychosis in the early stages. A secondary aim of this qualitative study is to explore the personal impact of different levels of CF.

- 5) Finally, the aim of Chapter 6 is to provide an overarching discussion of the findings in relation to this thesis.

Chapter 2. Are case formulations reliable? A systematic literature review with a narrative synthesis

“Two (or more) heads are better than one”.

– American Psychological Association (2014, para. 1).

2.1 Abstract

A ‘reliable’ CF is viewed (by some) as essential for the competent practice of evidence-based psychotherapy. Whilst several studies have investigated CF reliability, only three systematic literature reviews have been undertaken in the context of mental health. Flinn et al.’s (2015) systematic review focused solely on the reliability of CFs, across a range of psychological disorders, and theoretical modalities. Flinn et al. conducted their searches in 2014, meaning, it was timely to update and extend their findings. For the current review, a systematic search of four databases (plus Google Scholar) was undertaken in addition to citation searching, and reference list trawling. This yielded a total of 32 eligible articles. An assessment of methodological quality suggested that the quality of individual studies was not associated with CF reliability. Results overall, indicated that five different reliability methods had been used to assess CF reliability. Coefficient values ranged from ‘weak to perfect’ (.10 to 1.0), and ‘poor to excellent’ (<.40 to 1.0) both within, and across studies, in relation to individual formulation components. Owing to the disparate set of literature a meta-analysis was contraindicated. A narrative synthesis suggested that atheoretical formulations (e.g. four Ps) produced higher levels of CF reliability, than the other theoretical modalities. Another key finding provided a signal that as the number of raters/judges increased (e.g. from one to five), reliability levels also increased. This adds further weight to the importance of clinical supervision, and the potential benefits of team formulation. Having two (or more) ‘heads’ involved in CF, may help to counteract problematic heuristics that can affect the clinical decision making of a single clinician.

2.2 Introduction

Chapter 1 described the role of formulation in terms of the scientist-practitioner framework, its centrality within clinical practice, and its role as an alternative to diagnosis. Hence, CF is afforded a very high status, with, for example, clinicians and researchers ‘crowning’ it “*the heart of evidence-based practice*” (Bieling & Kuyken, 2003, p. 53). Yet, with limited scientific underpinnings for these claims, CF risks being dethroned (Bieling & Kuyken, 2003; Kuyken, 2006).

Psychological therapies such as CBT are empirically supported in the treatment of a wide range of psychological disorders (Easden & Fletcher, 2020). Despite this, little attention has been paid to the empirical foundations of formulation (Aston, 2009; Kuyken et al., 2005). Hence, there is a disparity between the psychological therapy evidence-base, and literature supporting the value of CF as providing a pivotal role. Owing to these tensions it is argued that CF should be open to scientific examination. Indeed, concrete evidence is needed to support its status (Flinn et al., 2015).

Whilst Chapter 4 of this thesis investigates whether CF plays a central role in treatment outcome, another area of scientific examination concerns the notion of CF reliability. This often refers to whether clinicians can: a) independently construct similar formulations based on the same set of clinical material (i.e. assuming they are using the same model, can therapists agree with the conceptualisation of a given case?), or b) independently construct formulations that match a gold standard ‘benchmark’ formulation, constructed by a single expert (or team of experts; Bieling & Kuyken, 2003; Eells, 2022; Kuyken, 2006; Mumma, 2011). The degree of consistency/extent of agreement that exists between the clinicians (as determined by raters) has been referred to as “interrater”, “inter-judge” (or more aptly termed “inter-clinician” or “inter-formulator” reliability). The degree of discrepancy represents the variance, and the degree of unreliability (Bartko, 1991; DeVellis, 2005).

It is argued that if psychotherapy aspires to be scientific, there should be consistency with which clinicians: 1) make sense of a SU’s problems, and 2) suggest ways of intervening (Eells, 2022). Indeed: “*it is reasonable to suppose that two practitioners asked to make inferences about the same case using the same theory and the same case formulation framework should construct similar formulations*” (Kuyken (2006, p. 21). Formulations, therefore, should not be entirely subjective. They should be

grounded in empirical data to ensure a high degree of replicability and consistency when evaluated (Flinn et al., 2015; McHugh, 2012). As such, there is an uncontested dictum (i.e. law) that a ‘reliable’ CF is essential for the competent practice of psychotherapy (Beck, 2020; Bieling & Kuyken, 2003; Eells, 2022; Kuyken et al., 2009).

One of the earliest investigations into CF reliability was conducted by Seitz (1966) who highlighted: “*the consensus problem in psychoanalytic (and psychotherapeutic) research*” (p. 209). Seitz reported that satisfactory consensus in relation to CF was minimal, with psychoanalysts unable to agree on the ‘focal conflict’ (i.e. what was of central importance; Seitz, 1966). Why might it be difficult for highly experienced and skilled clinicians (e.g. a group of psychologists) to have difficulty achieving formulation consensus? Seitz (1966) suggested this may have something to do with “*the inadequacy of our interpretive methods*” (p. 214). For example, the analysts in Seitz’s study made complex inferences at an overly deep level, without systematically checking that their interpretations aligned with the case material (Seitz, 1966). A second reason may have been owing to analysts focusing on different aspects of the case, hence: “*difficulties in reaching consensus are due, not so much to the various interpreters being all wrong, but to the fact that each of us tends to be only partly right*” (Seitz, 1966, p. 215). This led Seitz to conclude that CF reliability research is “*in need of consistent and systematic (...) methods that can make consensus possible*” (Seitz, 1966, p. 210).

Following the publication of Seitz’s (1966) paper, several researchers sought to improve the reliability of CFs using a variety of methods (Eells, 2022). The first to successfully achieve this was Luborsky (1977) using his Core Conflictual Relationship Theme (CCRT) method. Since then, more than 15 structured methods of CF have been proposed (Luborsky et al., 1993). Examples include: the Core Conflictual Relationship Theme Leipzig/Ulm (CCRT-LU) method (Albani et al., 2002), the Role-Relationship Model Configuration (RRMC) method (Horowitz, 1989, 1991b), the Plan Formulation Method (PFM; Curtis & Silberschatz, 1989), and the Case Formulation Content Coding Method (CFCCM; Eells et al., 1998, 2005). Whilst the majority of these have been utilised within a psychodynamic framework, methods from cognitive, cognitive behavioural, and integrative schools of thought have also been developed (Eells, 2022).

Nevertheless, only a small number of literature reviews have investigated the reliability of CF (e.g. see reviews by Aston, 2009; Bieling & Kuyken, 2003). These

reviews reported on CF reliability alongside other research questions, such as the validity of CF, and its relationship to outcome. Moreover, they were not *systematic*. To date, only three systematic reviews of CF have been undertaken in the context of mental health (see Easden & Kazantzis, 2018; Flinn et al., 2015; Rainforth & Laurenson, 2014). A recent systematic review also examined the reliability of forensic CFs (see Wheable & Davies, 2020). However, the predominant use of formulation in a forensic setting is focused specifically on understanding offending behaviour and risk, and the reliability of forensic CFs is beyond the scope of this thesis chapter.

The systematic review by Rainforth and Laurenson (2014) explored CF reliability, alongside efficacy and validity. However, their review was not restricted to peer reviewed articles, nor did their search terms contain the word ‘reliability’. Their search was restricted from 1999 to 2011, and this yielded 39 articles which were analysed using content analysis. As such, ‘reliability and validity’ was a theme that emerged from the CF literature (Rainforth & Laurenson, 2014). However, their findings (in relation to reliability) were not systematically reported.

The systematic review conducted by Easden and Kazantzis (2018) focused on the reliability and validity of CF, specifically in relation to CBT. No restriction was placed on the start date of their review, and the end date - February 2016, was the date they carried out their searches. Findings indicated that the level of reliability across 24 studies varied widely, from almost no agreement to 100% agreement (Easden & Kazantzis, 2018). Low levels of agreement were found in relation to underlying cognitive mechanisms (e.g. core beliefs about self, world, others, and dysfunctional attitudes, i.e. mean intraclass correlation coefficient [ICC] was .46, and the range was .07 to .70; Easden & Kazantzis, 2018). They also found that some therapists could agree on the content of a conceptualisation (or produce content that was similar to an expert), but the results were variable (Easden & Kazantzis, 2018). For example, high levels of agreement were found in relation to CF content (i.e. mean ICC was .83, and the range was .66 to .92), when averaged across five therapists (see Persons & Bertagnolli, 1999).

Flinn et al. (2015) were the first to conduct a systematic review of peer-reviewed articles that focused solely on investigating the reliability of CFs across a range of psychological approaches. No restriction was placed on the start date of their review, and the end date - April 2014, was the date they carried out their searches. In total, 18 studies were identified, with most studies focused on the extent to which clinicians

agreed with one other, or with an expert ‘benchmark’ formulation (e.g. Dudley et al., 2010; Kuyken et al., 2005). Interrater reliability was defined as the rate of consistency between clinicians on certain aspects of a case, with reliability estimates ranging from slight (.10 to .40) to substantial (.81 to 1.0; Flinn et al., 2015). In addition, one study also investigated the test-retest reliability of CFs (Collins & Messer, 1991), and this study demonstrated support for the stability of formulations over a 3-month period (Flinn et al., 2015).

Like the findings reported by Easden and Kazantzis (2018), and the non-systematic reviews (e.g. Aston, 2009; Bieling & Kuyken, 2003), Flinn et al. (2015) similarly found that cognitive therapists tended to agree on overt, descriptive aspects of a formulation, for example, SUs’ presenting issues (e.g. studies by Kuyken et al., 2005; Persons et al., 1995; Persons & Bertagnolli, 1999). In contrast, reliability was poorer in relation to theory-driven ‘inferential’ aspects of the case, for example, SUs’ underlying assumptions and core beliefs (e.g. studies by Kuyken et al., 2005; Dudley et al., 2010). Indeed, as cognitive therapists move to more inferential levels of CF, higher level formulation skills seem to be required (Kuyken et al., 2005). Flinn et al. also highlighted that training and greater clinical experience may lead to higher levels of agreement and reliability (e.g. Collins & Messer, 1991; Kuyken et al., 2005), particularly with inferential and theory driven aspects of a case (Flinn et al., 2015).

Non-systematic reviews of the literature have also consistently found that psychodynamic CFs show greater levels of reliability, over cognitive behavioural CFs (Aston, 2009; Bieling & Kuyken, 2003). Flinn et al. concurred with this finding, stating that: “...*psychodynamic formulations appeared to generate somewhat increased levels of reliability than cognitive or behavioural formulations...*” (Flinn et al., 2015, p. 266). This is an interesting finding, given that higher levels of inference would seem to be necessary with a psychodynamic approach to psychotherapy. It has been suggested that one reason for this greater reported reliability, may be owing to the methods used by psychodynamic researchers that: “...*may have inflated reliability, such as using standard categories and pooling the scores of judges*” (Flinn et al., 2015, p. 285).

In summary, the systematic review by Flinn et al. (2015) yielded mixed findings, which led the authors to describe difficulties in drawing definitive conclusions from the literature owing to 1) the wide variety of methods used, and 2) the various ways researchers had measured CF reliability; including the various statistical methods used

to report their findings (e.g. Cohen's kappa, weighted kappa, Person's correlation coefficient).

It is nine years since Flinn et al. reported their findings, and 10 years since their searches were undertaken. Therefore, it seems timely to update and extend the Flinn et al. review, by systematically reviewing the current state of the literature. Whether clinicians can achieve consensus to produce reliable formulations, remains a highly pertinent research question.

2.3 Aims

This systematic literature review aimed to investigate whether CF is a reliable process, and if so, whether higher levels of CF reliability are associated with a particular theoretical modality. A second aim considered whether methodological quality affected CF reliability.

2.4 Method

As discussed above, this review intended to provide an update and extension of Flinn et al., hence the inclusion/exclusion criteria stated here are broadly the same. Additions to the exclusion criteria (listed as the final three bullet points below) were made to tighten the focus of this review, and to keep it aligned with the overall aims of this thesis. The following inclusion/exclusion criteria were decided *a priori*, as per the study protocol.

2.4.1 Inclusion criteria

Studies were eligible if they:

- Examined inter-clinician/inter-formulator, and interrater/inter-judge, and/or the test-retest reliability of CFs. This required reporting the results of a reliability measure (e.g. percentage agreements, ICCs).
- Outlined the theoretical model, or atheoretical framework (e.g. five Ps) of the formulation, as psychological CFs may be based on a variety of different theoretical models.

- Included adult, and/or adolescent (aged 14+ years) mental health formulations.
- Investigated the reliability of CFs developed by any mental health professional, including studies that utilised a combination of clinicians and students.
- Were published in peer-reviewed journal articles, to control for quality.

2.4.2 *Exclusion criteria*

Studies were excluded if they:

- Had formulators recruited entirely from a student population, which would likely reduce the ecological validity of this review⁴.
- Consisted of a review of previous research (with no new research being undertaken).
- Focused on the reliability of assessment measures of formulation, or measures that may serve to inform the process of CF e.g. pre-therapeutic assessments.

In addition to the Flinn et al. criteria, the following studies were excluded if they:

- Included child formulations, as we aimed to tighten the inclusion criteria to reduce heterogeneity across the sample. Also, child formulations have not been the focus of the empirical work undertaken in this thesis.
- Focused on the formulation of physical/psychosomatic health conditions (e.g. chronic fatigue).
- Focused on the use of CF in a forensic setting.

2.4.3 *Search strategy*

The Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) 2020 updated guidance for the conduct of good quality systematic literature reviews

⁴ Formulations undertaken by clinicians-in-training (e.g. trainee clinical psychologists) were not excluded.

(see Page, McKenzie et al., 2021; Page, Moher et al., 2021; Rethlefsen et al., 2021) were followed. This helped to ensure methodological coherence/integrity.

Eligible studies were screened and identified via 1) database searching, 2) citation searching, and 3) reference list trawling. Four⁵ databases (PsycINFO, MEDLINE, CINAHL, and Web of Science) were searched in April 2023, with citation searches carried out between May and June 2023. Email AutoAlert (selective dissemination of information; SDI) searches for articles (and articles citing eligible studies already included in the review) were set up until 30 April 2024.

The databases utilised in this review, were the same as those used by Flinn et al. (2015) for the purposes of consistency. For additional thoroughness, Google Scholar (a web-based academic search engine) was also searched during these time frames. Google Scholar was used to supplement the database searches (limited to the first 300 results returned, as recommended by Haddaway, et al., 2015) owing to its comprehensive citation coverage (Gusenbauer & Haddaway, 2020). All databases (plus Google Scholar) were filtered by date range from April 2014 to the present date (which also applied to the citation searches). This was to avoid duplication of the original work by Flinn et al. (2015) who had already carried out their database searches, up to April 2014.

As recommended by Shamseer et al. (2015), the specific search strategies were developed with a Health Sciences Librarian⁶ who had expertise in systematic review searching. The following key search terms were used, alongside Boolean operators: "formulation" OR "case conceptualisation" OR "case conceptualization" OR "five Ps" OR "5Ps" AND reliability OR agreement OR consensus. The full line-by-line search strategies (i.e. 'as run' in each database/website) are presented in Appendices C to G. As outlined in Appendix C, the key search terms were searched for in the title, abstract, and keywords of papers.

As advised by the Health Sciences Librarian, it was not deemed necessary to include search terms such as "clinical", "psychosocial" or "psychological". Instead, use of the key search terms "formulation", OR "case conceptualisation", OR "case conceptualization", OR "five Ps", OR "5Ps" ensured that all types of papers containing

⁵ At the time of writing this thesis it was not possible to access AMED (the fifth database used by Flinn et al.), as Newcastle University did not subscribe to the database.

⁶ The Health Sciences librarian who provided input for this search strategy was based at Newcastle University, UK.

these words were identified. The benefits of this comprehensive search strategy meant that articles that related to clinical, psychosocial, and psychological formulations/conceptualisations were identified. However, the downside was that it also led to the identification of many ineligible articles (such as papers that related to the formulation of drugs, soil, fruit juices, fluids, chemicals etc.) which needed to be screened by the author (Helen Mary Spencer; H.M.S.).

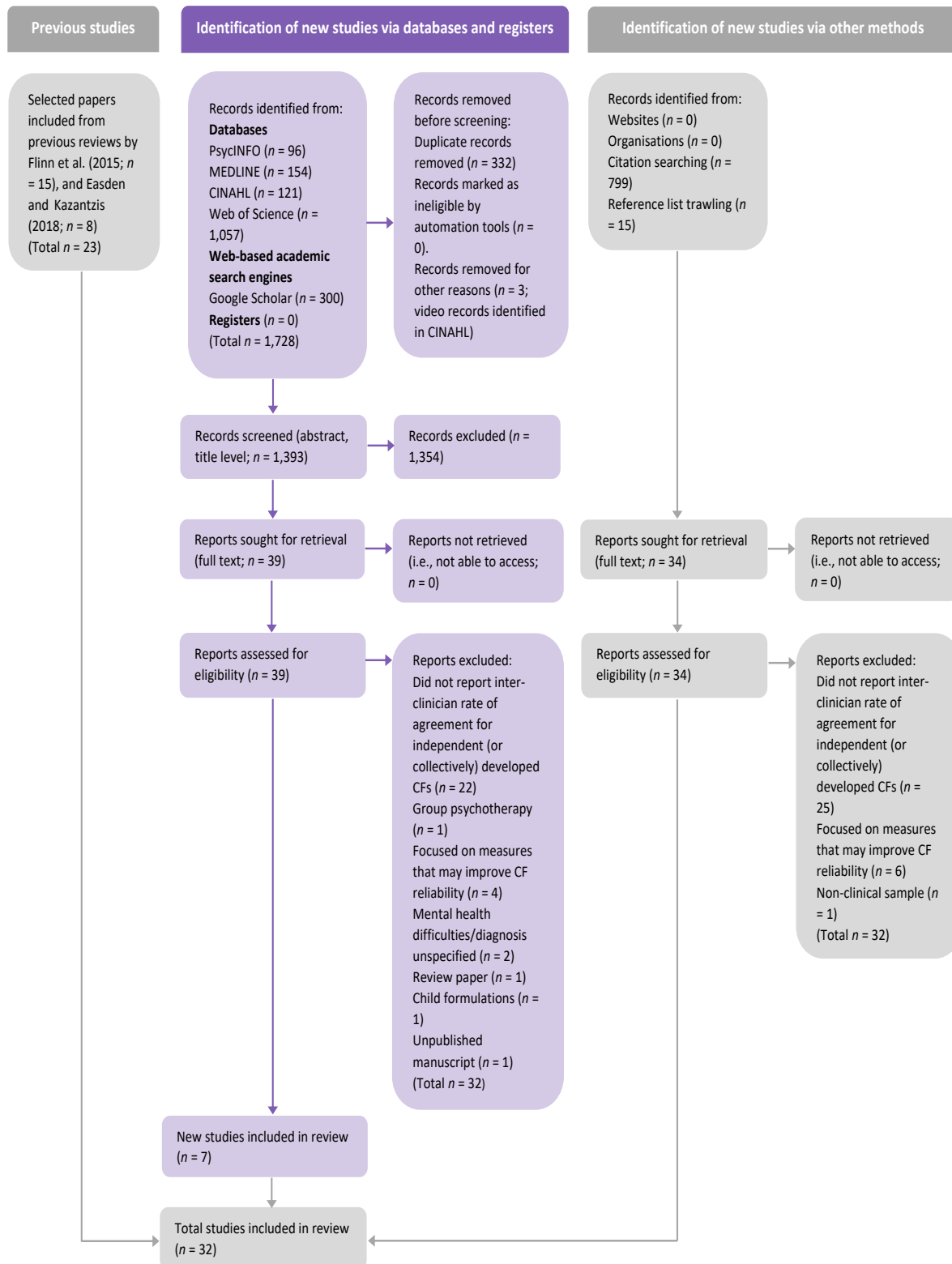
The search terms yielded a total of 1,728 records from all five sources. These records were downloaded into EndNote. Duplicate records ($n = 332$) were removed, prior to screening. If a record appeared relevant (or if its relevance could not be determined after screening the abstract and title), then the full article was sought for retrieval ($n = 39$ articles). After assessing these articles for eligibility (by applying the inclusion and exclusion criteria), seven articles remained. A citation search was conducted using all four databases (plus Google scholar) to check for later research that had cited these papers. This citation search retrieved zero eligible records. Finally, reference list trawling was undertaken. This involved checking the papers listed in the reference lists of the seven eligible papers, as well as trawling the reference lists of a more recent review of CBT formulation research (Easden & Kazantzis, 2018). This resulted in two additional articles being identified (including one which pre-dated 2014, but this paper had not been identified by Flinn et al). An additional eight papers were also identified from the Easden and Kazantzis (2018) review. These eight eligible papers pre-dated 2014 (hence, they were not picked up in the current database searches), yet they had not been identified in the review by Flinn et al.

Flinn et al. (2015) identified 18 eligible articles. However, three of these articles were excluded from the current review as they did not meet our inclusion criteria⁷. Robert Dudley (R.D.) who is an expert in the field of cognitive CF, independently assessed several articles to determine eligibility of papers to be included (or excluded) in the review process (particularly where there was ambiguity surrounding the methods or procedures). A total n of 32 articles were therefore included in this review. Figure 2.1 below, outlines the PRISMA (2020) flow diagram⁸ for this systematic literature search.

⁷ Studies by Popp et al. (1996), and Curtis et al. (1988), did not appear to focus on the formulation of mental health difficulties, and the study by Wilson & Evans (1983) focused on child formulations.

⁸ The flow diagram was based on the template provided by Page et al. (2021a).

Figure 2.1 PRISMA flow diagram



2.5 Assessment of methodological quality

As reported by Flinn et al. (2015), no pre-existing scale could be used to assess the methodological quality of the studies yielded from the CF reliability literature, owing to the wide variability in research designs. As such, Flinn et al. developed their own ‘quality assessment tool’ based on the Critical Appraisal Skills Programme (CASP, 2004), and the Newcastle-Ottawa Scale (Wells et al., 2010). The Flinn et al. quality assessment tool (adopted here, for the current review) is comprised of five questions, rated on a scale of 1-3. Total scores range from 5-15. In addition, a separate question allows for the reporting of ‘other potential sources of bias’, in relation to each study (see Table 2.1, below).

In relation to the ‘reliability measurement’ question, studies scored 3 points if they used an appropriate statistical measure of reliability, across all aspects of their data analysis (see section 2.6.4 below). To obtain a score of 3, studies could also incorporate percentage agreement, but they would also need to report an appropriate statistical measure (for the assessment of the same *method* of reliability; see section 2.6.3 below). A score of 2 was assigned if studies used a statistical measure for one method of reliability, and percentage agreement for another method. A score of 1 was assigned if studies used percentage agreement only.

Given the importance of para/non-verbal language, video recordings scored 3 points, audio recordings scored 2 points, and written transcripts/case vignettes scored 1 point, in relation to the ‘formulation data’ question. A score of 1 point was also assigned to studies that did not provide information regarding the format of their formulation data (Flinn et al., 2015). Several minor amendments were made by H.M.S. to the scoring of the ‘sample representativeness’ question. Using the Flinn et al. quality assessment tool, 1 point was assigned to those studies where: “*participants mainly consisted of students*” (Flinn et al., 2015, p. 280). However, ‘formulators recruited entirely from a student population’ was an exclusion criterion of the current review (see section 2.4.2 above), as well as the Flinn et al. review, therefore this item was not applicable. As such, a score of 1 point was instead given to studies where: “*sample representativeness was reported inadequately*” – which was relevant to several studies (see Table 2.1 below).

Two points were assigned to studies whereby: “*participants consisted of a range of clinicians and students*” (Flinn et al., 2015, p. 280). However, this item was

extended by H.M.S. to also include “trainee clinicians, RAs, and SUs” (see Table 2.1 below).

In line with Flinn et al., articles included in this review were not excluded following the assessment of quality. This was to ensure that all relevant studies were included (Flinn et al., 2015). However, the quality assessment tool assisted with interpretation of the findings.

2.6 Results

2.6.1 *Methodological quality findings*

Methodological quality ratings were conducted by H.M.S. The 32 studies included in this review, yielded ‘total quality’ scores ranging from 6 to 14 points ($M = 11.2$). This reflected a similar ‘total quality’ score range of 7 to 14 points ($M = 10.7$) reported by Flinn et al. Indeed, the lowest total quality score of 6 points was yielded from the current review, from a recently published study (3). Nevertheless, from the current review, four studies also yielded total quality scores of 14 points (1, 6, 7 and 9). This contrasts with only one study (22) having yielded 14 points (as determined by the Flinn et al. scoring in their review, but not found in the quality assessment of their papers here). This suggests perhaps, that the quality of CF reliability studies may be improving. It should be noted that owing to ambiguity in the write-up of some of the studies included in this review, there were discrepancies found between the findings (and quality ratings) of some of the studies reported by Flinn et al. and the findings (and quality ratings) reported here.

Table 2.1 Assessment of methodological quality

Study	Participant demographics	Sample representativeness	Formulation data	Blinding	Reliability measurement	Total score	Other potential sources of bias
<i>Identified within the current review:</i>							
1. Critchfield et al. (2015)	***	**	***	***	***	14	CF ‘templates’ (which imposed some structure) were provided to formulators.
2. Easden & Fletcher (2020)	***	**	***	*	***	12	CF diagrams (which imposed some structure) were provided to formulators.
3. Grandjean et al. (2021)	*	*	**	*	*	6	Formulators developed ‘prototypical’ CFs, which contained ‘default’ assumptions/hypotheses about the cases, based on diagnoses. Plan analysis takes para/non-verbal aspects

							into account, yet only audio recordings were used.
4. Hartley et al. (2016)	**	***	***	*	*	10	CF ‘templates’ (which imposed some structure) were provided to formulators. Only 10 of the 49 formulations were randomly chosen to be rated.
5. Hegarty et al. (2020)	***	***	*	*	***	11	Predefined standard categories for each formulation component were provided.
6. Nattrass et al. (2015)	***	***	**	***	***	14	Ratings were pooled over three judges.
7. Sørbye et al. (2019)	***	***	**	***	***	14	Raters were given two CFs (from a pool of 60 CFs) and asked to decide whether they were from the same SU (matched pair), or from a different SU (mismatched).

8. Tallberg et al. (2020)	***	***	**	*	***	12	Predefined standard categories for each formulation component were provided.
9. Zivov et al. (2013)	**	***	***	***	***	14	One of the authors of the study role-played the SU. A single case was formulated.
<i>Identified by Easden & Kazantzis (2018)</i>							
10. Eells et al. (2005)	***	**	**	***	***	13	Written vignettes of fictional cases were used. Each therapist was only given five minutes to “think aloud” and construct each CF. Ratings were pooled over three judges.
11. Eells et al. (2011)	***	**	**	***	***	13	Written vignettes of fictional cases were used. Each therapist was only given five minutes to “think aloud” and construct each CF. Ratings

								were pooled over three judges.
12. Haarhoff et al. (2011)	***	***	*	*	*	9	Written vignettes of fictional cases were used.	
13. Kendjelic & Eells (2007)	***	**	*	***	***	12	Training session provided for the ‘training group’ clinicians, emphasised the formulation of inferential material. Ratings were pooled over three raters.	
14. Mumma & Mooney (2007)	**	**	***	*	***	11	A single case was formulated. The novice clinician’s two formulation components were combined into a single component, owing to poor discriminant validity.	
15. Muran & Segal (1992)	***	**	*	*	***	10	A single case was formulated. Ratings were pooled over three judges.	

						The SU also participated as a rater.
16. Muran et al. (1998)	**	*	***	*	***	10 Ratings were pooled over four judges. Plausible, but less relevant CFs were provided (i.e. for the purposes of ‘matching’ to the case). SUs also participated as raters.
17. Muran et al. (2001)	*	*	***	*	***	9 Ratings were pooled over four judges. Plausible, but less relevant CFs were provided (i.e. for the purposes of ‘matching’ to the case). SUs also participated as raters.
<i>Identified by Flinn et al. (2015)</i>						
18. Barber et al. (1995)	***	***	*	*	***	11 Predefined standard categories for each

							formulation component were provided.
19. Caston & Martin (1993)	***	***	*	*	***	11	Formulations were developed by a team of clinicians. Ratings were pooled over four judges.
20. Collins & Messer (1991)	***	***	*	***	***	13	Formulations were developed by a team of clinicians. Ratings were pooled over five judges.
21. Crits-Christoph et al. (1988)	***	***	*	***	***	13	Predefined standard categories for each formulation component were provided. Each CF was initially rated on a scale of 1-5 for 'completeness' – only those that rated 2.5 (or more) were selected for scoring in the final analysis. Plausible, but less relevant CFs were provided (i.e. for the

							purposes of ‘matching’ to the case).
22. DeWitt et al. (1983)	***	***	***	*	*	11	Formulations were developed by a team of clinicians.
23. Dudley et al. (2010)	**	**	***	*	*	9	An actor role-played the SU.
24. Eells et al. (1995)	***	**	*	*	***	10	Formulations were developed by a team of clinicians.
25. Kuyken et al. (2005)	**	**	***	*	**	10	An actor role-played the SU. A single case was formulated. Notable differences were reported in relation to the content of the three training workshops. CF diagrams (which imposed some structure) were provided to formulators.
26. Mumma & Smith (2001)	***	**	***	*	***	12	Ratings were pooled over four judges.
27. Muran et al. (1994)	***	**	**	*	***	11	Ratings were pooled over three judges. Plausible, but

								less relevant CFs were provided (i.e. for the purposes of ‘matching’ to the case). SUs also participated as raters.
28. Perry et al. (1989)	***	**	***	***	*	12	Formulations were developed by a team of clinicians. Plausible, but less relevant CFs were provided (i.e. for the purposes of ‘matching’ to the case).	
29. Persons & Bertagnolli (1999)	**	**	**	*	**	9	A predefined list of overt presenting problems was provided. Ratings were pooled over five judges.	
30. Persons et al. (1995)	***	**	**	*	**	10	A predefined list of underlying cognitive mechanisms was provided. Data from the pilot (training) case was included in the final analysis.	

31. Rosenberg et al. (1986)	**	***	*	*	***	10	Formulations were developed by a team of clinicians. Plausible, but less relevant CFs were provided (i.e. for the purposes of ‘matching’ to the case). Ratings were pooled over four to five judges.
32. Shefler & Tishby (1998)	***	***	*	***	***	13	Plausible, but less relevant CFs were provided (i.e. for the purposes of ‘matching’ to the case).

Note. (1) **Participant demographics**⁹: ***participant demographics were reported clearly, **participant demographics were reported partially, *participant demographics were reported inadequately. (2) **Sample representativeness**: ***participants consisted of clinicians, **participants consisted of a range of clinicians, trainee clinicians, RAs, SUs, and students, *sample representativeness was reported inadequately. (3) **Formulation data**: ***the study used video recordings, **the study used audio recordings, *the study used written transcripts, written case vignettes, or did not define the formulation data. (4) **Blinding**: ***the study reported adequate blinding, **the study reported partial blinding, *there was no evidence of blinding. (5) **Reliability measurement**: ***the study used a statistical measure

⁹ Participant demographic information relates to the formulators/raters.

of reliability, **the study used a statistical measure of reliability in addition to percentage agreement, *the study used percentage agreement (or percentage agreement could be calculated from the study data).

2.6.2 Data extraction

The reliability data were extracted from each of the 32 studies and these data are highlighted in Tables 2.2 to 2.4. Table 2.2 details the data extracted in studies found in the current review. Table 2.3 details the data extracted from eligible studies found by Easden and Kazantzis (2018). Table 2.4 details the data extracted from eligible studies found by Flinn et al. (2015).

Reporting CF reliability was an outcome, but not always the main objective/primary focus for some of the studies included in this review. Just over half of the studies ($n = 18$) specifically aimed to focus on the reliability of CFs (with the most recent study published by Tallberg et al. in 2020). For the remaining $n = 14$ studies, CF reliability was reported as part of a wider study objective. For example, to investigate the relationship between CF and therapy outcomes (see Easden & Fletcher, 2020) or to investigate therapist reasoning in the construction of CFs (see Eells et al., 2011).

2.6.3 Methods used to assess CF reliability

Five different methods were used across the studies included in this review, to assess CF reliability. Most studies ($n = 24$) used one method, however $n = 8$ studies used two methods (7, 16, 20, 21, 24, 25, 29 and 30). The five methods are outlined below in Table 2.5 to indicate the total number of studies (and study numbers) that employed the use of each method across all 32 studies included in this review. Method 1 referred to whether clinicians can independently construct similar formulations based on the same set of clinical material. This was the most prevalent method used by a total n of 19 studies. Method 4 referred to whether clinicians can formulate the same set of clinical material in a similar way, at two different time points, referred to as test-retest reliability. This was the least common method used by 2 studies. For each study, the methods used to assess CF reliability are also outlined below in Tables 2.6 and 2.7, along with the interpretation of their reliability scores.

Using method 3, $n = 10$ studies (7, 15, 16, 17, 21, 24, 27, 28, 31 and 32) required formulators to combine their formulations (or formulation components) designed to be clinically representative of the case, with other formulations (or formulation

components) designed to be less relevant (i.e. purposefully mismatched) to the case. These formulations/components were then rated by separate participants. As suggested by Flinn et al. (2015), this method of ‘matching’ CFs (or their components) to the case could potentially be seen as a source of bias, given that the alternative formulation/components may be obviously seen as less relevant, therefore, easier to rate. In addition, $n = 4$ of these studies (15, 16, 17 and 27) were designed whereby the SU (whose difficulties were formulated) participated as one of the raters. Consequently, SUs were asked to rate how relevant they thought some of the formulation components were, in relation to themselves. Whilst consultation with SUs is useful to enhance CF reliability and collaboration in the real world (Persons & Bertagnolli, 1999), this is likely to have inflated reliability within these studies. Indeed, it may be hypothesised that this is the reason for their high rates of reliability (i.e. studies 15, 16, and 27 yielded ‘excellent’ rates of reliability, and study 17 yielded ‘fair to excellent’ rates of reliability; see Table 2.6 below).

Table 2.2

Study details: An update on Flinn et al. (2015) via database searching, citation searching, and reference list trawling ($n = 9$ studies)

Author(s), location, and area of CF studied	Participant characteristics	Method	Key findings
1. Critchfield et al. (2015) USA	• <i>SU/diagnosis formulated:</i> The population was described as CORDS (Comorbid, Often Re-hospitalized, Dysfunctional, and Suicidal). E.g. ‘treatment resistant’ inpatients diagnosed with major depression, generalized anxiety disorder, posttraumatic stress disorder ($n = 20$). • <i>Formulators:</i> Two groups: 1) undergraduate research assistants (RAs; $n = 7$), and 2) graduate-level	• <i>Materials:</i> 20 videos (90 mins each) of CF interviews conducted by Dr. Benjamin (the developer of IRT, and co-author of this research). Research templates. • <i>Theoretical modality:</i> Interpersonal reconstructive. • <i>Training:</i> Undergraduate RAs were provided with approximately four weeks training. Graduate-level IRT therapy trainees were provided with “months-long” formulation and intervention training with an expert IRT clinician. • <i>Procedure:</i> Formulators watched the 20 videos independently, then used a CF ‘research template’ to record their observations. • Interpersonal content was then rated based on the Structural Analysis of Social Behaviour (SASB; Benjamin, 1974, 1996).	• Cohen’s kappa coefficients concerning the rate of agreement between raters (in relation to the assessment of the four key elements of CF) ranged from .64 to .82 for the undergraduate RA formulations, and from .75 to .90 for the graduate-level IRT therapy trainee formulations. • Mean kappa coefficients concerning the rate of agreement for both the undergraduate RAs <i>and</i> the graduate-level IRT therapy trainees ranged from .71 to .88. • ‘Substantial’ to ‘almost perfect’ rates of agreement between the raters indicated that the IRT CF method

	Interpersonal Reconstructive Therapy (IRT) trainees ($n = 13$). • <i>Raters (referred to as judges)</i>: Graduate students and professionals receiving IRT training ($n = 33$), as well as undergraduate RAs ($n = 7$).	• Blind raters (who had no prior knowledge of the cases) independently scored the CF templates in relation to four key elements of the IRT CF method: 1) key figures, 2) copy processes, 3) key figures link to specific symptoms, 4) key figures link to a specific symptom, by a specific copy process.	was reliable amongst formulators with a range of clinical experience. • Formulators showed high levels of agreement about the presence of central components in the IRT CF, linking symptoms to current attachment relationships, and to relationships with early attachment figures.
2. Easden & Fletcher (2020) New Zealand & Australia	• <i>SU/diagnosis formulated</i>: SUs diagnosed with a first episode of major depressive disorder (MDD; $n = 28$). • <i>Formulators</i>: Intern psychologists in their penultimate or final year of training towards registration as a clinical psychologist ($n = 7$). • <i>Raters</i>: Postgraduate psychology students with a	• <i>Materials</i>: The Conceptualisation Rating Scale (CRS). J. Beck Case Conceptualisation Diagrams (CCDs; Beck, 1995). 255 individual video recordings of live therapy sessions. • <i>Theoretical modality</i>: Cognitive behavioural. • <i>Training</i>: Prior to this study, therapists had undertaken two postgraduate training courses in the theory and practice of CBT for depression. Therapists' training as part of this study involved the use of role-plays etc. and was delivered over five days. Raters received up to 13 hours training and orientation to the use of the CRS which was developed for this study.	• Cohen's kappa coefficients were used for dichotomous data sub-domains (e.g. integration, and importance), whereas ICCs were used for continuous data sub-domains (e.g. therapist competence, and fit/match). • Total score interrater reliability estimates for the integration sub-domain were excellent ($\kappa = .83$). This indicated that raters were able to agree whether therapists explicitly

minimum of one to two years postgraduate experience ($n = 6$).	Training was provided over three days (delivered by the first author, and a senior clinical psychologist). • <i>Procedure</i>: 28 SUs were recruited and offered up to 20 CBT sessions (plus 2 booster sessions) provided by the therapists. • Therapists wrote up CCDs in-between the therapy sessions. • Therapy sessions one to 10 were video recorded and rated by pairs of independent raters using the CRS. • The CRS consists of 12-items, each with four sub-domains: 1) integration (i.e. therapist selecting a situation and enquiring about related automatic thoughts, emotions), 2) importance, 3) therapist competence, 4) fit/match of the in-session use of formulation with the written CCD.	discussed different components of CF in-session, irrespective of the level of competence in doing so. • Total score interrater reliability estimates for the importance sub-domain were adequate ($\kappa = .65$). This indicated that raters were less able to agree on the aspects of CF that should be prioritised for discussion, and conversely when aspects of the CF were less relevant to the session (e.g. explicit identification of negative core beliefs is generally inadvisable in the first session of therapy). • Total score interrater reliability estimates for the therapist competence sub-domain was excellent ($ICC = .93$). • Total score interrater reliability estimates for the fit/match sub-
--	---	--

domain was also excellent (ICC = .86). This indicated that raters were able to agree when information relevant to the CF was captured in a written format (i.e. the CCD).

- In summary, raters (considered to be novices, in terms of their level of experience) were able to reach high levels of agreement concerning the content of therapist CFs, using the CRS.

3. Grandjean et al. (2021) Germany	• <i>SU/diagnosis formulated:</i> SUs diagnosed with narcissistic personality disorder (NPD) from an outpatient clinic ($n = 14$). Comorbidities included major depression ($n = 6$), substance abuse disorder ($n = 3$), somatoform disorder ($n = 2$), and histrionic personality	• <i>Materials:</i> Audio recordings of ($n = 14$) psychotherapy sessions with SUs diagnosed with NPD. • <i>Theoretical modality:</i> Plan Analysis (atheoretical). • <i>Training:</i> No training was provided as part of the study. • <i>Procedure:</i> Based on the audio recordings, the two formulators/raters independently developed 14 individual ‘plan structures’, which were formulated and drawn onto paper.	• Interrater reliability was randomly carried out on 14% of the data/plan structures ($n = 2$ cases). • Sufficient interrater reliability was achieved at 60.5% (calculated using Benkert's method; Benkert, 1997).
---------------------------------------	---	--	--

	disorder (PD; $n = 2$). One SU had no comorbidities. • <i>Formulators/raters</i>: Total $n = 2$; no demographic information provided.	• ‘Plan analysis’ (Casper, 2019) incorporates the conceptualisation of thoughts, beliefs, and emotions by taking verbal and para/non-verbal aspects into account. It is compatible with most therapeutic approaches. • A prototypical plan structure of a CF refers to a framework outlining what is frequently observed among SUs presenting with a particular diagnosis or clinical problem, and can therefore serve as ‘default’ assumptions/hypotheses. • The plan structure provides a basis/starting point from which individualised information can then be added. • Plans were scored based on ‘matching criteria’ between the two formulators/raters. A maximum of five points indicated that the formulators/raters had complete agreement.	
4. Hartley et al. (2016) UK	• <i>SU/diagnosis formulated</i>: An inpatient diagnosed with psychosis ($n = 1$).	• <i>Materials</i>: A CF template used in previous research (Dudley et al., 2010) that included predisposing factors, precipitants (or triggers), and maintaining factors. Case vignette information about a fictional SU	• The total mean CF score (i.e. the average number of ‘hits’ in relation to the expert benchmark formulation) was 31.3 (21.7%). This

• <i>Formulators</i>: Clinical staff ($n = 50$) with experience of psychosis, working in acute or long-stay mental health inpatient units (e.g. psychologists, mental health nurses, occupational therapists). • <i>Raters</i>: Co-authors of the study ($n = 2$).	diagnosed with psychosis. A 17-minute video outlining the SU's current presentation. A CF scoring manual adapted from Dudley et al. (2010). • <i>Theoretical modality</i>: We can infer from the CF template that the four Ps framework was used (atheoretical). • <i>Training</i>: No training was provided as part of the study. • <i>Procedure</i>: Clinicians read the fictional case vignette information, watched the video, and then completed the CF template. • The formulation templates were rated (using the scoring manual adapted from Dudley et al., 2010) against a 'benchmark' formulation developed by an expert panel of four clinical psychologists. • Each item identified on the clinician's formulation template, that matched an item present on the expert benchmark formulation, led to a score assigned to the participant on a scale of 0–2. Two points were assigned if the clinician identified the same theme as the expert's formulation. One point was assigned if the	indicated that clinicians had a low rate of agreement with the expert formulation.
---	--	---

		clinician correctly identified an item that was thematically relevant to that identified by the experts but was vague, and a score of 0 was assigned if the clinician did not mention an item that was in the expert formulation. • A total CF score was calculated for each clinician by calculating the total number of their ‘hits’ within a scoring range of 0-144. • Of the 49 formulation templates, 10 were randomly selected from the data set and scored independently by the raters, in relation to the expert formulation.	
5. Hegarty et al. (2020) Australia	• <i>SU/diagnosis formulated:</i> SUs diagnosed with major depression and comorbid PD ($n = 20$). • <i>Therapists:</i> Doctoral-level clinical psychologists ($n = 10$) trained in the time-limited supportive-expressive psychodynamic psychotherapy model, for the	• <i>Materials:</i> Verbatim transcripts of the third therapy session for each SU. The CCRT-LU method (Albani et al., 2002). • <i>Theoretical modality:</i> Psychodynamic. • <i>Training:</i> No training was provided as part of the study. • <i>Procedure:</i> SUs were offered 16 sessions of manualised time-limited supportive-expressive psychodynamic psychotherapy from the therapists (Luborsky et al., 1995).	• Interrater reliability was ‘fair’ for W ($\kappa = .55$) and RS ($\kappa = .57$), and ‘good’ for RO ($\kappa = .61$).

treatment of depression and PDs.

• *Formulators/Raters*
(referred to as judges):

Doctoral-level academic clinical psychologists ($n = 2$; authors of the study), with a psychodynamic research background.

• The CCRT-LU method was used (Albani et al., 2002). This provided a reformulation of the category structure of the CCRT method (Luborsky & Crits-Christoph, 1998).

• Judges independently read the verbatim therapy session transcripts for each SU, then they identified and scored $n = 728$ components, selected from lists of standardised categories (as described by Luborsky & Crits-Christoph, 1990).

• The categories were then converted to CCRT-LU base-level categories (Albani et al., 2002), then organised into 13 clusters based on the SUs' relationship patterns; specifically, their 1) wishes (W), 2) responses of others (RO), and 3) responses of self (RS).

• The “conflictual” aspect of the CCRT refers to the conflict between what the SU wishes for (W), and what they experience from others and oneself e.g. a wish to be admired, but perceives the other as rejecting (RO), which results in them experiencing low mood (RS).

• Interrater reliability for scoring of the CCRT-LU

		components was assessed at the cluster level using Cohen's kappa.	
6. Nattrass et al. (2015) UK	• <i>SU/diagnosis formulated:</i> SUs diagnosed with obsessive compulsive disorder (OCD; $n = 29$). • <i>Formulators:</i> British Association for Behavioural and Cognitive Psychotherapies (BABCP) accredited therapists with two to 10 years post-qualification CBT experience ($n = 8$). • <i>Raters:</i> Independent raters ($n = 3$), with prior experience of treating OCD with CBT.	• <i>Materials:</i> Audiotapes ($n = 70$) of two early formulation sessions per SU. The CFCCM (Eells et al., 1998, 2005). • <i>Theoretical modality:</i> Cognitive behavioural. • <i>Training:</i> Therapists received two CBT for OCD refresher workshops and attended a supervision group for one hour every fortnight throughout the duration of the study. Raters attended a three-hour training session on OCD-CBT CF and received two days of CFCCM training. Raters achieved a level of reliability of $>.75$ against the content codes on the example CFs (developed as a 'benchmark', by the research team). • <i>Procedure:</i> Using a validated coding manual (the CFCCM v2) CF content and quality were assessed by the first author from audiotapes of routine CBT sessions during which CFs were "presented" to the SUs during the "formulation phase". • The CFCCM categorises CF content according to three main categories: 1) descriptive, 2) diagnostic, and	• Interrater reliability across every CF obtained a Fleiss' multirater kappa coefficient (Fleiss, 1981; using the formula discussed in Siegel & Castellan, 1988) of .64, suggesting good agreement across the content <i>and</i> quality coding categories. • Interrater reliability for independent ICCs ratings ranged from .73 to .92, across the <i>quality</i> coding categories only.

		3) inferential. The quality coding system has six indices (e.g. coherence) and these were measured on a five-point scale (0 = not present, 1 = rudimentary, 2 = adequate, 3 = good, and 4 = excellent). • Three blind raters coded the same CF session independently and consensus was defined as at least two raters assigning the same code.	
7. Sørbye et al. (2019) Norway	• <i>SU/diagnosis formulated:</i> SUs diagnosed with depressive disorders, anxiety disorders, PDs, and interpersonal problems ($n = 100$). • <i>Formulators:</i> Psychiatrists ($n = 6$), and a clinical psychologist ($n = 1$). Total $n = 7$. All were experienced therapists and had been practising psychodynamic psychotherapy between 10 to 25 years.	• <i>Materials:</i> Audio recordings of semi-structured dynamic interviews ($n =$ not specified). Written narrative formulations: 30 pairs of matched formulations, and 30 pairs of mismatched formulations. The CFCCM (Eells et al., 1998, 2005). • <i>Theoretical modality:</i> Psychodynamic. • <i>Training:</i> Raters were trained by the first author. • <i>Procedure:</i> An independent clinician interviewed each SU in a two-hour semi-structured psychodynamic interview. • A minimum of two formulators listened to the audio recordings and were asked to independently write an unstructured psychodynamic CF (using non-technical language) based on the SU's clinical history,	• Raters were able to reliably agree whether a case was matched or mismatched. Mean ICC = .82. • Greatest agreement was in relation to the overt/descriptive elements of the formulation (ICC = .82), and less agreement related to the inferential elements (ICC = .61). • The first four raters (pairs 1 and 2) were experienced psychodynamic clinicians. The reliability of their ratings was ICC = .79.

	• <i>Raters:</i> Pair 1 ($n = 2$) both were psychiatrists and psychoanalysts with more than 20 years of clinical experience. - Pair 2 ($n = 2$) 1 psychiatrist and 1 specialist in psychology; both experienced clinicians. - Pair 3 ($n = 2$) both resident psychiatrists; clinically inexperienced with little knowledge of psychodynamic psychotherapy. Total $n = 6$.	diagnostic evaluation, and the psychodynamic interview. • 425 CFs were written (an average of 4.2 CFs per SU) and completeness of each CF was examined by the first author using the CFCCM. • Blind raters were given two CFs and they were asked to decide whether the two formulations were from the same SU (matched pair), or from different SUs (mismatched). • The degree of similarity of the CFs were rated on a Likert scale from one to seven (a rating of 7 = complete, or near complete agreement).	• Two raters (pair 3) had no experience in practicing dynamic psychotherapy. The reliability of their ratings was excellent, ICC = .91. • CFs (as written by experienced clinicians, without any specific structure or labelling of statements into components), could be rated reliably by both the experienced and inexperienced clinician raters.
8. Tallberg et al. (2020) Norway	• <i>SU/diagnosis formulated:</i> SUs diagnosed with depressive disorders, anxiety disorders, PDs, and interpersonal problems ($n = 52$).	• <i>Materials:</i> Audio recordings of semi-structured dynamic interviews ($n =$ not specified). The CCRT method. • <i>Theoretical modality:</i> Psychodynamic. • <i>Training:</i> No training was provided as part of the study.	• Results overall, indicated ‘moderate’ interrater reliability among five raters (which were compared two by two), with a CCRT total mean Cohen’s kappa of .41 (range .26 to .60).

	• <i>Formulators/raters:</i> Psychiatrists ($n = 6$) and a clinical psychologist ($n = 1$). Total $n = 7$. All were experienced therapists and had been practising psychodynamic psychotherapy between 10-25 years.	• <i>Procedure:</i> An independent evaluator interviewed each SU in a two-hour semi-structured psychodynamic interview. • SU's were helped to explore meaningful experiences and vignettes in detail and to give examples of interactions with others in their lives. • Using the CCRT method, two raters independently listened to the audio recordings and were asked to score the SU's interpersonal patterns; specifically, their 1) W, 2) RO, and 3) RS.	• Total mean Cohen's kappa varied across all three interpersonal patterns, and the values were as follows: W = .33 (range .26 to .52), RO = .44 (range .29 to .60) and RS = .45 (range .31 to .58).
9. Zivor et al. (2013) UK	• <i>SU/diagnosis formulated:</i> OCD ($n = 1$). • <i>Formulators:</i> A range of clinicians with different mental health professional backgrounds, and a range of CBT experience (e.g. clinical psychologists $n = 36$, CBT therapists $n = 13$, counselling psychologists $n = 12$, and other $n = 22$). Total $n = 83$.	• <i>Materials:</i> A 12-minute role-play video (produced by two of the authors), of a SU clinical assessment. The Formulation Assessment Tool and Rating (FATR) which was developed for this research. The Rating the Quality of Case Formulation for Obsessive-Compulsive Disorder (RQCFO) which was developed based on Fothergill and Kuyken's (2002) Quality of Cognitive Therapy Case Formulation (QCTCF) rating scale. • <i>Theoretical modality:</i> Cognitive Behavioural. • <i>Training:</i> A workshop (facilitated by one of the	• Pearson's correlation coefficient between the two raters (observers) in relation to agreement on the overall quality of clinicians' CFs (as measured by the RQCFO) was reported as .70. • The overall quality of clinicians' CFs fell within the adequate range for the post-workshop group ($M = 3.9$), whereas the pre-workshop group CFs fell within the poor-

• <i>Raters:</i> Clinicians described above ($n = 83$), and observers ($n = 2$); no demographic information provided.	authors) aimed to improve clinicians' formulation skills in CBT for OCD. The observers were also described as having been trained in the RQCFO. • <i>Procedure:</i> The clinicians were randomly assigned to one of two groups whereby their formulation skills were assessed either: 1) pre-workshop, or 2) post-workshop. • The clinicians were required to complete the FATR which consisted of four parts. • Part 1 required clinicians to watch a video of a clinical assessment (either pre- or post-workshop, depending on the group they had been randomised to), then draw a CF based on the video. • Part 2 required clinicians to rank three formulations that varied in specificity (general, anxiety specific, OCD) in relation to the video from: 1 = best formulation, to 3 = worst formulation. • Part 3 required clinicians to rate each of the three formulations in relation to six components. One of the components required clinicians to rate whether the 'best' formulation they had identified, was similar to	quality range ($M = 2.5$), as measured by the two raters (observers) using the RQCFO. • Ranking the three CFs, there was a significant main effect of group, with 94% of the post-workshop group ranking the OCD formulation as the best formulation for the case (rather than the general, and anxiety specific CFs), in comparison to 56% of the pre-workshop group, $F(1,80) = 12.9, p < .001$. • The post-workshop group also rated their own formulation as significantly similar to the OCD formulation, in comparison to the pre-workshop group, $t(72.2) = 3.6, p < .001$. • Overall, these findings suggest that time-limited, low-cost training can
--	---	--

the formulation they had constructed themselves (in part 1).

- Part 4 required blind raters (observers) to rate the quality of clinicians' formulations using the RQCFO. This measure had been designed for the study, based on Fothergill and Kuyken's (2002) scale.

be highly effective in improving formulation competence.

Table 2.3 Study details: Papers identified by Easden & Kazantzis (2018), not included in Flinn et al. (2015; $n = 8$ studies)

Author(s), location, and area of CF studied	Participant characteristics	Method	Key findings
10. Eells et al. (2005) USA	• <i>SU/diagnosis formulated:</i> SUs diagnosed with GAD, MDD, or borderline personality disorder (BPD; total $n = 6$). • <i>Formulators:</i> Three groups: 1) Novice therapists (total $n = 24$) that consisted of clinical psychology graduate-level student/trainee therapists with less than 1,500 hrs of supervised psychotherapy experience, of which $n = 11$ were psychodynamic in orientation, and $n = 13$ were	• <i>Materials:</i> Written case vignettes ($n = 6$) of fictional SUs diagnosed with either GAD, MDD, or BPD that varied in prototypicality. Audio recordings in the voice of Tracy D. Eells (two-minutes in length per vignette). The CFCCM (Eells et al., 1998, 2005). • <i>Theoretical modality:</i> Psychodynamic and cognitive behavioural. • <i>Training:</i> Raters were trained, which involved: 1) reading the coding manual, 2) attending a series of training sessions led by Tracy D. Eells, and 3) practicing on several formulations developed for training purposes. • <i>Procedure:</i> Therapists were provided with the written case vignettes, and a two-minute audio recording for each vignette.	• Reliability for the judges identifying “idea units” was calculated using two formulas proposed by Scott and Hatfield (1985) for this type of data. The conservative formula achieved 87% across all 6 vignettes, with a mean range of 86% to 88%. The less conservative formula achieved mean agreements ranging from 94% to 95%. These results indicated excellent agreement. • Rate of agreement for 10% of the codes for the three-member rater team was calculated. Fleiss’ multirater kappa coefficients (using

cognitive behavioural in orientation. 2) Experienced therapists (total $n = 19$) with 10 or more years of experience practicing as either psychodynamic ($n = 11$), or cognitive behavioural therapists ($n = 8$). 3) Expert therapists (total $n = 22$) that consisted of psychiatrists, or clinical psychologists recognised as national experts on CF (e.g. they had published one or more scientific articles on the subject), of which $n = 11$ were psychodynamic in orientation, and $n = 11$ were cognitive behavioural. Total $n = 65$.	• Each therapist was then given five-minutes (per case vignette) to “think aloud” and construct a CF (psychodynamic, or cognitive behavioural, depending on the therapist’s perspective) for each case. • The CFs (total $n = 390$) were transcribed, segmented into “idea units”, then content coded by independent blind raters that worked in two groups of three using the CFCCM. • Constructing the “idea units” involved judges (working in two teams) segmenting the formulation transcripts into small “idea units” (usually a sentence or less). Consensus was reached on what represented a “complete” idea. • The CFCCM (revised for this study) categorised CF content according to four main categories: 1) descriptive, 2) diagnostic, 3) inferential, and 4) treatment planning. • A code would be accepted if two of the three coders applied it. • CF content was then rated on eight dimensions of quality (adapted from Strupp, 1955, 1958) which	the formula discussed in Siegel & Castellan, 1988) were as follows: ‘descriptive’ $\kappa = .61$ ‘diagnostic’ $\kappa = .81$, ‘inferential’ $\kappa = .62$, and ‘treatment planning’ $\kappa = .69$. These values reflected good to excellent agreement. • Experts produced higher quality CFs than the novice and experienced therapists, regardless of SU diagnosis or theoretical orientation. • Interestingly, the CFs produced by the novice therapists were rated higher for ‘overall quality’, compared to the CFs produced by the experienced therapists.
--	--	--

	• <i>Judges (also referred to as text segmenters)</i>: Graduate students ($n = 4$), and Tracey D. Eells - the first author of this study. Total $n = 5$. • <i>Raters (referred to as coders)</i>: Clinical or counselling psychology graduate-level student/trainee therapists ($n = 6$).	included: 1) comprehensiveness, 2) formulation elaboration, 3) precision of language, 4) complexity, 5) coherence, 6) goodness-of-fit of the formulation to the treatment plan, 7) treatment plan elaboration, and 8) the extent to which the therapist appeared to follow a systematic formulation process across all six vignettes.	
11. Eells et al. (2011) ¹⁰ USA	• The SU/diagnosis formulated, formulators, judges, and rater information were the same as reported in Eells et al. (2005).	• The materials, theoretical modality, training, and procedure was the same as reported in Eells et al. (2005).	• Multirater Fleiss' kappa coefficients for the two teams of judges identifying "idea units" were the same as reported in Eells et al. (2005). • Experts generated more forward reasoning than the novice and experienced therapists, when forward reasoning is measured by

¹⁰ Eells et al. (2011) used the same data set as Eells et al. (2005), but with different research questions and analysis. Owing to this, the reliability data were the same as described in Eells et al. (2005).

the frequency of description-to-inference links in the CFs.

- Results yielded small to moderate effect sizes, with the experts generating more descriptive ($d = .33$), diagnostic ($d = .24$), inferential ($d = .35$), and treatment planning information ($d = .32$) in their CFs, compared to novice and experienced therapists.

12. Haarhoff et al. (2011) New Zealand	• <i>SU/diagnosis formulated:</i> SUs diagnosed with MDD ($n = 2$), or GAD ($n = 2$). Total $n = 4$. • <i>Formulators:</i> A range of novice CBT therapists (e.g. psychologists, nurses, psychiatric registrars, general practitioners) that had recently graduated with a postgraduate diploma in CBT	• <i>Materials:</i> Written case vignettes of fictional SUs diagnosed with either MDD or GAD. Four ‘benchmark’ CFs (provided by an expert), based on the four case vignettes. The CFCCM (Eells et al., 1998, 2005). The QCTCF rating scale (Fothergill & Kuyken, 2002). The CBT CC rating scale (Haarhoff, 2008). • <i>Theoretical modality:</i> Cognitive behavioural. • <i>Training:</i> No training was provided as part of the study. • <i>Procedure:</i> Therapists were provided with written case vignettes, and then given 20-minutes (per case	• The total mean quality score range for the CFCCM ‘comprehensiveness’ scale, was between 4.2 to 4.8. This indicated that therapists, on average, attended to less than half of the potentially useful aspects of CF information. • The QCTCF rating scale, indicated that between 50% to 61% of therapists produced “good-enough” CBT CFs (total mean score range
---	--	--	---

(total $n = 26$). • <i>Raters</i>: Two RAs, and the first author (total $n = 3$).	vignette) to “think on their feet” and construct a CF for each case. • The CFs ($n = 104$) were transcribed, segmented into “idea units”, then content coded by the three raters, using the CFCCM. • To assess CF quality using the CFCCM, only ‘comprehensiveness’ was rated, which included 10 subcategories (scored from 0 to 10). • The QCTCF rating scale was used, which measures inferential aspects of the CF on a four-point scale (1 = very poor, 2 = poor, 3 = good enough, 4 = good). • The CBT CC rating scale (developed for this study), contained four categories: 1) the problem list, 2) diagnostic information, 3) working hypothesis, and 4) treatment planning. Each category was rated on a 10-point scale (0 = absent, 10 = excellent). Total scores ranged from 0-40 with higher scores indicating better quality CFs. • The therapists’ CFs were also assessed for quality by calculating the percentage agreement for information matching the expert ‘benchmark’ CFs. Benchmark CFs	was between 2.7 to 3.0). • For the CBT CC rating scale, the range of the total mean score was between 17.2 to 18.8. • 10% ($n = 11$) of the 104 CFs, were randomly selected to be rated for quality by an independent RA. 100% consensus within .5 was achieved for the QCTCF rating scale between the independent RAs. • The highest percentage agreements between the therapists and the expert ‘benchmark’ CF (as a measurement of quality) related to: ‘relevant childhood data’, ‘Axis I diagnosis’, ‘compensatory strategies’, and ‘core beliefs about self’. However, actual percentage figures were not provided. • Broad omissions between the therapists and the expert
--	--	--

were structured by the experts under nine headings: 1) relevant childhood data, 2) core beliefs, 3) underlying assumptions, 4) compensatory behaviours, 5) presenting problems, 6) diagnosis, 7) therapy interfering behaviours, 8) treatment plan, and 9) implications for the therapeutic relationship.

‘benchmark’ leading to very poor agreement (i.e. ‘almost no agreement’) related to: ‘therapy interfering behaviour’, and ‘Axis II PD traits’. However, again, actual percentage figures were not provided.

- Agreement between the therapists and the expert ‘benchmark’ for ‘problem list’ was also low for all 4 vignettes (<30%).

- Most therapists had a ‘good enough’ grasp of the fundamental CBT components of the CF.

However, they attended to <50% of the potentially useful aspects of the formulation i.e. they excluded socio-cultural, biological, protective, and interpersonal factors from their hypothesis of the SU’s presenting issues.

13. Kendjelic & Eells (2007) USA	• <i>SU/diagnosis formulated:</i> SU diagnoses were not fully stated. However, it was reported that the most common diagnoses were mood disorders (e.g. major depression, and anxiety disorders; $n = 99$). • <i>Formulators:</i> A range of clinicians and students (e.g. third- or fourth-year psychiatry residents, clinical or counselling psychology graduate students, social workers). Total $n = 43$. • <i>Raters (referred to as coders):</i> Advanced doctoral psychology graduate students ($n = 3$).	• <i>Materials:</i> The CFCCM (Eells et al., 1998, 2005). Intake evaluations/interviews for each SU. • <i>Theoretical modality:</i> We can infer from the description of the CF training, that the four Ps framework was used (atheoretical). • <i>Training:</i> The clinicians were randomly assigned to one of two groups: 1) received CF training (training group; $n = 20$), 2) did not receive CF training (control group; $n = 23$). Training group clinicians received a two-hour training session in CF (delivered by the first author) which emphasised the formulation of inferential material. Control group clinicians did not receive any training. Raters received training in the CFCCM. • <i>Procedure:</i> ‘Intake evaluations/interviews’ were conducted by clinicians (in the training/control groups) with SUs that attended a university-based outpatient psychiatric clinic. • Two to three written CFs were produced by each clinician, based on the intake evaluations/interviews (that had been copied from each SU’s chart).	• Overall mean Fleiss’ multirater kappa scores for the coded formulations across all variables and raters was .86 (range .60 to 1.0). This indicated moderate to almost perfect agreement. • Clinicians in the training group produced overall higher quality CFs than clinicians in the control group, based on a large mean effect size ($d = 1.56$). • Training group clinicians produced CFs that were more elaborated, comprehensive, complex, and precise than the control group. They were also more likely to identify precipitants, predisposing factors, and an inferred mechanism to explain SU difficulties. • Control group clinicians tended to summarise descriptive data. Except
--	---	--	--

		• The CFs were coded by blind raters for content and quality using the CFCCM.	for ‘problems with interpersonal relationships’, no inferential information was present in more than 10% of their CFs.
14. Mumma & Mooney (2007) USA	• <i>SU/diagnosis formulated:</i> A SU diagnosed with MDD, dysthymia, and GAD ($n = 1$). • <i>Formulators:</i> A novice clinician (clinical doctoral trainee; $n = 1$), an expert clinician (first author of the paper; $n = 1$; no demographic information provided). Total $n = 2$). • <i>Raters:</i> N/A. The CFs were assessed using data from the SU only, external judges/ratings were not required.	• <i>Materials:</i> A video of the first author undertaking the Cognitive-Behavioural-Interpersonal Semi-Structured Assessment Interview (CBISSAI; Mumma, 2001; Mumma & Smith, 2001) with a SU. Two self-report measures of distress from the ‘Negative Affectivity General Distress’ scale of the Mood and Anxiety Symptoms Questionnaire (MASQ; Watson et al., 1995), and a third self-report measure that was primarily based on the SCL-90 Anger/Hostility items (Derogatis, 1983), and items obtained from the CBISSAI. • <i>Theoretical modality:</i> Cognitive. • <i>Training:</i> No training was provided as part of the study. • <i>Procedure:</i> Clinicians watched the SU video, then independently they each developed a CF. Each CF incorporated at least two idiosyncratic cognitive	• The expert clinician identified four ICSs, and the novice clinician identified two ICSs. However, the novice’s ICSs were later combined into a single ICS owing to inadequate discriminant validity. • The ICSs identified in the CF produced by the expert clinician, explained approximately twice the variance in distress scores compared to the novice clinician’s ICS (average <i>shrunk</i> $R^2 = .46$ vs $.23$, respectively). • This indicated that the ICSs in the expert clinicians CF were more strongly related to each of the three distress variables than the novice

		schemas (ICSs). • Each clinician independently predicted (via their CF), which ICSs would be related to which types of SU distress (e.g. depression, anxiety, anger). • The SU completed self-report measures on a day-to-day basis over 81 consecutive days (n of observations = 80). • Using a latent variable confirmatory dynamic factor analysis (Ferrer & Nesselroade, 2003) approach, the two CFs (from each of the clinicians) were compared to see how well their chosen ICSs predicted daily variability in the SU's distress and symptoms.	clinician's combined ICS, after adjusting for the number of predictors. • Good rate of agreement was reported between the two clinician's CFs, by examining relationships between their hypothesised ICSs. The average Pearson's correlation between the novice clinician's "combined" ICS, and the expert clinicians' four ICSs was .78 (range .55 to .92). This suggests that the novice clinician's "combined" ICS captured a core construct represented in each of the novice clinician's two ICSs. However, the expert clinician's CF was substantially better in helping to explain the SU's distress.
15. Muran & Segal (1992)	USA • <i>SU/diagnosis formulated:</i> A SU diagnosed with	• <i>Materials:</i> Assessment interviews ($n = 2$). Self-scenarios ($n = 5$).	• Interrater reliability estimates of the self-scenarios were established

dysthymia, and anxiety disorders NOS ($n = 1$). • <i>Formulator (referred to as a third-party observer):</i> A clinical psychologist with three years of training in cognitive therapy ($n = 1$). • <i>Raters:</i> A clinical psychologist (referred to as an interviewer) with six years of clinical training in cognitive therapy ($n = 1$), the clinical psychologist (third-party observer) described above ($n = 1$), the SU described above ($n = 1$; total $n = 3$).	• <i>Theoretical modality:</i> Cognitive. • <i>Training:</i> No training was provided as part of the study. • <i>Procedure:</i> A clinical psychologist (interviewer) conducted two assessment interviews with a SU. Another clinical psychologist (third-party observer) then constructed five ‘self-scenarios’ based on the information derived from the assessment interviews (prior to delivering therapy with the SU). • Two of the self-scenarios were purposefully designed to directly reflect the SU’s schemas. The other three self-scenarios were designed to be less relevant. • Each of the five self-scenarios consisted of four components that reflected schematic structure: 1) a stimulus situation, 2) affective response, 3) motoric response, and 4) cognitive response. This resulted in 20 items in total, which were rated by the therapist, SU, and third-party observer on a 9-point Likert scale across eight parameters (e.g. frequency: “how often has such a scenario occurred recently?”, preoccupation: “how concerned have you been about this happening	by calculating ICCs on these ratings. • Good to excellent rates of agreement were reported between the interviewer, SU, and the observer: stimulus situation (ICC = .92), affective response (ICC = .97), motoric response (ICC = .95), and cognitive response (ICC = .88). • Mean ICC = .93 when averaged across the three raters.
--	--	--

		recently?”), to assess the clinical relevance of each scenario.	
		• Interrater reliability was calculated after the first therapy session, once the SU, interviewer, and third-party observer had completed the 20-item ratings.	
16. Muran et al. (1998) USA	• <i>SU/diagnosis formulated:</i> Outpatients (total $n = 6$) were diagnosed with a mood disorder. Of these outpatients, $n = 2$ were also diagnosed with an anxiety disorder, $n = 2$ were also diagnosed with a PD, and $n = 4$ were also diagnosed with a PD NOS. • <i>Formulators (referred to as a third-party constructors):</i> The first and third authors (total $n = 2$); no demographic information provided. • <i>Raters (referred to as</i>	• <i>Materials:</i> Interpersonal scenarios ($n = 6$). Videos of semi-structured interviews ($n = 6$). • <i>Theoretical modality:</i> One model was interpretative and based on the principles of ego psychology (Pollack et al. 1992). The other model was based on an integration of cognitive-interpersonal traditions (Safran & Muran, 1995; Safran & Segal, 1990). Both models shared an interpersonal focus. • <i>Training:</i> Third-party raters (observers) received four-hours of training in relation to the semi-structured interviews. • <i>Procedure:</i> All SUs engaged in 30-sessions of therapy. • Third-party raters (observers) conducted the semi-structured interviews. SUs were asked to describe their interpersonal transactions with others: 1) “when I am at	• Interrater reliability estimates of the interpersonal scenarios were established by calculating ICCs on interviewer, observer, and constructor ratings of the relevance of the scenarios (and their respective components), in relation to each SU. • ICC range = .88 to .97 when averaged across all six scenarios were reported in relation to the clinical ‘relevance’ of each interpersonal scenario, for each SU. • ICC range = .87 to .97 when averaged across all six scenarios were reported for the component

<i>observers</i>): Graduate RAs acted as third-party raters and interviewers ($n = 4$), SUs ($n = 6$, information provided above), third-party constructors ($n = 2$, information provided above. Total $n = 12$.	my best”, and 2) “when I am at my worst”. The interviews were videotaped. • A ‘constructor’ then watched the video and constructed six interpersonal scenarios (two sets of three), for each SU. Three prototypical scenarios represented how the individual typically interacted with others when at their ‘best’, and three prototypical scenarios represented how the SU interacted with others when at their ‘worst’. • One to two scenarios per set, were purposefully designed to be clinically representative of the SU, whereas the other scenarios were designed to be less relevant. • Each scenario contained three components: 1) automatic thoughts/immediate feelings of the self, 2) interpersonal action of the self, and 3) interpersonal reaction of the other. • The SU, interviewer, observer, and constructor then rated (on a seven-point Likert scale), how relevant they thought each scenario was, in relation to the SU. • The three components were then broken down into a	ratings in relation to when each SU felt at their ‘best’. • ICC range = .80 to .94 when averaged across all six scenarios were reported for the component ratings in relation to when each SU felt at their ‘worst’. • Test-retest reliability for all six parameters across all 29 therapy sessions averaged resulted in an ICC range = .92 to .97 for the ‘best’ case scenario, and ICC range = .79 to .97 for the ‘worst’ case scenario.
--	---	--

	single sentence for all six vignettes. This resulted in 18 items (nine items anchored to when the SU felt at their ‘best’, and nine anchored to when they felt their ‘worst’). Again, these items were rated on a seven-point Likert scale by the SU, interviewer, observer, and constructor, in relation to how relevant they thought each item was, in relation to the SU. • These ratings were all used in different combinations to establish interrater reliability. • After every therapy session, each SU was asked to rate six parameters on a seven-point Likert scale (e.g. frequency, preoccupation), for the scenarios they felt were clinically relevant/representative of them (at their ‘best’, and at their ‘worst’). The repeated ratings on these parameters were used to establish test-retest reliability, and to measure change.		
17. Muran et al. (2001) USA	• <i>SU/diagnosis formulated:</i> A SU diagnosed with GAD and avoidant PD ($n = 1$). • <i>Formulators (referred to as a third-party constructor):</i>	• <i>Materials:</i> Video of a semi-structured interview. Interpersonal scenarios ($n = 6$). • <i>Theoretical modality:</i> Cognitive-interpersonal. • <i>Training:</i> No training was provided as part of the study.	• Between the interviewer, observer, and the SU, ‘best’ case ICC = .96, “worst” case ICC = .82, when averaged across all six scenarios were reported in relation to the

Total $n = 1$; no demographic information provided. • <i>Raters</i>: One interviewer (no demographic information provided), the SU (information provided above), one observer (no demographic information provided), and a third-party constructor (information provided above). Total $n = 4$.	• <i>Procedure</i>: The SU engaged with 40-sessions of therapy. • An interviewer conducted a semi-structured interview with the SU. The SU was asked to describe their interpersonal transactions with others: 1) “when I am at my best”, and 2) “when I am at my worst”. • A ‘constructor’ then watched the video and constructed six interpersonal scenarios (two sets of three), for the SU. Three prototypical scenarios represented how the individual typically interacted with others when at their ‘best’, and three prototypical scenarios represented how the SU interacted with others when at their ‘worst’. • One to two scenarios per set, were purposefully designed to be clinically representative of the SU, whereas the other scenarios were designed to be less relevant. • Each scenario contained three components: 1) automatic thoughts/immediate feelings of the self, 2) interpersonal action of the self, and 3) interpersonal reaction of the other.	clinical ‘relevance’ of each interpersonal scenario, for the SU. • Between the interviewer, observer, SU, and constructor, ‘best’ case ICC = .96, ‘worst’ case ICC = .89, when averaged across all six scenarios were reported in relation to the clinical ‘relevance’ of each interpersonal scenario, for the SU. • Between the interviewer and observer, ‘best’ case ICC range = .83 to .98, ‘worst’ case ICC range = .95 to 1.0, when averaged across all six scenarios for the component ratings. • Between the interviewer, observer, and SU, ‘best’ case ICC range = .86 to .99, ‘worst’ case ICC range = .55 to .87, when averaged across all six scenarios, for the component ratings. • Between the interviewer, observer, SU, and constructor, ‘best’ case ICC
---	---	--

-
- | | |
|---|---|
| <ul style="list-style-type: none"> • The SU, interviewer, observer, and constructor then rated (on a seven-point Likert scale), how relevant they thought each scenario was, in relation to the SU. • The three components were then broken down into a single sentence for all six vignettes. This resulted in 18 items (nine items anchored to when the SU felt at their ‘best’, and nine anchored to when they felt their ‘worst’). Again, these items were rated on a seven-point Likert scale by the SU, interviewer, observer, and constructor, in relation to how relevant they thought each item was, in relation to the SU. • These ratings were all used in different combinations to establish interrater reliability. • After every therapy session, the SU was asked to rate seven parameters on a seven-point Likert scale (e.g. frequency, preoccupation), for the scenarios they felt were clinically relevant/representative of them (at their ‘best’, and ‘worst’). The repeated ratings on these parameters were used to measure change. | <p>range = .80 to .97, ‘worst’ case ICC range = .53 to .88, when averaged across all six scenarios, for the component ratings.</p> |
|---|---|
-

Table 2.4 Study details: Papers identified in a review by Flinn et al. (2015; $n = 15$ studies)

Author(s), location, and area of CF studied	Participant characteristics	Method	Key findings
18. Barber et al. (1995) USA	• <i>SU/diagnosis formulated:</i> SUs diagnosed with major depression (total $n = 19$), and of these, $n = 11$ also had at least one PD. • <i>Formulators/raters (referred to as judges):</i> Two teams, with $n = 2$ judges in each team (total $n = 4$). The judges were all experienced psychodynamic clinicians • <i>Interviewer:</i> RA (total $n = 1$); no demographic information provided.	• <i>Materials:</i> Written transcripts from Relationship Anecdotes Paradigm (RAP) interviews. The CCRT. • <i>Theoretical modality:</i> Psychodynamic. • <i>Training:</i> Prior to the study, judges had been trained by L. Luborsky in the CCRT method (Luborsky & Crits-Christoph, 1990) method. • <i>Procedure:</i> The CCRT describes relationship patterns or conflict in terms of 3 components: 1) W, 2) RO, and 3) RS. • RAP interviews ($n = 19$) were undertaken by a RA with each SU, prior to receipt of therapy. The 30-minute RAP interviews were based around the identification of 10 relationship episodes. The interviews were audio recorded and transcribed. • CCRTs were extracted via narratives based on RAP interviews.	• Weighted kappa for the clustered categories ranged from .64 to .81 for CCRTs based on therapy sessions three and five. • Weighted kappa for the clustered categories ranged from .60 to .68 for CCRTs based on RAP interviews. • Weighted kappa for the clustered categories ranged from .40 to 1.0 when comparing CCRTs based on early therapy sessions, to the CCRTs based on RAP interviews. • The CCRTs extracted in pre-treatment interviews, were similar to the CCRTs extracted from early

		• SUs were then offered 16-sessions of time-limited supportive-expressive dynamic psychotherapy from an experienced therapist ($n = 17$ engaged in therapy). • CCRTs were then extracted via narratives based on transcribed therapy sessions three and five, in relation to the identification of 10 relationship episodes. • A team of judges independently read the RAP interview transcripts and identified/rated each of the three components (W, RO, and RS). A separate team of judges then independently read the therapy session transcripts and identified/rated each of the three components. • The two methods used to construct the CCRT formulations were compared. • There were eight clustered standard categories for each CCRT component, and interjudge agreement was reported. Two of the most frequent ratings for each of the CCRT components (from each judge) were chosen.	sessions of brief dynamic psychotherapy. • Results indicated moderate to good levels of agreement between two methods for the construction of CCRT formulations. • These findings increase researchers' confidence that RAP interviews can be used to determine SUs' psychodynamic themes, for the construction of formulations, independently of treatment.
19. Caston & Martin (1993) USA	• <i>SU/diagnosis formulated:</i> A SU experiencing low self-esteem, difficulty	• <i>Materials:</i> Verbatim transcripts of the first five analytic hours of each SU's therapy. • <i>Theoretical modality:</i> Psychoanalytic.	• For domains one to four, agreement on 'order' and 'magnitude' for the pooled four

experiencing pleasure, and a lack of sexual responsiveness. Total $n = 1$. • <i>Formulators (referred to as judges)</i>: Team 1 ($n = 2$) qualified psychoanalysts who were the authors of the study. • <i>Raters (also referred to as judges)</i>: Team 2 ($n = 4$) qualified psychoanalysts, referred to as the ‘textwise’ judges. • Team 3 ($n = 4$) qualified psychoanalysts, referred to as the ‘textless’ judges. Total $n = 8$.	• <i>Training</i>: No training was provided as part of the study. • <i>Procedure</i>: The team 1 judges generated five case-specific psychoanalytic domains from the transcripts: 1) conflictedness in the historical moment, and conflictedness in the telling, 2) defense-impulse configurations, 3) transferences in defense-impulse configurations, 4) historical linkages, 5) inferred wishes. • Items for conflictedness were drawn from the entire five therapy sessions. Items for all other domains were drawn from session five only. • The ‘textwise’ judges were given the verbatim text to read, then asked to independently rate the clinical propositions in the different domains (for both order, and magnitude), on a nine-point Likert scale. • The ‘textless’ judges were <i>not</i> given the verbatim transcripts to read and were asked to independently rate the same clinical propositions in the different domains (for both order, and magnitude), on a nine-point Likert scale. This was purposefully designed to	‘textwise’ judges yielded Spearman rho correlations that ranged from .61 to .96. For domain five (inferred wishes), Spearman rho correlations ranged from .02 to .91. • For domains one to four, agreement on ‘order’ and ‘magnitude’ for the four ‘textwise’ judges <i>and</i> the four ‘textless’ judges yielded Spearman rho correlations that ranged from .08 to .77. For domain 5 (inferred wishes), Spearman rho correlations ranged from .31 to .87. • In summary, relative ‘order’ of clinical intensity or frequency obtained higher levels of agreement, than comparisons of specific ‘magnitudes’. • Reliabilities of formulations made by ‘textwise’ psychoanalysts, passed
---	---	--

		carry out a ‘test of the mannequin’. • Clinicians that apply stereotypical notions when making sense of a case, are referred to as ‘mannequin’ formulations – clinically <i>stereotypical</i> formulations (i.e. based on a diagnosis) that do not incorporate idiosyncratic features. • ‘Order’ refers to the ‘rank-order’ of items whereby intervals between scale points do not represent equal units, but instead relate to levels that are greater or lesser than each other. • ‘Magnitude’ refers to the relative magnitude of the item, with agreement deemed sufficient if judges’ ratings on the nine-point scale fell within, but not more than, a two-point span.	the ‘test of the mannequin’ in most domains. This surpassed the reliability values obtained between the ‘textless’ and the ‘textwise’ analysts combined. • Psychoanalysts can agree on the essential features of formulations drawn from five analytic sessions. • Text information is necessary to arrive at agreements that are more case-specific than mere stereotypes or ‘mannequins’.
20. Collins & Messer (1991) USA	• <i>SU/diagnosis formulated:</i> Case A was diagnosed with mild depression. Case B was diagnosed with moderate depression. Total $n = 2$. • <i>Formulators/raters (referred to as judges):</i> Two	• <i>Materials:</i> Case material for each case. The PFM. • <i>Theoretical modality:</i> Cognitive-psychoanalytic. • <i>Training:</i> Rutgers clinicians were trained to use the PFM (in relation to object relations theory) over approximately 15-hours of practice. Prior to the study, Mt. Zion clinicians had already been trained to use the PFM (in relation to cognitive-psychoanalytic theory).	• ICC for pooled judges for the Rutgers ‘plan formulations’ in time 1 for both cases ranged from .81 to .93, and the Mount Zion ‘plan formulations’ in time 1 for both cases ranged from .81 to .95.

panels: 1) Mt. Zion judges ($n = 5$) were psychiatrists or psychologists with at least three years of experience using the PFM (Curtis & Silberschatz, 1989), 2) Rutgers judges ($n = 5$; profession not reported) had either 20 years' experience of providing short-term psychodynamic therapy ($n = 1$), two to five years of experience ($n = 3$), or three months' of experience ($n = 1$).	• <i>Procedure:</i> Both SUs completed short-term dynamic therapies with experienced doctoral-level psychotherapists. Case A was treated at the Rutgers College Counseling Center. Case B was treated at the Mt. Zion Medical Center. • A 'plan formulation' contains four components: 1) goals, 2) obstructions, 3) tests, and 4) insights. • The PFM had previously been devised by the 'Mt. Zion Psychotherapy Research Group' to test cognitive-psychoanalytic theory. For this study, the PFM was applied by the 'Rutgers Short Term Psychotherapy Research Group' to test the reliability of the PFM using object relations theory. • For Cases A and B, the case material was independently rated by the Rutgers and Mt. Zion panels using the PFM. • Blinding procedures involved the Mt. Zion judges rating the Rutgers items that had been selected for the Rutgers Plan, and the Rutgers judges rating the Mt. Zion items that had been selected for the Mt. Zion Plan.	• ICC for pooled judges for the Rutgers 'plan formulations' in time 2 for both cases ranged from .75 to .96. • Therefore, the reliability of both the Rutgers and the Mt. Zion 'plan formulations' for both cases were very high. • Pearson product-moment correlations were performed on the four formulation components between Rutgers judges' mean time 1 and time 2 ratings for Case A: goals, $r = .94$, obstructions, $r = .98$, tests, $r = .96$, and insights, $r = .96$. • Therefore, there was a nearly perfect positive relationship between Rutgers judges' evaluations of their Case A 'plan formulation'
--	--	--

-
- | | |
|--|---|
| <ul style="list-style-type: none"> • Rutgers ‘plan formulations’ were developed at time 1 and time 2 for each of the two cases, three-months apart (Rutgers ‘plan formulations’ total $n = 4$). A single Mt. Zion Plan was also developed for each case (Zion ‘plan formulations’ total $n = 2$). | <p>components at time 1 and time 2.</p> <p>Comparable calculations could not be made for Case B owing to time constraints.</p> <ul style="list-style-type: none"> • To verify these findings in a different way, percentages of time-1-selected items retained at time 2 were calculated for each component of the ‘plan formulations’ for both cases. Percentage agreement ratings indicated high levels of stability, based on test-retest reliability. • Between 85% and 97% of Case A items and between 90% and 96% of Case B items were retained at three-month follow-up. • In summary, the PFM can be reliably utilised in a different setting, with different case material, different interviewer-therapists, different |
|--|---|
-

			clinical judges, and with a different theoretical orientation.
21. Crits-Christoph et al. (1988) USA	• <i>SU/diagnosis formulated:</i> SUs diagnosed with a variety of different mental health difficulties (e.g. PDs, GAD, dysthymic disorder). Total $n = 35$. • <i>Formulators (referred to as CCRT judges):</i> Experienced psychodynamic clinicians, $n = 1$ psychiatrist, and $n = 1$ clinical psychologist. Total $n = 2$. • <i>Raters (referred to as judges):</i> Relationship episode (RE) selection judges ($n = 2$) were 1 psychiatrist, and 1 RA. CCRT judges ($n = 2$; described above). Similarity judges ($n = 4$) were RAs that	• <i>Materials:</i> The CCRT. Written therapy transcripts. • <i>Theoretical modality:</i> Psychodynamic. • <i>Training:</i> CCRT judges were trained in the use of the CCRT manual to formulate CCRT components. • <i>Procedure:</i> Independent RE judges identified a minimum of 10 REs in the psychotherapy transcripts (usually from sessions three and five). • Each RE was rated for ‘completeness’ on a scale of 1-5. • REs with an average of two RE selection judges’ ratings equal to 2.5 (or more) were put forward for CCRT scoring. • CCRT judges then independently read each RE and scored each SU’s interpersonal patterns, specifically, their: 1) W, 2) RO, and 3) RS to formulate a preliminary CCRT based on the most frequent types of each component. • Reliability was assessed in two different ways: Firstly, tailor-made (i.e. idiosyncratic wording) of the	• The reliability of RE selection judges identifying the REs was satisfactory. The ICC rating of RE ‘completeness’ for the pooled judges was reported as .68. • Interjudge reliability (for the similarity ratings in relation to tailor-made formulations that were either correctly matched or mismatched to the case), for the pooled similarity judges was reported as ICC = .79. • Interjudge reliability using the standard set of CCRT categories, were reported with weighted kappa values ranging from $\kappa_W = .61$ (for W and RS), and $\kappa_W = .70$ (for RO).

translated the tailor-made CCRTs into standard categories.

relationship themes were produced by the CCRT judges for each SU's CCRT. Blind similarity judges then rated the similarity of each tailor-made CCRT of one CCRT judge, to the tailor-made CCRT of the other CCRT judge. In addition, the similarity judges rated (on a scale of 1-7, 1 = completely different, 7 = identical) the first tailor-made formulation of each case, to seven other tailor-made formulations randomly chosen from different cases, scored by the second CCRT judge.

- A second method of assessing interjudge agreement involved judges coding the tailor-made CCRTs into 'ready-made' (standard) CCRT categories for nominal data. The purpose of this coding step was to eliminate any 'wording' differences between CCRT judges (as seen in the first method, whereby judges used idiosyncratic wording, tailored to suit the individual case). The judges then picked the standard category that was the closest in meaning, to the tailor-made wording.

22. DeWitt et al. (1983) USA	• <i>SU/diagnosis formulated:</i> SUs experiencing pathological grief reactions and diagnosed with either PTSD, adjustment disorder with anxiety, depression, or mixed features. Total $n = 18$. • <i>Intake evaluation clinicians:</i> Described as ‘experienced clinicians’ (no n or demographic information provided). • <i>Formulators (referred to as judges):</i> A mix of mental health professionals ($n = 6$ psychiatrists, $n = 2$ social workers, and $n = 1$ psychologist). All clinicians were psychodynamically trained, with at least five-years of post-degree	• <i>Materials:</i> Audiotapes or videotapes of SU intake evaluation interviews. Standardised record sheets. • <i>Theoretical modality:</i> Psychodynamic. • <i>Training:</i> No training was provided as part of the study. • <i>Procedure:</i> SUs engaged in 60 to 90-minute semi structured intake evaluation interviews, conducted by experienced clinicians. • To test the reliability of Malan's method (Malan, 1959, 1973, 1976), two teams of judges (A and B) viewed a taped intake evaluation interview for each SU. The two teams of judges worked independently from one another; they did not discuss the study cases. Using the record sheets, they developed a dynamic formulation that contained two elements: 1) a ‘hypothesis’ and 2) a set of ‘criteria’. • The ‘hypothesis’ is the neurotic conflict that is judged to be central to the SU’s difficulties, whereas the ‘criteria’ set proposes ways in which changes in the neurotic conflict may be observed via the SU’s behaviour, post-therapy e.g. idiosyncratic statements	• Teams A and B differed in their formulations. The mean rating for hypothesis and criteria sets combined, was 3.6 (51% agreement). This figure fell between the agreement scale for points 3 and 4, defined as ‘definite nonoverlap with some distinct areas of agreement or similarity in inclusion’ (for a score of a 3), and ‘equal amounts of overlap and nonoverlap’ (for score of a 4). • In summary, the agreement between psychodynamic formulations developed by independent teams of clinical judges was found to be poor.
---------------------------------	---	---	--

	experience. Total $n = 9$. • The nine judges were divided into three teams of three judges: team A, team B, and team C. Team C judges were used in another part of the study which is not reported here. • <i>Raters (referred to as agreement raters):</i> A mix of mental health professionals ($n = 1$ psychiatrist, and $n = 3$ psychologists). All clinicians were psychodynamically trained, with at least five-years of post-degree experience. Total $n = 4$.	such as: “inner feelings of confidence” which could then be rated as a specific criterion of change for that individual. • Levels of agreement between judges (A and B) in relation to the formulated hypothesis and criteria sets for each SU were rated by the agreement raters. • Two agreement raters judged ‘hypothesis’ agreement, and the other two agreement raters judged ‘criteria’ agreement. • The two sets of criteria (one set of criteria from each formulator team, per case), were rated via agreement ratings scored from (.00, .17, .33, .50, .67, .84, and 1.0, which converted to points 1-7). • The two hypotheses (one hypothesis from each formulator team, per case), were rated in terms of formulator agreement on a scale of 1-7, 7 = essential overlap, 1 = essential non-overlap).	
23. Dudley et al. (2010) UK	• <i>SU/diagnosis formulated:</i> A SU diagnosed with psychosis. Total $n = 1$. • <i>Formulators:</i> A mix of	• <i>Materials:</i> Demographic questionnaire. A 30-minute role-play video of an assessment session. A timeline that detailed the SU’s history. CF templates based on Beck’s (1976) generic cognitive behavioural therapy	• Mean percentage agreement between the clinicians’ and expert benchmark formulation (in relation to the formulation items) ranged

clinicians (e.g. psychiatric/mental health nurses $n = 36$, clinical psychologists $n = 5$, trainee clinical psychologists $n = 17$, social workers $n = 6$. Total $n = 82$. • <i>Expert panel</i>: A mix of highly experienced clinicians (total $n = 3$) recognised for their contribution to training/research in the field of psycho-social interventions, and CBTp. • <i>Role-play actor</i>: A clinical psychologist (total $n = 1$). • <i>Raters (referred to as the research team)</i>: Authors of the study (total $n = 4$); no demographic information provided.	model. A scoring manual. • <i>Theoretical modality</i>: Cognitive. • <i>Training</i>: Prior to the study, All the mental health professionals had undertaken at least a four-day basic course in CBT. • <i>Procedure</i>: The materials were all piloted in a pilot study, and a revised CF template was approved by the panel of experts. • The expert panel members independently watched the video and used the timeline information to develop a cognitive formulation. Items agreed by all three experts were included in the final ‘benchmark’ formulation. • A scoring manual was then developed by the research team, based on the benchmark formulation. • Formulators then watched a role-play video of an assessment and timeline, with an actor playing the role of a SU. The information was based on a real SU experiencing psychosis, who had provided consent. • Clinicians independently constructed a formulation of the case, using the CF template. • These formulations were then rated (using a scoring	from 32.2% to 91.6%. • There were higher rates of agreement amongst clinicians regarding the overt/descriptive level aspects of the formulation (e.g. behaviours, emotions), in comparison to the theory based inferential components (e.g. identification of core beliefs, dysfunctional assumptions). • Greater number of years of clinical experience amongst clinicians, improved reliability/agreement with the expert benchmark formulation overall. • CBT case experience was related to the reliability of clinicians formulating inferential aspects of the case.
--	---	---

		manual) to compare the clinician's formulations to the 'benchmark' formulation provided by the experts. • Formulators then completed a demographic questionnaire in relation to their professional background, academic qualifications, skills, and experience. • Raters then rated the clinician's formulations using the manual. Each formulation component was scored either '0' = inaccurate, '1' = theme identified, or '2' = accurate. Clinicians who correctly identified a component from the 'benchmark' formulation, scored either 1 or 2 on that item. Formulation items were summed to produce a total score. • The scoring manual also included an 'inferential' subscale (e.g. core beliefs, and dysfunctional assumptions), and a 'stress-vulnerability' subscale (e.g. early experiences, stressors). • An acceptable level of interrater reliability between the raters was achieved (>.85).	
24. Eells et al. (1995)	• <i>SU/diagnosis formulated:</i> A SU diagnosed with	• <i>Materials:</i> Transcripts from the first five psychotherapy sessions for each SU.	• ICCs for the RRM quadrants between the 20 clinical judges,

pathological grief disorder (including PTSD, and MDD), and a SU diagnosed with social phobia. Total $n = 2$. • <i>Therapists</i>: Highly experienced psychotherapists, each with more than 20 years of clinical experience ($n = 2$). • <i>Formulators</i>: Team one from Yale University consisted of 10 graduate-level students and two clinical psychologists ($n = 12$). Team two from the University of California, San Francisco (UCSF) consisted of two clinical psychologists, and one psychiatrist ($n = 3$). Total $n = 15$.	• <i>Theoretical modality</i>: Described as integrative (based on psychodynamic, interpersonal, and cognitive models). • <i>Training</i>: No training was provided as part of the study. • <i>Procedure</i>: Both SUs were seen weekly for psychotherapy, and these sessions were videotaped. • Two teams (Yale and UCSF) independently read the therapy transcripts and constructed a CF for each SU, following the RRM method. • The RRM method (Horowitz, 1989, 1991b) is based on person schemas theory. It attempts to integrate psychodynamic, interpersonal, and cognitive psychotherapy. • Mapping out a RRM in a diagram, there are four quadrants: 1) dreaded, 2) desired, 3) problematic compromise, and 4) adaptive compromise. Each quadrant contains a Role-Relationship Model (RRM), that relates to the roles of self, other, and a transactional script. • The RRM ‘desired’ organises states of potential	indicated good reliability, with the two pairs of formulations generating a mean ICC of .74 for the ‘desired’ RRM quadrant, and a mean ICC of .87 for the ‘dreaded’ RRM quadrant. • ICCs for the ‘correctly matched’ items between the 20 clinical judges, indicated good reliability, with the two pairs of formulations generating a mean ICC of .74 for RRM items, and a mean ICC of .89 for the RRM items. • The Yale and UCSF formulation teams captured similar information in their RRM for the two cases, despite using different sub-step procedures to construct their formulations.
--	--	---

• *Raters (referred to as clinical judges):* Clinical Psychology graduate students (total $n = 20$).

pleasure and satisfaction. Whereas the ‘dreaded’ RRM organises states of intense imagined (or real) suffering, and loss of control.

- The goal of the Yale and UCSF formulators was to identify the important person schemas and relationship scripts, then arrange them into RRMCs. However, both teams of clinicians used different sub-step procedures to achieve this aim.

- Both teams agreed upon a set of relationships in the SU’s life, for which RRMCs would be constructed. The relationships chosen were those most frequently discussed in the psychotherapy sessions.

- After the RRMCs were formulated, they were rated for similarity by a separate group of clinical judges. Their task was to compare the similarity of the Yale and UCSF sets of RRMCs.

- Each RRMC (and the RRM that comprised it) were translated from the diagrams into a sentence/item format. To compare the Yale and UCSF formulations, these items were then rated as to whether they represented a ‘correct match’ or ‘incorrect match’ with

		one another. • Similarity ratings were also made on a six-point Likert scale from: 1 = extremely dissimilar, to 6 = extremely similar.	
25. Kuyken et al. (2005) UK	• <i>SU/diagnosis formulated:</i> A SU diagnosed with major depression, and personality difficulties. Total $n = 1$. • <i>Formulators:</i> A mix of clinicians (e.g. clinical psychologists, $n = 35$, psychiatric nurses, $n = 19$, pre-qualification students, $n = 29$, counsellors, $n = 14$). All clinicians had (at least) a basic level knowledge and experience of cognitive therapy (total $n = 115$). • <i>Raters (referred to as judges):</i> ($n = 3$); no	• <i>Materials:</i> A five-minute video of a SU named 'Anna' (pseudonym). A 'benchmark' formulation provided by J. Beck. CCDs (Beck, 1995). Intake assessment information of 'Anna' (including an account of Anna's presenting problems, psychosocial history, completed psychometric assessment measures etc.). • <i>Theoretical modality:</i> Cognitive. • <i>Training:</i> Clinicians participated in a training workshop (provided by the first author) to train them how to use the CCD framework. • <i>Procedure:</i> At the end of the training workshop, the clinicians were given the intake assessment information. They then watched a five-minute role-play video of 'Anna' (role-played by an actor), and the first author (who acted as the cognitive therapist). • The clinicians were then asked to independently	• Agreement amongst clinicians CFs, in relation to the formulation categories using Cohen's kappa, ranged from $\kappa = .63$ to $\kappa = .91$. • The third 'experienced' judges' ratings of agreement using Cohen's kappa, ranged from $\kappa = .79$ to $\kappa = .97$. • Overall percentage agreement on the individual formulation components (between the individual clinicians and the expert 'benchmark' formulation) ranged from 7% to 73%. • Results indicated that clinicians could agree with each other (and with the expert 'benchmark')

demographic information provided.	formulate the case using a blank copy of the J. Beck CCD. • The clinicians' formulations were compared with one another, as well as with a 'benchmark' formulation (provided by J. Beck). • Clinicians' formulations were also assessed for quality using the QCTCF rating scale (Fothergill & Kuyken, 2002). This scale categorises CFs as either 'good', 'good enough', 'poor', or 'very poor'. • Finally, clinicians completed a demographic questionnaire to provide information about their professional background, amount of clinical experience, and any accreditation credentials. • To establish percentage agreement amongst clinicians, content analysis was used by two judges. Statements in the clinicians' CFs were put onto cards, and organised into categories (e.g. core beliefs, compensatory strategies). • A third, more experienced judge was given the category titles. They then checked the analysis by conducting their own content analysis of the CFs.	formulation), on the descriptive aspects of the case (e.g. presenting issues). However, agreement decreased for the more theory-driven inferential aspects (e.g. inferring dysfunctional assumptions). • Generally, the pre-qualified group were less likely to identify an important aspect in the benchmark formulation, and in formulations developed by the qualified clinicians. • Quality of the formulations ranged from 'very poor' to 'good', with only 44% rated as being (at least) 'good enough'. This suggested there was considerable variation. • Higher quality CFs were positively associated with number of years post-qualification experience, and BABCP accreditation status.
-----------------------------------	---	---

26. Mumma & Smith (2001) USA	• <i>SU/diagnosis formulated:</i> SUs met diagnostic criteria for mood disorders, anxiety disorders, or both. Three SUs met criteria for MDD (one of these SUs also met criteria for GAD. Another SU also met criteria for panic disorder with agoraphobia, dysthymic disorder, and PTSD). A fourth SU met diagnostic criteria for Bipolar II disorder. Total $n = 4$. • <i>Interviewing clinician(s):</i> No n or demographic information provided. • <i>Formulators:</i> A clinical psychologist ($n = 1$), and trainee clinical psychologists ($n = 3$). Total $n = 4$. • <i>Raters:</i> Clinical	• <i>Materials:</i> Videotapes of semi-structured assessment interviews with each SU. Cognitive-Behavioural-Interpersonal scenarios (CBISs; $n = 21$). A set of interpretive guidelines (based on Beck, 1995; Mumma, 1998; and Persons, 1989). • <i>Theoretical modality:</i> Cognitive behavioural, and cognitive-behavioural-interpersonal. • <i>Training:</i> No training was provided as part of the study. • <i>Procedure:</i> Two semi-structured assessment interviews with each SU focused on cognitive, affective, and motoric behaviours included in self-scenarios. • Three to four salient, distressing, recent, and typical events were also explore, with the interviewer eliciting a) anticipated (e.g. desired, feared) and actual responses of others, b) the SU's response to others, and c) the meaning of the event to the SU. • Pairs of formulators watched the videotapes of the interview assessments, then (using the interpretive	• ICC mean reliability rating (pooled over four similarity raters) was .90. This indicated a highly reliable between the formulators on the similarity of the situational component of the CBISs. • Situationally similar CBISs (constructed by different formulators) tended to have similar cognitive, affective, symptomatic, and interpersonal content.
------------------------------------	--	---	---

	psychologists ($n = 8$), and counselling psychologists ($n = 2$). All raters had doctoral degrees and worked clinically from a cognitive behavioural, or cognitive-interpersonal perspective. Total $n = 10$. • <i>Similarity raters</i>: Clinical psychology doctoral trainees (total $n = 4$).	guidelines) they independently formulated two or three CBISs for each of the four SUs. • CBISs are situation-level idiographic CFs. They are narrative descriptions of ‘prototypical’ cognitive, affective, behavioural, and interpersonal responding, in relation to distressing situations. Such patterns are common to two or more events and reflect how the SU typically responds & interprets these events. • Similarity raters then independently rated the situational components of pairs of CBISs (developed by different formulators) on a 9-point Likert-type scale from 0 (not at all similar) to 8 (very highly similar).	
27. Muran et al. (1994) USA	• <i>SU/diagnosis formulated</i>: SUs were outpatients diagnosed with either a mood disorder (e.g. dysthymia, major depression, and double depression; total $n = 5$), an anxiety disorder (e.g. GAD and social phobia; total $n = 2$), or both a mood <i>and</i> an	• <i>Materials</i>: Two audiotapes for each of the assessment interviews. Self-scenarios ($n = 19$). 20-item rating forms. • <i>Theoretical modality</i>: Cognitive. • <i>Training</i>: No training was provided as part of the study. • <i>Procedure</i>: An interviewer conducted a two-session assessment interview with each SU.	• Mean rates of agreement (averaged across all eight cases) between the SUs, therapists, and third-party observers, in relation to the four components of the self-scenarios: 1) stimulus situation, ICC = .92, 2) affective responses, ICC = .90, 3) motoric responses, ICC = .91, and 4) cognitive responses, ICC = .93.

anxiety disorder (e.g. dysthymia and social phobia; total $n = 1$). Total $n = 8$. • <i>Therapists (also referred to as interviewers)</i>: Qualified clinicians with an average number of 3 years-experience in cognitive therapy. Total $n = 5$. • <i>Formulators (referred to as third-party observers)</i>: One PhD-level clinician with three years-experience of the cognitive model, and one MA-level clinician with one-year experience of the cognitive model. Total $n = 2$. • <i>Raters</i>: Therapists ($n = 5$, information provided above), SUs ($n = 8$, information provided above), and third-	• Third-party observers then independently created up to five self-scenarios for each SU (M number of self-scenarios per SU = 2.5) based on audiotaped semi-structured assessment interviews. • Self-scenarios capture dysfunctional self-schemas in terms of ‘nuclear’ scenes. These highly distressing events (ideographically constructed for each SU) consist of four components, which reflect schematic structure: 1) a stimulus situation, 2) affective responses, 3) motoric responses, and 4) cognitive responses. • One to three of the self-scenarios were purposefully designed to reflect the SU’s schemas. Whereas the other self-scenarios were designed to be less relevant. • Each of the four components were separated into a single sentence or statement. This resulted in $n = 20$ items. These items were then rated by the therapist, SU, and third-party observer on a nine-point Likert scale across seven parameters (e.g. frequency: “how often has such a scenario occurred recently?”, preoccupation: “how concerned have you been about
--	--

	party observers ($n = 2$, information provided above) Total $n = 10$.	this happening recently?”), to assess the clinical relevance of each scenario. • Interrater reliability was calculated after the first therapy session, once the SU, interviewer, and third-party observer had completed all 20-item ratings.	
28. Perry et al. (1989) USA	• <i>SU/diagnosis formulated:</i> SUs diagnosed with either BPD ($n = 8$), antisocial PD ($n = 6$), or bipolar type II affective disorder ($n = 6$). Total $n = 20$. • <i>Interviewers:</i> Experienced clinicians with at least 10 years post-training experience. Two of the interviewers were psychoanalysts (total n not reported). • <i>Formulators (referred to as clinician raters):</i> Clinicians with at least five years post-	• <i>Materials:</i> Videotapes of SU interviews. • <i>Theoretical modality:</i> Psychodynamic. • <i>Training:</i> Clinicians received three training sessions and were given feedback on their written formulations from the authors of the study. • <i>Procedure:</i> Each SU participated in an unstructured psychodynamic interview which was videotaped. • The interviewers focused on the SU’s pattern of psychosocial functioning in relation to their: a) relationships with others, b) occupation, c) leisure activities, and d) other important aspects of their life. • The Ideographic Conflict Formulation (ICF) is a method used to formulate an individual’s psychodynamic conflicts. • Clinician raters watched the videotaped interviews, then collectively they discussed and formulated each	• Type A correctly matched formulation pairs, had higher mean similarity ratings than the mismatched type B and type C formulation components. • The type A mean overall similarity of correctly ‘matched’ formulation was reported as 4.4 (63% agreement). • Mean ratings for each of the four core components, were scored above 4, meaning, they were rated somewhere between ‘equal amount of overlap and non-overlap’ and ‘definite overlap with some distinct disagreements or differences in

doctoral experience. Total $n = 2$. • <i>Raters (referred to as similarity raters):</i> Clinical psychology graduate students grouped into two independent teams of two. Total $n = 4$.	SU's conflicts, using the ICF method. • Written ICFs were four to eight pages in length, with clinician raters required to list evidence in the data to support each assertion. • For the next part of the study, a second independent formulation was written by two pairs of clinicians who also viewed the video-taped interviews and produced written ICFs. They were blind to all other information. • Independent raters then rated the similarity of each pair of ICFs ($n = 20$ pairs in total) on a seven-point Likert scale from 'essentially similar' to 'essential dissimilar'. • The ICFs were divided into three possible types: 1) Type A formulations were correctly 'matched', and expected to be similar, 2) Type B formulations were correctly 'matched' for sex and diagnosis, but less similar on some conflict elements, 3) Type C formulations were 'mismatched' for subject and diagnosis, but correctly matched for sex. These formulations were expected to have the lowest similarity scores.	description'. Therefore, on average the formulations were rated as definitely similar, but with some distinct areas of differences in description. • In summary, results indicated that two independent psychodynamic formulations of the same SU, were rated as more similar than dissimilar. Whereas mismatched formulations were rated as more dissimilar than similar. • These findings indicate modest but substantial evidence for reliability of the ICF method for constructing psychodynamic formulations.
---	---	---

		• The main analysis focused on pairs of raters making similarity ratings on pairs of ICFs, based on the following four core components: 1) wish, 2) fear, 3) resultants, and 4) an overall rating of the core components.	
29. Persons & Bertagnolli (1999) USA	• <i>SU/diagnosis formulated:</i> One SU (that served as the practice/training case) met diagnostic criteria for major depression and GAD. A second SU (pseudonym 'Megan') met diagnostic criteria for major depression, dysthymia, and PD NOS. A third SU (pseudonym 'Lisa') met diagnostic criteria for dependent PD, and avoidant PD (total $n = 3$). • <i>Interviewer:</i> The first author of this study ($n = 1$); no demographic information	• <i>Materials:</i> An audiotape (12-minutes) of an initial session with the practice/training case. Audiotapes (35-minutes each) and typed transcripts of the first two initial sessions of therapy with 'Megan' and 'Lisa'. Multiple choice 15-item schema questionnaires. Demographic questionnaires. • <i>Theoretical modality:</i> Cognitive behavioural. • <i>Training:</i> Clinicians received two hours of training (conducted by the first author). This involved teaching on cognitive behavioural CF, then listening to a 12-minute audiotape of an initial session with the practice/training case and practising the CF measures as per the procedures described below. • <i>Procedure:</i> Two initial sessions of therapy (also conducted by the first author) with the SUs 'Megan' and 'Lisa' were audio recorded and the interviews	• Interrater reliability of the clinicians' problem lists was reported as high (two random checks resulted in 83% and 93% agreement). • Clinicians correctly identified 67% of SUs' overt presenting problems. • PhD trained clinicians were more accurate than non-PhD trained clinicians in identifying SU's overt problems. • When schema ratings were averaged over five judges, ICCs ranged from .44 to .91. However, single judges showed poor interrater agreement on schema ratings, ICCs

provided. • <i>Formulators (referred to as clinicians)</i>: A mix of health professionals (e.g. clinicians with a PhD = 45%, clinicians with an MA, or MSW = 45%, students = 13%). Most clinicians (63%) had previously received training in CF, prior to participating in the study. Total $n = 47$. • <i>Benchmark experts</i>: The authors of the study ($n = 2$), and a graduate student ($n = 1$). Total $n = 3$; no other demographic information provided. • <i>Raters</i>: The authors of the study (total $n = 2$); no demographic information provided.	were transcribed. • Clinicians were asked to independently listen to the audio recordings/read the transcripts of the sessions with ‘Megan’ and ‘Lisa’, then identify each SU’s 1) overt presenting problems (maximum of eight problems per case, based on a specific list of problem domains developed by Nezu & Nezu, 1993), and 2) underlying schemas associated with those problems. • Clinicians completed a multiple-choice questionnaire in relation to the SU’s view of self, the world, and others. • They were asked to rate the strength of each schema for each SU on a scale of 0 = ‘no belief’, to 10 = ‘very strong belief’. • Clinicians then rated the schemas for each SU a) within two specific contexts (one context was provided for ‘Megan’, and another context was provided for ‘Lisa’. No context was provided for the practice case), and b) in the absence of any context (e.g. a public speaking situation commonly activates ‘self’ schemas about inadequacy and ‘other’ schemas about criticism).	ranged from .13 to .66. •These findings indicated that as the number of judges upon which a rating is based increased (from one to five), the reliability of the ratings increased. • Providing clinicians with specific contexts in which to make the schema ratings, did not improve the reliability of schema agreement. • No demographic training or clinical experience variables predicted clinicians’ tendency to agree with the other clinicians on the schema ratings.
---	---	---

	• <i>Judges:</i> Described as ‘clinicians who provided the ratings’ (total $n = 5$); no demographic information provided.	• Clinicians completed a demographic questionnaire in relation to their training and clinical experience. • Expert ‘benchmark’ problem lists were provided for ‘Megan’ and ‘Lisa’ from which the clinicians’ problem lists could be rated for comparison. Raters gave the clinicians a score of one if their problem list included a problem identified in the expert benchmark list, and a score of 0 if it did not. • Data from all three SUs (including the practice/training case) were included in the final analysis.	
30. Persons et al. (1995) USA	• <i>SU/diagnosis formulated:</i> SUs diagnosed with anxiety and depression. Total $n = 2$. • <i>Formulators (referred to as clinicians):</i> A mix of health professionals (e.g. clinicians with a PhD = 57%, clinicians with an MA or MSW $n = 39\%$, students = 4%). Total $n = 46$.	• <i>Materials:</i> An audiotape (12-minutes) of an initial session with the first SU (practice/training case). An audiotape of a full initial session of therapy with the second SU. Multiple choice (61-item) questionnaire. Demographic questionnaires. • <i>Theoretical modality:</i> Cognitive behavioural. • <i>Training:</i> Half a day’s training was provided to the clinicians (conducted by the first author). This involved receiving teaching on cognitive behavioural CF, then listening to a 12-minute audiotape of an initial	• Interrater reliability of the clinician’s problem lists was reported as 96% agreement for the first SU and 89% agreement for the second SU. • Percentage agreement for clinicians correctly identifying overt presenting problems (in relation to the ‘benchmark’) ranged from 13% to

• <i>‘Benchmark’ experts:</i> Both experts (total $n = 2$) were authors of the study. One expert had written extensively about CF and had treated both SUs. The other expert had eight years’ experience of conducting cognitive therapy and training cognitive therapists. • <i>Raters:</i> The authors of the study ($n = 2$); demographic information provided above (see section titled ‘benchmark experts’). • <i>Judges:</i> Described as ‘clinicians who provided ratings’ ($n = 5$); no further demographic information provided.	session with the practice/training case and practising the CF measures as per the procedures described below. Then, the first author provided some feedback about the practice case, and the CF. • <i>Procedure:</i> Clinicians were asked to independently listen to an audio recording of a full initial CBT session (also conducted by the first author) in relation to the second SU. • Clinicians were then asked to identify the SU’s 1) overt presenting problems (maximum of six problems per case), 2) underlying cognitive mechanisms associated with the presenting problems from a predefined list (i.e. schemas in relation to the SU’s view of self, the world, and others), and 3) conditional beliefs or dysfunctional attitudes (i.e. “if...then” statements). • Expert ‘benchmark’ problem lists were provided for each SU, from which the clinicians problem lists could be rated for comparison. Raters gave the clinicians a score of 1 if their problem list included a problem	98% for the first SU, and 67% to 100% for the second SU. • When schema ratings and dysfunctional attitudes were averaged over five judges, interrater reliability was high (except for one rating of dysfunctional attitudes for the second SU, where the ICC was .27). ICCs ranged from .27 to .92. • However, a single randomly chosen judge showed poor interrater agreement. ICCs ranged from .07 to .70. • These findings indicated that as the number of judges increased (from one to five), the reliability of the ratings increased. • No demographic training variables predicted clinicians’ ability to reach agreement on the identification of 1) overt presenting problems, 2)
--	--	---

		identified in the expert benchmark list, and 0 if it did not. • Clinicians completed a demographic questionnaire in relation to their previous training in CF and CBT. • Data from both SUs (including the practice/training case) were included in the final analysis.	schemas, or 3) dysfunctional attitudes.
31. Rosenberg et al. (1986) USA	• <i>SU/diagnosis formulated:</i> SUs diagnosed with neurosis and PDs. Total $n = 5$. • <i>Formulators (referred to as clinicians):</i> ‘Experienced’ clinicians that held a cognitive-psychoanalytic orientation to psychotherapy. Total $n = 4$-5. • <i>Raters (referred to as reliability judges):</i> All raters were described as having a cognitive-psychoanalytic orientation to psychotherapy (total $n = 4$); no additional	• <i>Materials:</i> Verbatim transcripts of each SU’s intake interview, and the first two hours of their psychotherapy sessions. A nine-point Likert scale. • <i>Theoretical modality:</i> Cognitive-psychoanalytic. • <i>Training:</i> No training was provided as part of the study. • <i>Procedure:</i> SUs received brief psychodynamic psychotherapy (16-sessions). • A team of clinicians read the interview and therapy session transcripts, then worked together to develop a written ‘plan formulation’ for each SU. • ‘Plan formulations’ contain the SU’s presenting problems and history, as well as four main components: 1) goals for therapy, 2) obstacles preventing the attainment of goals, 3) tests that may be	• To assess the degree of agreement between the clinicians and reliability judges, Pearson’s correlations between the average team ratings (for each of the four components) were compared. • Findings showed good agreement between the two teams of judges. Pearson’s correlations ranged from $r = .62$ to $r = .97$ across all five SUs.

	demographic information provided.	used in therapy to work with unhelpful beliefs, and 4) insights that would be helpful to the SU. • Each component (i.e. goal, obstruction, test, and insight) was then reduced to a list of 15-30 idiosyncratic statements. These relevant statements were then combined with a comparable number of alternative a) less relevant (but plausible) statements, and b) irrelevant statements, for each SU. • The statements for each component were then rated independently by the clinicians in terms of their relevance to each SU on a nine-point Likert scale from 0 = 'not relevant' to 8 = 'highly relevant'. • A team of reliability judges also read the interview and therapy session transcripts, then independently rated the relevance of each statement (in relation to each SU), on the nine-point Likert scale.	
32. Shefler & Tishby (1998) Israel	• <i>SU/diagnosis formulated:</i> SUs either had no diagnosis on Axis I of the DSM III-R ($n = 5$), or they were diagnosed with an anxiety	• <i>Materials:</i> The Accuracy Rating Scale (developed for this study). Transcribed intake interviews ($n = 15$); two 50-minute sessions per SU. • <i>Theoretical modality:</i> Psychoanalytic. • <i>Training:</i> Judges received training which involved	• Interrater reliability for each case (based on ratings of three judges) were calculated using ICCs (Rosenthal, 1987).

disorder ($n = 3$), depressive disorder ($n = 2$), or an adjustment disorder ($n = 5$). On Axis II, SUs either had no diagnosis ($n = 14$), or a diagnosis of avoidant PD ($n = 1$). Total $n = 15$. • <i>Formulators (referred to as clinicians)</i>: A mix of senior mental health professionals (e.g. clinical psychologists $n = 5$, social workers $n = 3$, and a senior psychiatrist $n = 1$). Clinicians were psychoanalytically trained, with between seven to 16 years of clinical experience. They had also participated in 12-months training in Time Limited Psychotherapy (TLP; Mann, 1973; Mann &	didactic teaching, as well as practical exercises in relation to the formulation of a SU's central issue (CI). • <i>Procedure</i>: The SUs each received 12-weekly sessions of TLP. Their intake interviews were audio recorded, then transcribed. • The intake transcripts for each SU were read by three clinicians who were then asked to independently formulate each SU's CI, then produce a second less relevant (but plausible) CI. • Each CI was two to three sentences long and consisted of four components: 1) SU strengths, achievements, and areas of coping, 2) incorporation of longitudinal aspects of the SU's suffering, 3) awareness of the SU's emotional pain, and 4) the SU's relation to oneself. • All formulations were checked by the first author for adherence to the TLP model. • Using the Accuracy Rating Scale, judges (who were blind to the therapy outcomes) also read the intake interview transcripts and were asked to rate a) the	• The mean ICC for all 15 cases ranged from $R = .13$ to $R = .85$. • To investigate whether agreement between raters for each case was greater than could be expected by chance, the Lawlis and Lu (1972) chi-square test was calculated. In this test, agreement was defined as a discrepancy of no more than 1-point between judges. In 10 out of 15 cases interrater agreement was greater than would be expected by chance. • In sum, judges agreed on approximately two thirds of the CIs. • For the four CI components, judges agreed on these ratings in 10 out of 15 cases.
--	--	--

Goldman, 1982). Total $n = 9$. • <i>Raters (referred to as judges)</i>: A mix of experienced mental health professionals (e.g. psychologists $n = 10$, and social workers $n = 5$. Judges had all participated in a four-year seminar on TLP. Total $n = 15$.	accuracy (or goodness-of-fit) of the CI, as well as b) the four components of the CI (described above). • Judges provided one global rating, and four component ratings on a scale from 1 = ‘highly inaccurate’ to 5 = ‘highly accurate’. • Judges were provided with three CIs per case. One CI ‘matched’ the case, and two additional CIs purposefully ‘mismatched’ the case.
--	---

Table 2.5 Methods used to assess CF reliability across all $n = 32$ studies included in the review

Method	Total n of studies (and study numbers)
<i>As agreed by the raters:</i>	
1) can clinicians independently construct similar formulations based on the same set of clinical material (i.e. assuming they are using the same model, can therapists agree with the conceptualisation of a given case?).	19 (1, 2, 3, 5, 6, 7, 8, 9, 14, 18, 19, 20, 21, 22, 24, 25, 26, 29 and 30)
2) can clinicians independently construct formulations, similar to a gold standard ‘benchmark’ formulation (i.e. constructed by a single expert, or team of experts?).	6 (4, 12, 23, 25, 29 and 30)
3) can clinicians agree that an already constructed formulation/its components ‘match’ a set of clinical material (or ‘match’ another formulation/its components) that have been constructed for the same SU?	10 (7, 15, 16, 17, 21, 24, 27, 28, 31 and 32)
4) can clinicians formulate the same set of clinical material in a similar way, at two different time points? (i.e. referred to as test-retest reliability).	2 (16 and 20)

5) can raters deconstruct/segment the CFs (produced by the clinicians) and agree about the “idea units” in the CF? (i.e. so that the formulation can be content coded to assess for quality).	3 (10, 11 and 13)
---	-------------------

2.6.4 Reliability measurement

High levels of reliability reflect excellent levels of agreement between clinicians, and this may be measured in different ways (Grant et al., 2010). Most studies in this review ($n = 26$) used one measure of reliability. However, $n = 6$ studies used two (2, 10, 25, 29 and 30), or three (20) measures. Often this was owing to studies having adopted several different *methods* from which to assess reliability (see section 2.6.3 above).

The most common reliability index used (across $n = 13$ studies; 2, 8, 15, 16, 17, 20, 21, 24, 26, 27, 29, 30 and 32) was the ICC. The second most common measurement was Cohen's kappa coefficient (k ; Cohen, 1960; 1, 2, 5, 8 and 25), Fleiss' kappa (Fleiss, 1981; 6, 10, 11 and 13), or Cohen's weighted kappa (K_w ; Cohen, 1968; 18 and 21) used by a total of $n = 11$ studies. This was followed by percentage rate agreement, which was also used by $n = 11$ studies (3, 4, 10, 12, 20, 22, 23, 25, 28, 29 and 30). Five studies used either Pearson's correlation coefficient (r ; 9, 14, 20 and 31), or Spearman's Rho correlation coefficient (ρ ; 19).

The ICC is the most appropriate measure of reliability for continuous data (Shrout & Fleiss, 1979; Bartko, 1991). This can be used to estimate interrater reliability of any number of formulators/raters (Mumma, 2011). For categorical data, k is the most appropriate reliability measurement between two formulators/raters, whereas Fleiss' kappa is more appropriate for three or more formulators/raters (McHugh, 2012). K_w allows for different 'weights' to be allocated to different levels of agreement between formulators/raters. As explained by Tallberg et al. (2020) "...*weighted kappa is [also] more permissive of varying measures, and therefore as a rule, presents higher kappa values*" (p. 5).

Paradoxically, unreliable measures of reliability for continuous data include the Pearson and Spearman correlation coefficients. This is because they measure the linear association between raters, *not* the level of agreement (Bartko, 1991). Correlation coefficients such as Pearson's r may therefore poorly reflect the amount of agreement between formulators/raters resulting in extreme over (or under) estimates of the true level of agreement (Stemler, 2004).

Historically, for categorical data, interrater reliability was measured using percentage rate agreement as this is easily calculated and interpreted (McHugh, 2012). However, it has since been argued that percent agreement is an inappropriate measure of

reliability, because it does not consider the possibility of chance – that raters might accurately guess some scores, owing to uncertainty (Bartko, 1991; McHugh, 2012). Percent agreement may therefore overestimate the ‘true’ rate of agreement amongst formulators/raters, whereas kappa and ICC are more reliable chance-corrected agreement statistics (McHugh, 2012).

There are no universal guidelines outlining ‘acceptable’ levels required to ascertain reliability (Graham et al., 2012; Koo & Li, 2016). Nonetheless, commonly cited thresholds within the field of psychology have been proposed by Cicchetti (1994), and Cicchetti and Sparrow (1981). Their guidance has been used to interpret levels of ICC and kappa (including weighted kappa, and Fleiss’ kappa). The guidance states that when the reliability coefficient is below .40, the level of clinical significance is ‘poor’; when it is between .40 and .59 the level of clinical significance is ‘fair’; when it is between .60 and .74 the level is ‘good’; and when it is between .75 and 1.0 the level of clinical significance is deemed to be ‘excellent’ (Cicchetti, 1994; Cicchetti & Sparrow, 1981). Whilst this guidance was helpful for evaluating (and synthesising) data across all 32 studies, it does imply that a score of .41 might be acceptable. It is helpful to note that McHugh (2012) has argued that this is too lenient for healthcare/clinical research proposing more conservative cut-offs for the kappa statistic, stating: *“any kappa below 0.60 indicates inadequate agreement among the raters and little confidence should be placed in the study results”* (McHugh, 2012, p. 279).

Psychology guidelines proposed by Dancey & Reidy (2020) were also followed for the interpretation of Pearson and Spearman’s correlation coefficients. They suggest that if the value is near ± 1 , then it indicates a ‘perfect’ correlation (i.e. as one variable increases, the other variable tends to also increase [if positive] or decrease [if negative]). If the correlation value lies between $\pm .70$ and $.90$, then it is said to be a ‘strong’ correlation. If the value lies between $\pm .40$ and $.60$, then it is a ‘medium’ correlation. If the value lies between $\pm .10$ and $.30$, then it is said to be a ‘weak’ correlation, and when the value is 0, then it is said to be an inverse or ‘zero’ correlation (Dancey & Reidy, 2020).

For percentage agreement, guidance by Luborsky and Diguier (1998) was followed, in line with Flinn et al. (2015). Luborsky & Diguier (1998) have proposed that 70% or higher indicates ‘good’ reliability. However, others have suggested that anything over 75% (Hartmann, 1977; Stemler, 2004), or 80% (McHugh, 2012)

demonstrates ‘acceptable’ levels of agreement. As McHugh’s reminds us, when agreement is less than 80%, over 20% of the data are unreliable (McHugh, 2012). To answer the research question posed by the current review, the reliability findings from all 32 studies will be considered in relation to these levels of agreement.

2.6.5 Study location

Most studies ($n = 20$; 60%) were conducted in the USA. Five studies were conducted in the UK (4, 6, 9, 23 and 25), two studies were conducted in Norway (7 and 8), and one study was conducted in each of the following countries: New Zealand *and* Australia (2), Germany (3), Australia (5), New Zealand (12), and Israel (32).

2.6.6 Theoretical modality

Fourteen studies formulated from a cognitive (14, 15, 16, 17, 23, 25 and 27), or cognitive behavioural (2, 6, 9, 12, 26, 29 and 30) modality. Nine studies formulated from a psychoanalytic (19 and 32), or psychodynamic (5, 7, 8, 18, 21, 22 and 28) modality. Six studies used an integrative approach (e.g. cognitive-psychoanalytic; 1, 20, 24 and 31), or a combination of different models (e.g. psychodynamic, *and* cognitive behavioural; 10 and 11). Three studies were atheoretical (e.g. four Ps framework; 4 and 13; or plan analysis; 3).

2.6.7 Participant demographics

The 32 studies included in this review, used data from a total of 525 SUs, and between 948 and 949 formulators/raters. The exact number of formulators/raters is not known, as one study (31) described their formulator participant numbers within a range of 4 to 5. Collectively, the studies formulated SUs with a wide range of presenting issues. The most prevalent diagnoses fell into the category of neurosis (e.g. GAD, MDD), followed by PDs (e.g. NPD, BPD). Only two studies formulated psychosis (4 and 23).

Most studies ($n = 22$) outlined demographic information in relation to their formulators and raters. This included providing details about their clinical experience

and professional role. However, $n = 10$ studies (3, 4, 9, 14, 16, 17, 23, 25, 29, and 31) had missing demographic information in relation to their formulators and/or raters. For example, formulators were referred to as ‘highly experienced clinicians’, and/or formulators/raters were referred to as ‘the research team’ (i.e. authors of the study) without providing any further information.

2.6.8 Training

Half of the studies ($n = 16$) reported that training was offered to the formulators and/or raters as part of their study (1, 2, 6, 7, 9, 10, 11, 13, 16, 20, 21, 25, 28, 29, 30 and 32). Whereas $n = 3$ studies referred to formulator/rater training having taken place *prior* to their participation in the study (e.g. postgraduate training in CBT, or training in the CF method as part of their professional role; 18, 20¹¹ and 23). The remaining studies ($n = 14$) did not refer to having trained their formulators/raters (3, 4, 5, 8, 12, 14, 15, 17, 19, 22, 24, 26, 27 and 31).

Across the studies included in this review, training did not appear to improve reliability scores. Studies that yielded the lowest levels of reliability, either *had* offered training (29, 30 and 32), or had *not* offered training (8 and 19) to their formulators/raters (see Table 2.6). For example, the lowest reliability values across all studies were reported as .07 (in relation to a study where formulators/raters *had* received training; 30), and .02 (in relation to a study where formulators/raters had *not* received training; 19). In contrast, studies that yielded the highest levels of reliability, also *had* either offered training (13, 16 and 25), or had *not* offered training (15 and 31) to their formulators/raters. For example, one of the highest reliability values was reported as .97 in relation to studies where formulators/raters *had* received training (16 and 25). A reliability value of .97 was also reported in relation to several studies where formulators/raters had not received training (15 and 31).

Whilst there does not seem to be a general association between training and CF reliability, individual studies did appear to demonstrate some value in training. For example, a perfect reliability value of 1.0 was reported (at the upper end of their range) in relation to study 13, where clinicians were randomly assigned to either receive training in the CFCCM (Eells et al., 1998, 2005), or not receive the CFCCM training

¹¹ For study 20, some of the clinicians were trained as part of the study, whereas some of the clinicians had already been trained prior to their participation in the study.

(control group). Findings revealed that the training group clinicians produced CFs that were more elaborate, comprehensive, complex, and precise than the control group. Training group clinicians were also more likely to identify precipitants, predisposing factors, and an inferred mechanism to explain SU symptoms and problems, whereas control group clinicians tended to summarise descriptive data (see Kendjelic & Eells, 2007). Despite this, a perfect reliability value of 1.0 was also reported (at the upper end of their range) in relation to study 18, where clinicians had *not* received training.

2.6.9 Formulator/rater participant sample

Some studies ($n = 12$) used qualified clinician formulators/raters to determine the reliability of CFs (4, 5, 6, 7, 8, 9, 18, 19, 20, 21, 22, 31 and 32). Several studies ($n = 18$) used varying combinations of qualified clinicians, graduate-level trainee therapists/psychologists, undergraduate/postgraduate students, SUs, and RAs (1, 2, 10, 11, 12, 13, 14, 15, 16, 17, 23, 24, 25, 26, 27, 28, 29 and 30). In addition, one study (16) referred to the formulators as ‘authors of the study’. However, no demographic information was provided. Similarly, two studies (3 and 17) only provided the n in relation to their formulators/raters. No further demographic information was provided, apart from diagnostic information in relation to a SU rater (who acted as a rater) to triangulate (and determine) the reliability of their own CF components (17).

As indicated in Table 2.6 below, study 1 found ‘good to excellent’ rates of agreement were achieved between the raters in relation to two groups of formulators: 1) undergraduate RAs and 2) graduate-level IRT trainees. They concluded that the IRT CF method was reliable amongst formulators with a range of clinical experience (see Critchfield et al., 2015). However, study 4 found that qualified clinicians (e.g. psychologists) with experience of psychosis, indicted a low rate of agreement (22%) with the expert benchmark formulation (see Hartley et al., 2016). Study 7 found that CFs (written by experienced clinicians, without any specific structure or labelling) could be reliably rated by experienced, as well as less experienced clinicians (see Sørbye et al., 2019). Study 10 found that experts produced higher quality CFs than novice and experienced therapists, regardless of SU diagnosis or theoretical orientation (see Eells et al., 2005). Study 11 found that experts generated more forward reasoning (measured by the frequency of description-to-inference links in the CFs), than novice and experienced therapists (see Eells et al., 2011). Study 12 found that ‘novice’ CBT

therapists (i.e. clinicians that had recently graduated with a postgraduate diploma in CBT) constructed CFs that had a ‘good enough’ grasp of the fundamental CBT components of the CF. However, broad omissions between the therapists and the expert ‘benchmark’ CFs lead to very poor agreement for certain components (e.g. ‘Axis II PD traits’; see Haarhoff et al., 2011). Study 14 reported a ‘moderate to strong’ rate of agreement between two clinicians’ CFs (a novice vs an expert), that focused on ICSs. However, the expert clinician’s CF was substantially better in helping to explain the SU’s distress (see Mumma & Mooney, 2007). Study 23 found that a greater number of years of clinical experience amongst clinicians, improved reliability/agreement with the expert benchmark formulation overall. Moreover, experience was also related to the reliability of clinicians’ ability to formulate inferential aspects of the case (see Dudley et al., 2010). Study 25 found that in comparison to BABCP accredited formulators, pre-qualification student formulators were least likely to identify an important aspect of the benchmark formulation. However, interestingly, for a small number of the inferential aspects of the formulation (e.g. core beliefs, compensatory strategies) pre-qualification students demonstrated a higher percentage rate of agreement (when compared to the benchmark formulation), in comparison to the BABCP accredited formulators (see Kuyken et al., 2005). Finally, study 29 found that PhD trained clinicians were more reliable than non-PhD trained clinicians, in identifying SU’s overt problems. However, training, and clinical experience did not predict clinicians’ tendency to agree with one another, in relation to schema ratings, when formulating each case (see Persons & Bertagnolli, 1999).

2.6.10 Formulation data

Most studies provided the formulators with one source of information from which to construct their formulations. This included: a) written transcripts (e.g. of early therapy sessions, and/or intake evaluations/interviews; 5, 13, 18, 19, 21, 24, 29, 31, 32), b) fictional written case vignettes (12), c) audio recordings (e.g. of initial therapy sessions, early CF sessions; 3, 6, 7, 8, 10, 11, 22, 27 and 30), d) video recordings (e.g. of therapy sessions, CF interviews, semi-structured assessment interviews, intake evaluation/interviews; 1, 2, 4, 9, 14, 16, 17, 22, 26 and 28), and e) case material (e.g. a referral letter, completed assessment measures, psychosocial history, a timeline detailing the SU’s history; 23 and 25). One study did not state the format in which their case material was provided (20). Finally, study 15 reported that the formulator had

based their CF on ‘two assessment interviews’ with the SU, and although not explicitly stated, it may be inferred that these assessments were completed face-to-face, with the presence of a second clinician (rater) observing.

Six studies provided the formulators with two sources of information (e.g. fictional written case vignettes alongside an audio [or video] recording of the vignette; 4, 10, 11, 23, 25 and 29). This may be seen to increase ecological validity, by reflecting the realities of clinical practice. However, using two sources of information did not appear to increase levels of reliability in comparison to other studies included in this review. For example, study 4 only achieved 22% agreement, and the lower range for percent agreement in study 25, fell to 7%). This may be because factors that may have served to decrease ecological validity were also present, such as the use of fictional case vignettes (4, 10 and 11), and having an actor role-play a SU (23 and 25).

Several studies ($n = 6$) asked raters to identify and score CF components from: a) predefined lists of standardised categories (e.g. using the CCRT and CCRT-LU methods; 5, 8, 18 and 21), b) predefined lists of overt presenting problems (e.g. using a specific list of problem domains; 29), or c) a predefined list of underlying cognitive mechanisms (30). Whilst it may be hypothesised that higher rates of reliability would be more likely with a systematic approach to CF, the results from Flinn et al. (2015) and the current review do not necessarily support this. Reliability values for these six studies ranged from ‘poor to excellent’ (29 and 30) for cognitive behavioural formulations, and ‘fair to good’ (5), ‘poor to good’ (8), ‘fair to excellent’ (18), and ‘good’ (21) for psychodynamic formulations.

2.6.11 Blinding

Studies reported that their formulators either worked entirely independently ($n = 20$; 2, 3, 4, 5, 6, 7, 8, 12, 13, 14, 15, 16, 17, 21, 23, 25, 27, 29, 30 and 32), or as part of an independent team ($n = 12$; 1, 9, 10, 11, 18, 19, 20, 22, 24, 26, 28 and 31). Most studies ($n = 27$) also reported that the CFs (produced by the formulators) were assessed by independent raters. However, for a small number of studies ($n = 6$; 8, 15, 16, 17 and 27) the formulators also acted as raters. For $n = 4$ studies (15, 16, 17 and 27) this related to study designs whereby the interviewer, SU, and formulator’s ratings were triangulated to assess the reliability of self-scenarios, or interpersonal scenarios. These scenarios

were designed to either be relevant (or less relevant) to the SU. Finally, several studies ($n = 11$) referred to blinding procedures (1, 6, 7, 9, 10, 11, 13, 20, 21, 28 and 32) to protect against bias. For example, the raters had no prior knowledge of the cases, or they were blind to group membership (i.e. they had no prior knowledge as to whether the formulation had been produced by a novice, or expert therapist).

2.6.12 Overall key findings

This systematic literature review yielded an additional 17 studies, to provide an update and extension of the systematic searches carried out by Flinn et al. in 2014. Results from the current review demonstrated a wide *range* of reliability values (both within, and across studies). The overall findings (across all 32 studies [and all five reliability methods]) demonstrated coefficient values that ranged from ‘weak to perfect’ (.10 to 1.0), and ‘poor to excellent’ (<.40 to 1.0), hence scores were yielded at either end of each scale (depending upon the interpretive measure of reliability that was used, i.e. ICC and kappa; Cicchetti, 1994; Cicchetti & Sparrow, 1981; Pearson’s and Spearman’s correlation coefficients; Dancey & Reidy, 2020). A range of values were often reported within individual studies, because these values related to the reliability of different components of CF. This reflected the range of reliability values also reported by Flinn et al. (2015).

Multiple variables meant that it was hard to synthesise the literature, and like-for-like comparisons were contraindicated owing to: a) the range of methods used to assess CF reliability, b) the range of reliability measures used (e.g. Cohen’s kappa, Pearson’s correlation), c) the range of theoretical modalities (e.g. CBT, psychodynamic), d) the range of disorders, e) the range of training provided to formulators/raters as part of the study, f) the range of experience and training of the formulators/raters, prior to their participation in the study, and g) the range of ways in which the formulation data was provided to formulators, from which to construct their formulations (e.g. written transcripts, videos).

The range and values of reliability (on each of these measures) across 27 studies can be seen in Table 2.6. The full range of percent agreement scores across 12 studies can be seen in Table 2.7. Owing to some studies having used more than one method from which to assess reliability, some studies appear twice in Table 2.6, or the study

appears in both Tables 2.6 and 2.7, alongside the corresponding reliability method that was used. Based on the number of studies (and the different methods used to determine reliability) there was a greater number of cognitive/cognitive behavioural studies/methods used ($n = 17$), followed by integrative/combined ($n = 11$), psychoanalytic/psychodynamic ($n = 10$), and atheoretical ($n = 3$).

The greatest number of studies ($n = 9$) yielded 'good to excellent' levels of reliability (1, 2, 6, 7, 10, 11, 13, 24 and 25), followed by $n = 7$ studies that yielded 'excellent' levels of reliability (7, 15, 16, 20, 21, 26 and 27). Three studies demonstrated 'poor to excellent' levels of reliability (29, 30 and 32), with $n = 2$ studies demonstrating 'fair to excellent' levels (17 and 18). Single studies yielded 'strong to perfect' (20), 'strong' (9), 'good' (21), 'moderate to perfect' (31), 'moderate to strong' (14), 'fair to good' (5), 'poor to good' (8), and 'weak to perfect' (19) levels of reliability. The full range of percent agreement scores for the remaining 5 studies/methods can be seen in Table 2.7. Again, these varied widely, for example, from '10% to 100%' agreement (25 and 30). One study (12) reported their percent agreement scores as between <30% to <50%. However, they did not report the specific values, hence they are reported in Table 2.7 within the range of '0% to 49%'. Tables 2.6 and 2.7 also indicated that the method used to determine reliability, did not have a bearing on the level of reliability that was yielded (i.e. each method 1-5, appear to have yielded a wide range of reliability scores).

Despite the mixed findings, it may be noted that over half of the studies (including the different reliability methods used; 21/31; 67.7%) yielded levels of reliability that fell within a lower range of $\geq .60$; as per the McHugh (2012) recommendations for more conservative cut-offs applied to clinical research. However, only a small number of studies (2/11 studies; 18.2%) yielded percentage agreement scores that fell within a lower range of $\geq 80\%$ (again, as recommended by McHugh, 2012).

When considering the coefficient values across different theoretical modalities, three quarters of the cognitive/cognitive behavioural studies (including the different reliability methods used; 9/12 studies; 75%) yielded values that fell within a lower range of $\geq .60$. In comparison, less than half of psychoanalytic/psychodynamic studies (including the different reliability methods used; 3/8 studies; 37.5%) yielded lower values that fell within $\geq .60$. Just over three quarters of the integrative/combined studies

(including the different reliability methods used; 7/9 studies; 77.8%) yielded lower values that fell within $\geq .60$). Finally, in relation to the atheoretical modality, one study reported coefficient values that fell within a lower range of $\geq .60$ (1/1 studies; 100%).

When considering percentage agreement values across the different theoretical modalities, none of the cognitive/cognitive behavioural (0/5 studies; 0%), psychoanalytic/psychodynamic (0/2 studies; 0%), or atheoretical studies (0/2 studies; 0%) yielded percentage agreement scores that fell within a lower range of $\geq 80\%$. Meaning, that in general, they were unreliable, as per the more stringent level of guidance proposed by McHugh (2012). In contrast, the integrative/combined studies (of which there were only two studies that reported percentage agreement), *did* yield percentage agreement scores that fell within a range of $\geq 80\%$ (2/2 studies; 100%).

When considering the different theoretical modalities, atheoretical formulations, followed by integrative/combined formulations, appeared to generate greater levels of reliability (however, only three atheoretical studies, and six integrative/combined studies were included in the current review), so this may have had a bearing on the findings. Cognitive/cognitive behavioural formulations also generated greater levels of reliability than psychoanalytic/psychodynamic formulations. This differs from the overall finding reported by Flinn et al. (2015), who found psychodynamic formulations to have greater levels of reliability in contrast to the other theoretical modalities.

Considering the integrative/combined studies that yielded lower values that fell within $\geq .60$, or $\geq 80\%$ (1, 10, 11, 20 and 24), and the atheoretical study that yielded a lower value that fell within $\geq .60$ (13), then the next step was to consider whether the higher rates of reliability were associated with the methodology quality of these studies? As can be seen from Table 2.1, study number one produced a total quality score of 14 (out of 15) points; studies 10, 11 and 20 scored 13 points; study 13 scored 12 points; and study 24 scored 10 points. Whilst these scores are relatively high, there were also another three studies in the review (6, 7, and 9) that scored 14 points for their methodological quality assessment, and another 2 studies that scored 13 points (21 and 32), but their reliability scores fell below the thresholds outlined by McHugh (2012). Hence, there does not seem to be a common denominator as to what factor(s) appear to be associated with improved levels of CF reliability.

Across theoretical modalities, when reliability scores were pooled over three (6, 10, 11, 13, 15 and 27), four (16, 17, 19 and 26), four to five (31), or five (20 and 29) judges/raters, a large proportion of studies (including the different reliability methods used; 12/15 studies; 80%) yielded levels of reliability within a lower range of $\geq .60$. Similarly, over half of the studies (including the different reliability methods used; 2/3 studies; 66.7%) yielded percentage agreement scores within a lower range of $\geq 80\%$.

Overall, this systemic review of 32 studies yielded variable findings. The findings do not appear to be associated with any of the five methods employed, used to determine CF reliability, nor the training and experience of those that formulated/conducted the ratings. The many conceptual and methodological differences across studies should be noted. Given the level of reliability (across and between studies) varying from almost no agreement to 100% agreement, the review does not lend itself to a robust meta-analytic approach. A close narrative synthesis of studies grouped in terms of their theoretical modality are therefore presented below.

Table 2.6 The range of reliability scores (and method(s) used) across all studies included in the review

Interpretation - reliability range	Value range ¹²	Study numbers (and reliability method(s) used)	Total n of studies
Overall			
<i>Excellent</i>	.75 to 1.0	7 (3), 15 (3), 16 (3 and 4), 20 (1), 21 (3), 26 (1), 27 (3)	8
<i>Strong to perfect</i>	.70 to 1.0	20 (4)	1
<i>Strong</i>	.70 to .80	9 (1)	1
<i>Good to excellent</i>	.60 to 1.0	1 (1), 2 (1), 6 (1), 7 (1), 10 (5), 11 (5), 13 (5), 24 (1 and 3), 25 (1)	10
<i>Good</i>	.60 to .74	21 (1)	1
<i>Moderate to perfect</i>	.40 to 1.0	31 (3)	1
<i>Fair to excellent</i>	.40 to 1.0	17 (3), 18 (1)	2
<i>Moderate to strong</i>	.40 to .90	14 (1)	1
<i>Fair to good</i>	.40 to .74	5 (1)	1
<i>Poor to excellent</i>	<.40 to 1.0	29 (1), 30 (1), 32 (3)	3
<i>Poor to good</i>	<.40 to .74	8 (1)	1
<i>Weak to perfect</i>	.10 to 1.0	19 (1)	1

¹² Value ranges based on the interpretation of kappa and ICC (according to Cicchetti, 1994; Cicchetti & Sparrow, 1981), and the interpretation of Pearson's and Spearman's correlation coefficients (according to Dancey & Reidy, 2020).

Cognitive/cognitive behavioural			
<i>Excellent</i>	.75 to 1.0	15 (3), 16 (3 and 4), 26 (1), 27 (3)	5
<i>Strong</i>	.70 to .80	9 (1)	1
<i>Good to excellent</i>	.60 to 1.0	2 (1), 6 (1), 25 (1)	3
<i>Fair to excellent</i>	.40 to 1.0	17 (3)	1
<i>Poor to excellent</i>	<.40 to 1.0	29 (1), 30 (1)	2
Psychoanalytic/psychodynamic			
<i>Excellent</i>	.75 to 1.0	7 (3), 21 (3)	2
<i>Good to excellent</i>	.60 to 1.0	7 (1)	1
<i>Good</i>	.60 to .74	21 (1)	1
<i>Fair to excellent</i>	.40 to 1.0	18 (1)	1
<i>Fair to good</i>	.40 to .74	5 (1)	1
<i>Poor to excellent</i>	<.40 to 1.0	32 (3)	1
<i>Poor to good</i>	<.40 to .74	8 (1)	1
<i>Weak to perfect</i>	.10 to 1.0	19 (1)	1
Integrative/combined			
<i>Excellent</i>	.75 to 1.0	20 (1)	1
<i>Strong to perfect</i>	.70 to 1.0	20 (4)	1
<i>Good to excellent</i>	.60 to 1.0	1 (1), 10 (5), 11 (5), 24 (1 and 3)	5

<i>Moderate to perfect</i>	.40 to 1.0	31 (3)	1
<i>Moderate to strong</i>	.40 to .90	14 (1)	1
<i>Atheoretical</i>			
<i>Good to excellent</i>	.60 to 1.0	13 (5)	1

Note. Reliability method(s) used: (1) Clinicians being asked to independently construct similar formulations based on the same set of clinical material. (2) Clinicians being asked to construct formulations, which were then compared to a gold standard ‘benchmark’ formulation (i.e. constructed by a single expert, or team of experts). (3) Clinicians being asked to determine whether a constructed formulation/its components ‘match’ a set of clinical material (or ‘match’ another formulation/its components) that have been constructed for the same SU. (4) Clinicians being asked to formulate the same set of clinical material in a similar way, at two different time points (i.e. test-retest reliability). (5) Raters being asked to deconstruct/segment the CFs (produced by clinicians) to determine the ‘idea units’ in the CF.

Table 2.7 The range of percentage agreement scores (and method(s) used) across all studies included in the review

Percentage agreement	Study numbers (and reliability method(s) used)	Total n of studies
<i>Overall</i>		
80% to 100%	10 (5), 20 (4)	2
60% to 100%	3 (1), 28 (3), 29 (2)	3
50% to 59%	22 (1)	1
30% to 100%	23 (2)	1
10% to 100%	25 (2), 30 (2)	2
0% to 49%	4 (2), 12 (2)	2
<i>Cognitive/cognitive behavioural</i>		
60% to 100%	29 (2)	1
30% to 100%	23 (2)	1
10% to 100%	25 (2), 30 (2)	2
0% to 49%	12 (2)	1
<i>Psychoanalytic/psychodynamic</i>		
60% to 100%	28 (3)	1
50% to 59%	22 (1)	1

<i>Integrative/combined</i>		
80% to 100%	10 (5), 20 (4)	2
<i>Atheoretical</i>		
60% to 100%	3 (1)	1
0% to 49%	4 (2)	1

2.6.13 Cognitive/cognitive behavioural formulations

Fourteen studies formulated from a cognitive (14, 15, 16, 17, 23, 25 and 27), or cognitive behavioural (2, 6, 9, 12, 26, 29, and 30) modality. This was an additional seven cognitive/cognitive behavioural papers, since the Flinn et al. review. The reliability of cognitive/cognitive behavioural formulations mainly yielded ‘excellent’ levels of reliability using methods 1 (26), 3 (15, 16, and 27), and 4 (16). Using method 1, one study also yielded ‘strong’ levels of reliability (9), three studies yielded ‘good to excellent’ levels of reliability (2, 6, and 25), and two studies yielded ‘poor to excellent’ levels of reliability (29 and 30). Finally, using method 3, one study yielded ‘fair to excellent’ levels of reliability (17).

In terms of percentage agreement scores, the following studies utilised method 2 and yielded the following rates of agreement: <30% to <50% (12); 32% to 92% (23); 7% to 73% (25); 67% to 93% (29); and 13% to 100% (30). Although not accounting for agreement based on chance, studies that employed purely percentage agreement for all aspects of the formulation (12 and 23), found less than a third of their CF components met the (liberal) threshold of >70%, as proposed by Luborsky and Diguer (1998). Owing to the limitations associated with percentage agreement outlined above in section 2.6.4, it is likely that actual rates of agreement may be even less (Flinn et al., 2015).

Providing clinicians with specific contexts in which to make the schema ratings did not improve the reliability of schema agreement. Moreover, no demographic training or clinical experience variables predicted clinicians’ tendencies to agree with other clinicians on the schema ratings (Persons & Bertagnolli, 1999).

In general, there was higher agreement for the overt/descriptive components of cognitive/cognitive behavioural formulations (e.g. presenting issues, behaviours; 23, 25, 29, and 30) and less agreement for the theory based inferential components (e.g. core beliefs, dysfunctional assumptions; 23, 25, and 29). For example, study 25 involved 115 therapists independently formulating a case using J. Beck’s (1995) CCD. The clinicians’ formulations were compared with one another, and with a ‘benchmark’ formulation provided by J. Beck. Results indicated that clinicians could reliably agree with one another, and with the expert on the descriptive aspects of the case (i.e. rates of

agreement where high for e.g. presenting issues). However, rates of agreement were moderate for easy-to-infer information (e.g. compensatory strategies), and poor for difficult-to-infer information (e.g. dysfunctional assumptions; Kuyken et al., 2005). One exception to this was study number 30, which reported 'moderate' rates of agreement in relation to clinicians identifying 'overt problems', and except for one type of belief (dysfunctional attitudes) for one SU, they reported high ICCs in relation to underlying cognitive mechanisms (e.g. schema ratings) when averaged over five judges (mean ICC = .76). However, a single randomly chosen judge showed poor inter-rater agreement (mean ICC = .46; Persons et al., 1995).

One study (16) assessed the test-retest reliability of cognitive formulations (16). Findings yielded 'excellent' levels of test-retest reliability for 6 parameters across 29 therapy sessions, in relation to prototypical scenarios of how SUs interacted with others when at their 'best' and 'worst'. An average ICC of .95 was reported for when SUs felt at their 'best', and an average ICC of .91 was reported for when SUs felt at their 'worst' (Muran et al., 1998).

2.6.14 *Psychoanalytic/psychodynamic formulations*

Nine studies formulated from a psychoanalytic (19 and 32), or psychodynamic (5, 7, 8, 18, 21, 22 and 28) modality. This was an additional three psychoanalytic/psychodynamic papers, since the Flinn et al review. Two studies yielded 'excellent' levels of reliability (7 and 21) using method 3. However, another study (32) yielded 'poor to excellent' levels using method 3. Single studies that used method 1, yielded 'good to excellent' (7), 'good' (21), 'fair to excellent' (18), 'fair to good' (5), 'poor to good' (8), or 'weak to perfect' (19) levels of reliability, therefore, demonstrating wide variability. In terms of percentage rate agreement, study 28 used method 3 and yielded 63% agreement, and study 22 used method 1 and yielded 51% agreement.

Reflecting the findings reported in the reliability of cognitive/cognitive behavioural formulations, study 7 found higher rates of agreement in relation to the more overt/descriptive elements of the psychodynamic formulation (ICC = .82) with less agreement found in relation to the more inferential elements (ICC = .61). Interestingly, the reliability of experienced psychodynamic formulators yielded an ICC of .79, and for those formulators who had no prior experience in practising

psychodynamic psychotherapy, the reliability of their ratings was even higher with an ICC of .91.

It should be noted that scores in studies 19 and 21 were pooled over four judges/raters (in each study) which Flinn et al. (2015) suggests may have served to inflate reliability. Interestingly, study 19 found that when formulators were given verbatim therapy transcripts to read (referred to as ‘textwise judges’), they surpassed the reliabilities obtained between the ‘textless’ (referred to the group that were *not* given the verbatim transcripts to read) and the ‘textwise’ formulators combined. This led the authors to conclude that text information is necessary to arrive at agreements that are more case-specific than mere stereotypes or ‘mannequin’ formulations (i.e. clinically stereotypical CFs that do not incorporate idiosyncratic features; Caston & Martin, 1993).

2.6.15 Integrative/combined formulations

Six studies either used an integrative approach to formulation (1, 20, 24 and 31), or a combination of different models (e.g. psychodynamic, *and* cognitive behavioural; 10 and 11) where formulators could construct a psychodynamic or cognitive behavioural CF for each case, depending on their psychological perspective. This was an additional four integrative/combined papers, since the Flinn et al review.

As suggested by Flinn et al. (2015), combining different theoretical modalities (e.g. cognitive-psychoanalytic; 24 and 31), may have inflated the reliability of CFs through theoretical bias. When ratings were pooled and averaged over several judges (e.g. five judges/raters were used in studies 20 and 31), correlations for integrative formulations demonstrated ‘good to excellent’ (1 and 24), ‘excellent’ (20), or ‘moderate to perfect’ (31) levels of agreement. However, when an average score was taken for a single judge, reliability appeared to be in the ‘poor to excellent’ range (31). This finding tentatively suggests that pooling the judges/raters scores may inflate reliability.

One study (20) assessed the test-retest reliability of integrative formulations 3-months apart. Findings yielded ‘strong to perfect’ coefficients ($r = .94$ to $.98$) for mean Time 1 and Time 2 ratings for Case A (comparable ratings could not be made by the authors for Case B owing to time constraints; Collins & Messer, 1991). To verify these findings in a different way, percentages of Time-1-selected items retained at Time 2 were calculated. Percentage rate agreements indicated high levels of stability with 85%

to 97% of Case A items, and 90% to 96% of Case B items retained from Time 1 to Time 2 (i.e. at the 3-month follow-up; Collins & Messer, 1991).

For studies 10 and 11 that used a combined theoretical approach, ‘good to excellent’ levels of reliability were yielded for both ‘descriptive’ ($\kappa = .61$) and ‘inferential’ ($\kappa = .62$) components of the CF (using method 5; Eells et al., 2005; Eells et al., 2011). This was coupled with high percentage rates of agreement (80% to 100%) when judges were asked to segment the formulation transcripts into small “idea units” (again, using method 5; Eells et al., 2005).

2.6.16 *Atheoretical formulations*

Only three studies used an atheoretical formulation, therefore comparisons are limited. These three atheoretical papers were all an addition, since the publication of Flinn et al. For the two studies that used the four Ps framework, study 4 demonstrated a low percentage rate of agreement (22%) when clinician CFs were compared to an expert benchmark formulation (method 2; Hartley et al., 2016). In contrast, study 13 demonstrated ‘good to excellent’ levels of reliability (.60 to 1.0) when the raters were required to agree on the “idea units” (method 5) using the CFCCM. Ratings had also been pooled across multi-raters (Kendjelic & Eells, 2007). Interestingly, in study 13 a similar finding was highlighted (as found using CFs with other modalities) in that the ‘training’ group clinicians were more likely than the ‘control’ group clinicians to identify an inferred mechanism to explain SU difficulties. Whereas the ‘control’ group clinicians tended to summarise descriptive data (Kendjelic & Eells, 2007).

In relation to the one study that used plan analysis as its atheoretical framework (3), the authors described a ‘sufficient’ level of interrater reliability (60.5%) using method 1 (Grandjean et al. 2021). However, this did not meet the threshold for acceptable levels of percentage rate agreement (as explained in section 2.6.4 above).

2.7 Discussion

This systematic review with a narrative synthesis has updated and extended the Flinn et al. review, investigating the reliability of CFs across different models and disorders. Moreover, this was the first systematic review of CF to differentiate between type of reliability *methods* used when reporting the reliability statistics – an important aspect of the literature which has seemingly been overlooked in previous reviews (see Easden & Kazantzis, 2018; Flinn et al., 2015; Rainforth & Laurenson, 2014).

The current systematic review extended the literature search from 2014 onwards, to avoid duplication of the original work published by Flinn et al. This yielded an additional nine studies, including one study (9) which was identified via the trawling of reference lists and which pre-dated 2014, but was missed by Flinn et al. An additional eight studies were also identified from a review of the CBT CF literature by Easden and Kazantzis (2018). Again, these eight studies pre-dated 2014, but had been missed by Flinn et al. Given the disparate set of literature, which at times was hard to decipher (i.e. there were difficulties disentangling and extracting some of the relevant information), then it is perhaps unsurprising that these studies were missed. Moreover, as previously mentioned, CF reliability was reported as an outcome, but not always the main objective/primary focus for some of these studies; hence Flinn et al. may be forgiven for this understandable oversight.

Overall, the broad range of reliability scores (both within, and across studies) is indicative that studies are reporting poor to high levels of agreement. Meaning, that certain CF components (or aspects of the formulation) were more reliably formulated by clinicians, than others. For example, as previously mentioned, overt/descriptive components were more reliable than inferential components (Dudley et al., 2010; Kuyken et al., 2005; Persons et al., 1995; Persons & Bertagnolli, 1999).

Overall, the findings suggest that atheoretical formulations, followed by integrative/combined formulations, appeared to generate greater levels of reliability. However, only three atheoretical studies, and six integrative/combined studies were included in the current review, so this may have had a bearing on the findings. One hypothesis for higher rates of agreement in relation to atheoretical formulations (e.g. four Ps), may be owing to the framework/model being generic and simpler. Consequently, reliability may be enhanced with formulations that a) have broader categories (i.e. with wider scope to ‘capture’ many components), and b) that are not

overly complicated. Regarding integrative approaches (e.g. interpersonal reconstructive; 1), the researchers/clinicians were the pioneers and developers of the model. Undoubtedly, they will have been invested in this newer model. Therefore, as suggested by Flinn et al. researchers/clinicians from an integrative modality are likely to have had a stronger theoretical bias, which may have inflated levels of reliability. Moreover, some of the models where two theories are integrated (e.g. cognitive-psychoanalytic), may enable two-three related aspects of the formulation to be captured in multiple places (e.g. appraisals and thoughts, may also be captured under the heading of 'reciprocal-roles'). So, perhaps, there are multiple places where an integrative CF component can be scored. Meaning, (like the four Ps framework and the point mentioned above), these models more accommodating/suited to producing higher levels of reliability.

For the combined studies (which were published by the same first author; 10 and 11) the method, and reliability findings were the same across both studies. Each of these studies compared novice, experienced, and expert formulators from either a psychodynamic, or cognitive-behavioural perspective, and they were asked to formulate each case in relation to the CFCCM.

The results also showed that cognitive/cognitive behavioural formulations generated greater levels of reliability than psychoanalytic/psychodynamic formulations. This differs from the overall finding reported by Flinn et al., who found psychodynamic formulations to have greater levels of reliability in contrast to the other theoretical modalities. However, perhaps it is less about whether we can say one modality is better than another in terms of rate of agreement, it is more the case that we can say that the studies do not show that formulation is totally reliable, nor is it totally unreliable.

Findings from the methodological quality assessment, demonstrated wide variability in the quality assessment ratings across five different parameters, with a mean total quality score not dissimilar to that reported by Flinn et al. This is despite minor discrepancies found between some of the individual quality assessment ratings reported by Flinn et al., and some of the rating of these items, reported here.

2.7.1 *Methodological limitations of the studies*

One methodological limitation concerning the studies included in this review, relates to the use of students (24), or undergraduate RAs (1) as formulators, when their demographic information (e.g. training, and clinical experience) was *not* used for the purposes of comparison (i.e. with more experienced formulators, of which this may have been appropriate; 25, 29 and 30). As mentioned in Chapter 1, CF is recognised as a core skill within the profession of clinical psychology (Beck, 2020; DCP, 2010, 2011; Eells, 2022). However, there is less emphasis (if indeed any) in the use of CF in undergraduate level psychology. Therefore, it is unlikely that students and RAs have the skill set to formulate competently. As Flinn et al. (2015) stated: “*it is therefore questionable at what level a psychology student or graduate could formulate*” (p. 285). This is especially important when thinking about CF as it moves from more descriptive to explanatory levels, which requires much greater theory-based inference (Kuyken et al., 2009). Indeed, formulating inferential aspects of a case calls for higher-level training and clinical expertise, with the reliability (and quality) of CBT CFs deteriorating as therapists move from descriptive to inferential levels of formulation (with several studies offering support for this view, e.g. Dudley et al., 2010; Eells et al., 2005; Kendjelic & Eells, 2007; Kuyken et al., 2005). Despite some studies specifically implementing CF training to the formulators as part of their research (1, 2, 6, 9, 13, 21, 25, 28, 29, 30 and 32), it is questionable as to whether relatively brief training in the process of CF can be comparable to clinicians with years of clinical experience in a particular therapeutic modality (Flinn et al., 2015).

Another methodological limitation relates to the use of written transcripts, written fictional case vignettes, and audio recordings (in the absence of videos) for the construction of CFs. These sources of information may be seen as less ecological valid, owing to the omission of para/non-verbal language. Given that approximately 60-65% of interpersonal communication is non-verbal (Burgoon et al., 2009), this would seem important when formulating clinical cases. However, the use of video recordings did not seem to substantially increase rates of reliability. Three quarters of the studies (including the different reliability methods used; 6/8; 75%) yielding levels of reliability that fell within a lower range of $\geq .60$. However, none of the studies (0/3; 0%) yielded percentage agreement scores that fell within a lower range of $\geq 80\%$.

A final limitation noted across studies, relates to the inconsistent use of terminology across studies, which meant that reading and extracting relevant information was a challenging part of this review process. For example, “raters” were sometimes referred to as “judges”, and “formulators” may have been referred to as “clinician raters”, “third-party observers”, or “CCRT judges”, with a separate set of clinicians/therapists who had also delivered therapy.

2.7.2 Clinical implications

When clinicians are asked to formulate the same case, some of the literature points to low/poor levels of reliability (e.g. Persons et al., 1995). This lack of consensus suggests that clinicians may hold different views on the essential components of a CF, and what it should contain, even when they are formulating from the same perspective (Butler, 2006; Eells et al., 1998; Bieling & Kuyken, 2003). Chapter 3 attempts to address consensus in CF when making sense of psychosis. Nevertheless, it is argued that (even for the same case), several formulations could be ‘valid’. Butler (2006) argues that low levels of reliability are inevitable because there is no single ‘right’ or ‘correct’ way to formulate: *“I think it is wrong to use the language of ‘correctness’ when talking about formulations – which is not the same as saying that a formulation could not be obviously wrong”* (Butler, 2006, p. 11). Failure to find consistently good evidence of reliability (except for the descriptive components of a CF; Bieling & Kuyken, 2003) may be because: 1) the emphasis for understanding and treatment may be focused in different places particularly if the case is complex with multiple presenting issues, 2) owing to the different levels of CF, some clinicians may decide to focus their formulation on maintaining factors in the ‘here and now’, whereas other therapists may prefer to start with a longitudinal formulation first, focusing on early life experiences and underlying core beliefs/schemata, and 3) Therapists differ from each other in different ways (e.g. age, gender, cultural background, life experiences), and so they are likely to approach their CFs from different angles, and to see things in different ways (Butler, 2006).

In training, supervision, and practice, it is possible to formulate the case material in more than one way (Butler, 2006). As Flinn et al. (2015) have pointed out, the literature suggests that CF is both a ‘science’ (that emphasises replicability, reliability, etc.), and an ‘art’ (that emphasises an idiographic approach, that is outside the purview

of scientific investigation) and, so on what basis do we look to assess it – its reliability? or its value? Perhaps instead, we should be asking – ‘which formulation will be more useful?’ and ‘what core ingredients/components should we seek to include?’.

Another clinical implication is that reliability is not the same as *quality* (Bieling & Kuyken, 2003). Even in studies where there is a high degree of agreement amongst clinicians, this does not necessarily mean that they are producing good quality formulations (Kuyken et al., 2009). Establishing a way of ensuring clinicians can reliably formulate the same case (using the same model) may be of limited value if the formulation is incoherent and unhelpful in planning treatment/interventions.

Consequently: “...it may be more important for a formulation to be a coherent and justifiable account of a person’s presenting problems than for it to be replicable” (Bieling & Kuyken, 2003, p. 191). It is also likely that a CF for the same SU would differ between therapists of different psychotherapeutic modalities. However, both CFs may be of good quality in generating verifiable hypotheses, and having good treatment utility (Bieling & Kuyken, 2003). So, reliability is perhaps not the only ‘marker’ we should be attending to. CFs and their relationship to treatment outcome, have also been argued as a better marker by which to assess CF, and this is investigated in Chapter 4.

It is argued that one possible explanation for the poor levels of reliability in relation to inferential aspects of a formulation may be owing to cognitive shortcuts that clinicians unknowingly take, such as availability and anchoring heuristics (Flinn et al., 2015; Kuyken et al., 2005; Kuyken, 2006). These become more likely when problems are particularly complex, therapists are inexperienced, and when there are competing demands (Garb, 1998). For example, mental shortcuts may be used by therapists to speed up the process of finding a satisfactory solution. Heuristics include using trial and error, a rule of thumb or an educated guess. They explain how people make come to make decisions, use their judgement, and solve problems. These rules typically come into play when people face complex problems (e.g. there is no optimal solution), or there is incomplete information, which can work well under some circumstances, but can lead to systematic errors or cognitive biases (Gigerenzer, 1991). Heuristic errors that affect clinical decision making can lead to erroneous CFs.

Heuristics are an essential part of clinical decision making, yet it is important to identify and remedy problematic heuristics to improve the reliability of CF (Kuyken et al., 2009). Therapists need to ask themselves whether they are basing their CFs on

hunches or habits, rather than on theoretically based hypotheses (Butler, 2006). Reflective practice both in therapy and supervision can boost awareness of these heuristics to help avoid such problems (Kuyken et al., 2009). For example, therapists benefit from explicitly developing written CFs and sharing these with SUs and supervisors, a process that focuses and engages the rational decision-making system that is a counterbalance to intuitive decision making (Kahneman, 2003). Writing CFs in session, in therapy case notes, and for supervision purposes encourages a therapist to make intuitive processes *explicit*. The process of writing out the CF often puts a spotlight on gaps in understanding or inconsistencies in thinking. This also seems relevant in relation to the tentative signal that pooling the scores of formulators/raters provides higher levels of CF reliability. This finding was also highlighted by Flinn et al. (2015) and is one that adds further weight to the importance of clinical supervision in constructing CFs (Cromarty & Gallagher, 2023), and the benefits of team formulation. For example, by clarifying hypotheses (DCP, 2011; Johnstone, 2014b; Short et al., 2019), suggesting that two (or more) heads are better than one.

2.7.3 Limitations, and recommendations for future research¹³

The CF reliability findings indicate a high degree of heterogeneity between the study designs, including a variety of different methods (see Table 2.5), and measurements used to assess CF reliability (see section 2.6.4). Therefore (as highlighted by Flinn et al.), like-for-like comparisons were contraindicated. One of the limitations of this review is that it included all five methods from which researchers used to assess CF reliability. This may explain why the range of reliability values are so variable (i.e. ranging from weak/poor, to excellent/perfect). Future systematic reviews in this area could seek to focus on *one* (or perhaps two) method(s) of CF reliability (see Table 2.5) to reduce heterogeneity.

As referred to in section 2.6.4 several measurements are deemed inappropriate for the assessment of reliability, and the consequences of unreliable measurements leads to poor quality research (see Bartko, 1991). Therefore, a recommendation for future

¹³ Flinn et al. (2015) outlined six recommendations for future research. These recommendations do still apply, therefore (owing to the word count restrictions of this thesis) H.M.S. decided not to duplicate those recommendations here. Instead, the main recommendations provided in this chapter, relate to the conduct of future systematic reviews in this area.

reviews would be to exclude studies that had used ‘unreliable’ measures of reliability (e.g. Pearson and Spearman correlation coefficients, and percent agreement).

Whilst it would have been interesting to consider the levels of CF reliability across different disorders, this was not possible owing to time and thesis word count restrictions. This is a clear limitation of the current review, hence, is an area recommended for future research. Indeed, it should be noted that the reliability for CFs in psychosis, for example, (4 and 23) did not yield agreement scores that fell within a lower range of $\geq 80\%$; meaning, they were unreliable.

Whilst Flinn et al. have suggested that pooling the scores of formulators/raters may serve to artificially inflate levels of reliability, it is still an interesting signal, and one whereby future research studies could look to explore the level of CF reliability produced by a team of clinicians, versus a single clinician, or compare levels of CF reliability produced by different sized teams. This would help clarify the question, as to whether two (or more) heads are better than one?

2.8 Conclusion

This systematic review with a narrative synthesis updated and extended the findings of Flinn et al. (2015). Owing to the disparate set of literature it was not possible to conduct a meta-analysis to produce a precise estimate of the reliability value/rate of agreement, across all the studies. Instead, a close narrative synthesis of studies was undertaken. This indicated that formulations constructed using an atheoretical modality (e.g. four Ps) produced higher levels of reliability. The assessment of methodological quality indicated that the quality of the papers was not associated with CF reliability. Nevertheless, the quality of studies has improved over recent years. In summary, the current review suggests that the reliability of CFs is still an under-researched area for a process crowned (over twenty years ago) as: *“the heart of evidenced-based practice”* (Bieling & Kuyken, 2003, p. 53). The reliability of CF requires much further investigation.

Chapter 3. What are the essential ingredients of a CBT case formulation for auditory hallucinations and persecutory delusions in schizophrenia spectrum disorders? Using the Delphi method to establish an expert consensus

“Science is really about individual experts reaching a consensus”.

– Alan Stern (2017, para. 8).

The main findings outlined in the following thesis chapter have been published in Schizophrenia Research¹⁴ and the reference for the publication is:

Spencer, H. M., Dudley, R., Freeston, M. H., & Turkington, D. (2020). What are the essential ingredients of a CBT case conceptualization for voices and delusions in schizophrenia spectrum disorders? A study of expert consensus. *Schizophrenia Research*, 224, 74-81. <https://doi.org/10.1016/j.schres.2020.09.026>

This research was also presented at a conference symposium and the reference for the abstract is:

Spencer, H. M., Turkington D., Freeston, M., & Dudley, R. (2021, May 6 - 28). *Essential ingredients of CBTp case formulation* [Paper presentation]. Aaron T. Beck Annual CBT for Psychosis Conference 2021, Amsterdam, The Netherlands.

¹⁴ Impact Factor at the time of publication was 4.93.

3.1 Abstract

Evidence supports the use of cognitive behavioural therapy (CBT) for the treatment of schizophrenia spectrum disorders. A CF (or case conceptualisation) is seen as the keystone of CBT in terms of making sense of a person's difficulties, to guide and inform such treatment. Despite the importance placed on CF there is no known consensus amongst experts as to the essential ingredients involved in this fundamental process. This study used the Delphi method to establish expert consensus for the components of a CF when working to treat auditory hallucinations, and persecutory delusions. An international panel of 78 CBT for psychosis (CBTp) experts from 12 different countries participated in the main stage of this study. This 3-stage process involved producing and rating statements that addressed key areas of CF in terms of: presenting issues, predisposing, precipitating, perpetuating, and protective factors. One presenting issue and 7 perpetuating factors were endorsed as essential by >80% of the expert panel. The exact same items were endorsed for both auditory hallucinations, and persecutory delusions. The findings are unique in that a large panel of international experts reached consensus that case formulations (CFs) should be parsimonious and focused on the perpetuating (maintaining) factors to facilitate change. Overall, the proposed recommendations should lead to core guidance for the process of developing CFs, and improvements in training for clinicians that conceptualise auditory hallucinations, and persecutory delusions in CBT for schizophrenia spectrum disorders.

3.2 Introduction

An estimated 70-80% of SUs with a diagnosis of schizophrenia report auditory hallucinations (Hugdahl, 2015). For those experiencing a FEP, over 70% report persecutory delusions which are characterised by unfounded suspicion and mistrust of others (Coid et al., 2013). Given the prevalence of these distressing and disabling symptoms (which often co-occur) treatment and research has typically focused on voice hearing and paranoia.

CBTp helps to reduce distressing symptoms such as voice hearing and persecutory beliefs (Bighelli et al., 2018), and whilst evidence for the use of CBT in treating a wide range of psychological disorders is well established (Butler et al., 2006), findings from meta-analyses suggest that CBTp produces smaller treatment effects. For example, a network meta-analysis showed that CBT produced a small to medium effect

size for the treatment of positive symptoms in schizophrenia, in comparison to treatment-as-usual (standardised mean difference, *SDM* = -0.30; 95% confidence interval, *CI* = -0.45 to -0.14; *n* = 1,508; *n* = 18 RCTs; Bighelli et al., 2018). Whereas a large effect size was found in a network meta-analysis of face-to-face CBT for depression, compared to treatment-as-usual (standardised difference in mean change, *sDIMC* = -1.11; 95% credible interval, *CrI* = -1.62 to -0.60; *n* = 2,169; *n* = 19 RCTs; López-López et al., 2019). Meta-analyses such as these suggest that CBT is less effective when working with more complex presentations such as psychosis, and it is likely that there are multiple reasons to account for these differences.

As explained in Chapter 1, the role and value of CF has been identified as an important one. However, one reason to account for the differences in treatment effects, may be owing to the vast array of psychological models that may be drawn upon when treating psychosis. Models of psychosis are also not as advanced, therefore the formulations arising from these models may be incomplete or lacking the specificity that is a key feature of models for anxiety and depression. Another reason may be owing to the nature of psychosis itself. For example, individuals with psychosis often have a comorbid diagnosis of anxiety and depression (Wilson et al., 2020), and a history of childhood trauma (Varese et al., 2012). There is also substantial evidence that people with psychosis have difficulties in a wider range of areas of their lives, in comparison to those with common mental health difficulties. For example, individuals with psychosis are more likely to have dropped out of education, to be unemployed, to experience poverty, debt, social isolation, and to have poorer physical health (Ashworth et al., 2017; Marwaha et al., 2007; Meltzer et al., 2002; Mind, 2008). They are also less likely to form romantic relationships (Redmond et al., 2010), and to manage tasks that are necessary for independent living (Holloway, 2005). Protective factors (such as employment) may be included in a formulation (Kuyken et al., 2009). However, people with psychosis often have less of these. Consequently, they may not benefit as much from CBT, as their mental health is more severe and complex to treat.

In mental health services, the ability to be able to provide effective CBTp treatment is further reduced owing to a limited number of therapists, meaning cognitive therapists need to work efficiently to meet demand. The independent Schizophrenia Commission established that only 1 in 10 who might benefit from CBT, have access to it (Schizophrenia Commission, 2012). Data reported from the second National Audit of Schizophrenia (NAS2) audited 100 SUs living in the community from each of the

Mental Health Trusts/Health Boards in England and Wales. The findings from this report identified significant gaps in the provision of CBT, with 39% of SUs having been offered it (the range was 14% to 67% across all Trusts/Health Boards), and only 18% of SUs reported they had received it (Royal College of Psychiatrists, 2014). Another study has argued that only 5-10% of SUs with psychosis in the UK, receive CBTp (Haddock et al., 2014). Clearly the number of people that have access to (and receive) evidence based psychological intervention remains low, despite NICE guidance which recommend that 1:1 CBTp is offered to all people with schizophrenia (NICE, 2014). Given the shortfall in provision, an important question is: how can we make CBTp more efficient with the limited time therapists have? It has been argued that learning to use a CF effectively, is one of the biggest challenges CBT therapists face (Kuyken et al., 2009), meaning any improvement in the efficiency of this process, will likely help increase the quality of therapy provided.

Viewed as the keystone of CBT, a CF helps cognitive therapists to work efficiently by integrating a SU's unique experience within a psychological theory or model, to understand the origins, development, and maintenance of the problem (Tarrier & Calam, 2002). It is seen as the link between theory, research, and practice (Butler, 1998), as well as a collaborative, ever evolving process, whereby the therapist and SU work together first to describe, and then to make sense of, a person's difficulties. This 'sense making' process, can be particularly helpful in psychosis, when trying to understand what at first might appear to be bizarre, nonsensical, or perplexing symptoms, which is often a strange and frightening experience for the individual (Spencer, 2019).

A CF-driven approach to treatment recognises the individuality of each person, so that those presenting with the same symptoms/diagnosis will have distinctive, ideographic formulations that have personal meaning for them (Persons et al., 2006). Nevertheless, there may be some common features or core ingredients of a CF, that are essential to include.

The primary function of a CF is to guide and inform therapy, to relieve a person's distress, and to build resilience (Kuyken et al., 2009), and several authors have argued that this process needs to be as parsimonious as possible (Butler, 1998; Dudley et al., 2015; Kuyken et al., 2009; Kuyken et al., 2016; Morrison, 2007; Padesky et al., 2011). Kuyken et al. (2016) stated that "*parsimony in the conceptualization helps the client understand his/her presenting issues and use the conceptualization for change*"

(p. 182), and Kuyken et al. (2009) stated “*the best conceptualizations are as simple as possible without losing essential meaning*” (p.139). Similarly, Butler (1998) outlined parsimony as one of three key principles that guide the practice of CF, stating that:

it is always better, and more useful, to keep the formulation as simple as possible. The temptation to elaborate a formulation is strong, especially when dealing with complex cases. However, the simpler and clearer it is the more readily will its implications be seen and the easier it will be to use (p. 4).

Indeed, the Collaborative Case Conceptualization Rating Scale (CCC-RS; which has found to be a reliable measure with adequate face, content, and convergent validity for assessing therapist competence in the use of CF (see Kuyken et al., 2016)) rates item 4 on the basis that the CF is as simple as possible. The highest number of points (a score of a 3) is awarded for a conceptualisation that “*conveys complex ideas concisely, distilling information into the essential parts necessary to describe or explain what is needed at this stage of therapy for this client*” (Padesky et al., 2011, p. 14). Whilst the therapist may hold a complex conceptualisation in mind:

in the session this is distilled into a simple and functional conceptualisation.... if the conceptualisation is too complex there may be evidence the client does not understand, feels overwhelmed, or is humouring the therapist.... what is simplistic and obvious to one, may take another client several sessions to assimilate (Padesky et al., 2011, p. 14).

As such, the therapist must use their clinical judgement to pitch the level of the CF to the individual. Furthermore, the CF should be presented in a succinct way that organises information clearly, particularly if the CF is represented within a diagram (Spencer et al., 2023).

As explained in Chapter 1, the five Ps framework (Dudley & Kuyken, 2014) is a commonly used framework to facilitate the construction of CFs by eliciting and organising information in terms of: 1. Presenting issues (the key problems or difficulties the person is experiencing at that particular moment in time), 2. Predisposing factors (those factors which make the individual vulnerable to the problem, and therefore at greater risk for developing psychosis), 3. Precipitating factors (events that happened in the person's life that preceded/triggered the onset of psychosis), 4. Perpetuating factors (the things that keep a problem going, to maintain the psychosis), and 5. Protective factors (skills, strengths, resources, supports, or coping strategies, that have helped the

person to deal more effectively with stressful life events, to help mitigate the impact of psychosis; Dudley & Kuyken, 2014).

In CBT, there are also a variety of models specifically for voice hearing (for example, Kinderman, 2011; Morrison, 1998), and persecutory delusions (for example, Freeman et al., 2002; Turkington et al., 2011) which help guide the conceptualisation and treatment of individual psychotic symptoms. Whilst these models are well argued, they often have varying levels of empirical support. As such, there can be confusion for the therapist as to which model is best, and how to marry the SU's individual experiences *within* the chosen model (rather than forcing the SU's experiences to fit the model). Different models also emphasise different processes. For example, in relation to paranoia and persecutory beliefs, roles have been identified for 'jumping to conclusions' (Dudley et al., 2011), worry (Startup et al., 2016), rumination (Lebert et al., 2021), sleep (Waite et al., 2016), and safety behaviours (Freeman et al., 2007). Whilst incorporating these evidence-based components may be desirable, their inclusion risks cluttering the formulation, making it cumbersome and less focused on change. Formulations that reflect multiple components can therefore be complex, confusing, or even counterproductive. An investigation into the essential ingredients of a CF (from the many models available), is arguably an important first step towards clarity, refinement, and parsimony.

In mental health services, the ability to formulate is considered a regular feature of daily practice (Sperry et al., 1992), despite being a high-level clinical skill (Dudley et al., 2015; Kuyken et al., 2005) that is often under-taught (Ben-Aron & McCormick, 1980). CF as a process, is also under researched (Sperry et al., 1992) and the evidence for its impact on treatment outcome is also limited. For example, as detailed in Chapter 4, the use of a longitudinal formulation in the early stages of CBTp, led to poorer treatment effects (Spencer et al., 2018). Nevertheless, a recent meta-analysis of CBTp (Turner et al., 2020) reported that formulation-based treatment was more effective than non-formulation-based treatment for hallucinations, but not for delusions. This suggests that CBT CFs may contribute to the treatment of psychosis, but this is clearly a complex process.

To assist with the process of developing CFs, The BPS developed good practice guidelines (DCP, 2011), and several rating scales have also been designed to evaluate the content and quality of CFs (Eells et al., 1998; Haarhoff et al., 2011; Kuyken et al., 2005; Padesky et al., 2011). Whilst many of these scales are specific to CBT, others are

designed for use across different psychotherapeutic modalities, to facilitate treatment for a wide range of psychological disorders. As such, there are no standardised CBT guidelines that are tailored to the CF of individual psychotic symptoms (such as voices, or persecutory delusions).

Perhaps therefore, it is unsurprising to find variability in the content of clinician's CFs (Dudley et al., 2010; Kuyken et al., 2005). In part this may be owing to the fact there is no known consensus as to the essential ingredients or core components involved in this fundamental process in CBTp. Indeed, it has been suggested that one reason why CFs have not been studied more, is because of a lack of consensus as to what a CF should contain (Eells et al., 1998). What a formulation 'contains' is surely of great importance, and it would be helpful to understand what essential components therapists listen out for, as they gather information from the SU to develop the formulation. How do therapists know what to include (or what to omit)? If cognitive therapists cannot agree on the content of a CF, then it is unlikely they will agree on the best treatment plan (Dudley et al., 2015), and treatment outcomes are likely to differ.

Consequently, there are many reasons why consensus with regards to the content of CFs could be useful. A clear framework from which CBTp clinicians could refer to, would be incredibly useful in terms of providing 'best practice guidelines' with regards to the 'core' content of a CF for the treatment of particular psychotic symptoms. Butler (1998) argued that use of the word 'core' suggests that formulations may be thought to have certain basic elements which should be identified, as the process of formulation *"might be facilitated if there was agreement over which were the basic elements of a formulation and an atheoretical way of linking them together"* (Butler, 1998, p.18). Furthermore, a CF that contains these basic (essential) elements, would be more likely to lead to consistency of understanding, and of treatment (because practice can vary widely, as differences of opinion exist).

Whilst not trying to restrict therapists from making sense of a person's unique experience, a CF based on a consensus view would minimise the opportunity for problematic decision making that might compromise validity (Kuyken et al., 2005). It may be argued that no 'one size fits all', and that CF is a non-replicable process (Dudley et al., 2010). Nevertheless, standardised empirically supported guidelines help to ensure a 'quality control check' on the process of CF, and the therapy we provide.

The Delphi method is one way of achieving such consensus, and several studies have utilised this methodology in relation to the understanding and treatment of

psychosis (see Byrne & Morrison, 2014; Langlands et al., 2008; Law & Morrison, 2014; Marshall et al., 2004; Morrison & Barratt, 2010), the key components of a clinical psychology formulation (Thrower et al., 2024), factors that constitute a good CBT treatment manual (Duncan et al., 2004), as well as the key characteristics of a CF for those diagnosed with a personality disorder (Völlm, 2014). Nevertheless, the Delphi method has never been used to develop consensus for the key ingredients of a CF for the treatment of auditory hallucinations, or persecutory delusions in CBTp.

3.2.1 The Delphi method

The Delphi method was originally defined as: *“a method used to obtain the most reliable consensus of opinion of a group of experts by a series of intensive questionnaires interspersed with controlled feedback”* (Dalkey & Helmer, 1963, p. 458).

Formal consensus methods such as the Delphi method may be used to develop clinical guidance in health care settings where there is a) insufficient scientific evidence, b) lack of clarity, and/or where c) other research methods that provide higher levels of empirical evidence cannot be used (Humphrey-Murto & de Wit, 2019; Niederberger & Spranger, 2020). As explained by Niederberger and Spranger (2020): *“Delphi techniques have proven useful in the explorative or theoretical phase of the research process because they generate knowledge that can increase the evidence for the desired effect of an intervention - and thus possible insights into its potential effectiveness”* (p. 2).

Hence, the focus of the Delphi method lies where unanimity of opinion does not exist in the literature (Jones & Hunter, 1995) but where ‘experts’ are thought to hold the collective wisdom and knowledge to reach a consensus-based decision (Jünger et al., 2017; Thompson, 2009). Owing to this, it is argued there should be an explicit and clear rationale for its use as the most appropriate method (Jünger et al., 2017).

Ironically, there is little consensus with regards to how to undertake the method itself (Hasson et al., 2000), and this has resulted in considerable variation in terms of how the Delphi method is applied (Iqbal & Pilon-Young, 2009). As such, there are challenges when describing ‘best practices’ for each methodological step, within the Delphi process (Humphrey-Murto & de Wit, 2019).

Despite this, guidance on Conducting and REporting Delphi Studies (CREDES) have been developed to assess the methodological quality and robustness of Delphi

research (Jünger et al., 2017). CREDES guidance suggests that the rigorous conduct of the Delphi method requires high levels of quality in the study design, and transparency in reporting of the findings (Jünger et al., 2017).

Diamond et al. (2014) also identified 4 criteria considered to be key indicators of a high-quality Delphi study, namely 1) the selection criteria of participants should be reproducible, 2) the number of stages to be undertaken should be stated (e.g. particularly if the number of stages have been predetermined), 3) the criteria for dropping items at each stage should be made clear, and 4) the stopping criteria (other than number of stages) should be justified (e.g. a decision to stop based on consensus having been reached).

Modifications of the classic Delphi method often use a mixed methods approach, overlapping both positivist/quantitative and interpretative/qualitative methods via a structured, sequential, and interactive process (Hasson et al., 2000; Hsu & Sandford, 2007). Regardless of the modifications adopted, it is argued that “*the appropriate use of this technique requires a high degree of methodological precision and research rigour*” (Hasson et al., 2000, p. 1009) to instil confidence and trustworthiness in the findings.

The method relies on a panel of experts (with expertise in the chosen area of interest) answering surveys that are usually sent out electronically via email, or via the use of survey software (Brady, 2016). The use of multiple, successive stages is indicative of concurrent validity (Sharkey & Sharples, 2001), with the experts deciding on the importance of individual survey items with each iteration (Hasson et al., 2000). At the end of each stage, the expert panel is informed of the collective opinion, and there is an opportunity for the experts to reflect on their individual responses (and change them) as the collective opinion is repeatedly fed back; this process brings about the group consensus (Hasson et al., 2000; Iqbal & Papon-Young, 2009).

In stage 1, the traditional Delphi method invites the experts to generate all of the statements (or items; Keeney et al., 2011). However, it is argued that this can lead to the production of large amounts of qualitative data, that is time consuming for the researchers to analyse (Jenkins & Smith, 1994). As such, there is growing support for the researchers themselves providing the items to the experts in stage 1, by comprehensively extracting and synthesising the information, via a thorough review of the literature (Iqbal & Papon-Young, 2009; Keeney et al., 2011). This can act as a provisional starting point - making the consensus process easier by standardising the

knowledge base of the experts from the outset (Jacoby, 1988; Jünger et al., 2017; Murphy et al., 1998). Furthermore, the experts can arrive at justifiable, logical, and reliable decisions, based on the best empirical evidence available, alongside their experiential expertise (Jünger et al., 2017); this is viewed as a common and acceptable modification to the classic Delphi method (Hasson & Keeney, 2011; Hsu & Sandford, 2007).

Step-by-step guidance published by Iqbal & Papon-Young (2009) outline how to use the Delphi method within clinical psychology. This guidance recommends that stage 1 should incorporate an initial qualitative ‘idea generation stage’ whereby the experts are invited to review the items (provided by the researchers) and contribute any additional items (Iqbal & Papon-Young, 2009) or suggest the removal of any redundant items (Law & Morrison, 2014; Morrison & Barratt, 2010). Consensus is the aim of the Delphi method, so the more experts that mention an item in stage 1 signifies its importance for either inclusion or exclusion, in construction of the stage 2 survey (Brady, 2016). Pilot testing of the survey is recommended to reasonably predict the time frame for its completion, as well as checking the layout of the survey, and its readability (Hasson, 2000; Iqbal & Papon-Young, 2009; Jünger et al., 2017).

Stage 2 and subsequent stages are analysed quantitatively, with the items rated by the experts in terms of their perceived level of importance (Thangaratinam & Redman, 2005). Systematic reviews of the literature have found that 5- or 9-point Likert scales (Likert, 1932) are the most commonly used in Delphi research (Keeney et al., 2011; Niederberger & Spranger, 2020), and CREDES guidance recommend that the level of consensus should be predefined *a priori* (Jünger et al., 2017).

Different studies have predefined different levels of consensus. A systematic review of 100 Delphi studies found that the median threshold for consensus was 75% (range 50-97%). However, the preferred level of consensus for the majority of studies was set at >80% (Diamond et al., 2014), and a number of psychosis Delphi studies use the >80% level (see for example Byrne & Morrison, 2014; Langlands et al., 2008; Law & Morrison, 2014; Morrison & Barratt, 2010).

Many Delphi studies also predetermine their number of stages *a priori* (Diamond et al., 2014), primarily for pragmatic reasons (Thangaratinam & Redman, 2005). Humphrey-Murto and de Wit (2019) urge caution in continuing with multiple stages, as this can increase attrition rates owing to the time commitments involved. Also, those with opposing views may start to drop out if the process is too ‘drawn out’,

and this can lead to ‘false’ consensus. Too many stages can also ‘force’ the consensus, with experts succumbing to the majority opinion, in order to make the process end.

The classic Delphi method has 4 stages. However, it is argued that 3 stages are both sufficient and preferable to provide meaningful results that reduce attrition rates, and sample fatigue (Hasson et al., 2000; Iqbal & Pilon-Young, 2009; Thangaratinam & Redman, 2005). Past psychosis Delphi studies have also used 3-stages (see for example Byrne & Morrison, 2014; Langlands et al., 2008; Law & Morrison, 2014; Morrison & Barratt, 2010), and 12 systematic reviews of the Delphi literature found that Delphi research is typically carried out in either 2 or 3 rounds (Niederberger & Spranger, 2020). Furthermore, recommendations from the Health Technology Assessment (HTA) for the use of consensus development methods, suggest that 2 or more ‘rating’ rounds are likely to result in the convergence of individual judgements, as the dispersion of experts’ views lessen with each round of rating (Murphy et al., 1998).

Quasi-anonymous, controlled feedback usually takes place in the ‘evaluation phase’ (stage 3) whereby the experts are provided with the panel’s group response and asked to re-consider their original responses in light of the group feedback (Iqbal & Pilon-Young, 2009). The HTA argue there is little research as to what type of feedback is best, however in the majority of studies there is limited exchange of information (Murphy et al., 1998). Recommendations by Iqbal and Pilon-Young (2009) suggest feeding back group percentages and providing individual scores to enable panellists to assess the diversity or convergence of the responses. Indeed, this is the only form of communication between the expert panel members; they do not meet or interact directly.

One of the main strengths of the Delphi method (in comparison to other consensus methods) is that individual expert opinions can be elicited confidentially to eliminate the impact of counterproductive group dynamics (Murphy et al., 1998). These counterproductive group dynamics include a) the group discussion going off topic, b) dominant personalities taking over and leading the group discussion, and c) pressure to agree with the opinions of those of a higher status, or those aligned with the same profession (Akins et al., 2005; Brady, 2016; Hsu & Sandford, 2007; Thangaratinam & Redman, 2005). The Delphi method therefore provides an opportunity for all voices to be heard. This is important, because a group of experts providing independent judgements are more likely to produce a valid result, than a single judgement by an individual expert (even if they are the most highly esteemed expert in the field;

Niederberger & Spranger, 2020). Indeed, the Delphi method asserts that “*two heads are better than one*” (Dalkey, 1972, p. 15) as many individuals are less likely to arrive at the wrong decision (Hasson et al., 2000), and this finding was also highlighted in Chapter 2.

In terms of the number of experts recruited into the Delphi study, there are no strict guidelines regarding the optimal sample size (Iqbal & Pilon-Young, 2009; Keeney et al. 2011). Panels of up to 50 experts have been recommended (Turoff, 2002) although most range between 10 to 100 experts based on convenience, purposive, or criterion sampling strategies (Akins et al. 2005). It has been argued that larger panels of experts, equate to a higher degree of reliability and representativeness, as a larger sample is more likely to reflect the opinion of the expert population as a whole (Hasson & Keeney, 2011; Murphy et al., 1998). Therefore, the selection of particular experts is likely to have little impact on the overall group decision, as long as the sample is of sufficient size to carry enough ‘weight’ (Murphy et al., 1998; Niederberger & Spranger, 2020). Nevertheless, the risks inherent within a larger sample, include high attrition rates (Hsu & Sandford, 2007), and Sumsion (1998) has argued that a 70% response rate should be maintained throughout the Delphi study, as this level of retention is indicative of the preservation of rigour in utilising the method, and confidence in the validity of the findings. However, achieving this level of response, requires substantial effort on behalf of the research team (Keeney et al., 2011)

Traditional Delphi research tends to use the same pool of experts in stage 1 and subsequent stages (Thompson, 2009). However, variations of the Delphi method have used a smaller sample in the first stage (see Byrne & Morrison, 2014; Law & Morrison, 2014; Morrison & Barratt, 2010) primarily for pragmatic reasons, to reduce the amount of qualitative data that is generated when consulting a large sample of experts. Delphi guidance within clinical psychology, suggest that the panel in stage 1 can be expanded, by opening it up to invite more experts to participate in stages 2 and 3 (Iqbal & Pilon-Young, 2009).

3.2.1.1 *How to define an expert?*

There is much debate over what constitutes an expert, and how to identify a professional as an ‘expert’ (Hasson & Keeney, 2011; Murphy et al., 1998). Experienced clinicians working within the field have clinical expertise, senior academics have

scientific/research expertise, and individuals that have experienced the condition or intervention are ‘experts by experience’ (Murphy et al., 1998). This led Niederberger and Spranger (2020) to state: *“the definition of an expert is either based on their individual’s scientific/professional expertise or life worldly experience”* (p. 3).

Delphi panels usually consist of a mixture of both clinicians and academics, and it is this diversity within the panel that allows for the consideration of a wide variety of expert opinion (Murphy et al., 1998). The composition of the panel is therefore important, and in a homogenous group of one speciality, the group decision tends to reflect the majority view. However, there may be a moving together (or moving apart) depending upon whether the experts represent different disciplines (e.g. psychiatry, clinical psychology, or nursing) within a particular speciality (e.g. CBTp; Murphy et al., 1998).

Ultimately the Delphi method seeks input from a sample of high-quality experts with specific knowledge and expertise on the topic of inquiry, and participants are identified on the basis of predetermined ‘expert criteria’ (Brady, 2016; Hasson et al., 2000). As Hsu and Sandford (2007) stated: *“choosing the appropriate subjects is the most important step in the entire [Delphi] process because it directly relates to the quality of the results generated”* (p. 3).

As such, the researchers must carefully select the most qualified individuals, with specialist knowledge of the topic under investigation, so the experts can make the judgements and then help to implement the findings (Hsu & Sandford, 2007). CBT experts in the UK have been broadly defined as:

experienced CBT therapists whose research or clinical activities focused on the assessment and evaluation of CBT competence...identified as experts through their involvement in the development and provision of British Association for Behavioural and Cognitive Psychotherapies-accredited CBT training courses and/or recent publications related to training or assessment of CBT therapists (Muse & McManus, 2016, p. 248).

Rector and Cassin (2010) have also stated that *“continuous professional reflection is what distinguishes CBT therapists with more expertise from those with less expertise”* (p. 159). Specific indices of CBT expertise also include *“individuals who have published treatment manuals or used them in published research”* (Duncan et al. 2004, p. 201), and for those deemed to have expertise in formulation, defined as having

“led one or more workshops for professionals on how to construct case formulations” (Eells & Lombart, 2003, p. 189). Despite the definitions above, this more holistic definition of a CBT ‘expert’ is preferred:

Experts in the field of CBTp (i.e. Clinical Psychologists, Cognitive Therapists, Psychiatrists and other mental health professionals) who work (or have worked) in a clinical or research capacity to implement CBT for the treatment of psychosisidentified from a list of clinical academics and trial therapists who a) have been delegates at National and/or International CBTp conferences, b) have written treatment manuals, or c) have worked as trial therapists on randomised control trials (RCTs) of CBTp (Morrison & Barratt, 2010, p. 137).

3.2.2 Aims

This study aimed to achieve consensus as to what experts considered to be core components, for the formulation of voices, and persecutory delusions in CBT for schizophrenia spectrum disorders. Whether it is possible to reach consensus about how we conceptualise individual psychotic symptoms, seems a highly relevant research question.

3.3 Methods

3.3.1 Participants

3.3.1.1 Pre-defined expert criteria

The pre-defined ‘expert’ criteria specified that all participants (at each stage of the study), had to have undertaken one or more of the following: 1. Written a CBTp treatment manual, 2. Conducted a RCT or treatment trial of CBTp as a Chief or Principal Investigator, 3. Involved in a RCT or treatment trial of CBTp by providing expert supervision or training, 4. Involved in a RCT or treatment trial of CBTp as a trial therapist, 5. Supervised clinicians on how to construct CBTp CFs in clinical practice, 6. Led one or more workshops/teaching sessions for clinicians on how to construct CBTp CFs.

3.3.1.2 *Expert identification and recruitment*

Experts recruited in stage 1 of the study, were predominantly some of the most eminent international CBTp leaders in the field. They were identified via the literature as a lead author of a CBTp treatment manual, and/or as a Chief Investigator on a major RCT of CBTp (for example, authors of major CBTp RCTs that were listed in a systematic review by Bighelli et al., 2018). This approach to participant recruitment in stage 1, followed the Morrison and Barratt (2010) approach to expert identification, in that experts with ‘high esteem indicators’ (pioneers/developers/innovators of CBTp) were primarily recruited in stage 1.

Experts recruited in stage 2 of the study, were also identified via the literature as a lead author of a CBTp treatment manual, and/or as a Chief Investigator on a major RCT of CBTp. In addition, experts were identified as clinical academics/trial therapists that had attended a recent International CBTp conference (e.g. Beckfest, Oxford, 2018; the 9th World Congress of Behavioural and Cognitive Therapies (WCBCT), Berlin, 2019), or as an expert in the field of CBTp that was well known. Again, this approach to participant recruitment in stage 2, followed the Morrison & Barratt (2010) approach to expert participant identification, in that experts with ‘high applied indicators’ (implementers/deliverers/followers of CBTp) were primarily recruited in stage 2.

The experts were approached via several recruitment methods including: personal email invites (see Appendices H, I, and J for examples of email invites that invited experts to participate in stages 1-3 of the study), circulation of the study information via email distribution lists (e.g. CBTp experts working at the Psychosis Research Unit (PRU) in Manchester led by Prof. Anthony Morrison), advertisements of flyers at International CBTp conferences (see Appendix K for a copy of the flyer that was circulated at the 9th WCBCT in Berlin, 2019), advertisements on social media (see Appendix L for an example of a recruitment invite, circulated by H.M.S. on Twitter), a PowerPoint presentation delivered by H.M.S. outlining the Delphi study protocol to CBTp experts that attended the psychosis specialist interest group in the North East of England, and (in stage 2) invites to participate in the study from senior academics/clinicians distributed to their colleagues, via snowball sampling techniques. Recruitment was further facilitated by offering a book prize draw to participants, with a 1 in 3 chance of winning a copy of a new CBTp book, edited by Douglas Turkington (D.T.) and H.M.S. (Turkington & Spencer, 2019).

Forty-one initial invitations were sent out via email to potential participants in stage 1, and 111 email invitations were sent out to potential participants in stage 2. However, it is not possible to report the actual response rates for stages 1 and 2, owing to the use of snowball sampling techniques, as the author (H.M.S.) cannot be sure how many experts were approached to participate. Only those experts that participated in stage 2 of the study, were invited to participate in stage 3. Some of the initial email invitations sent out in stages 1 and 2, included a personal message from D.T. It was presupposed that personal email invites to participate from an eminent expert in the field, would be more likely to facilitate higher rates of recruitment and retention in the study.

At each stage of the study, up to 3 email reminders were sent if the potential participant had not responded to the original invite.

3.3.1.3 Eligibility screening

All participants were required to complete a screening form in Qualtrics (web-based software, used to conduct survey research) prior to commencing the study, to ensure they met the expert eligibility criteria. The eligibility screening questions were embedded in the demographics form of the Qualtrics survey in stage 1 of the study (see Appendix M, Q.13, pg. 4 of the survey). These questions were presented at the start of the Qualtrics survey in stage 2 of the study (see Appendix N, pg. 2 of the survey) to prevent participants consenting to take part in the study (and completing the survey) that were not eligible to do so. If certain individuals did not meet the eligibility criteria ($n = 7$ in total), then a rule was set up in Qualtrics whereby the survey explained to the person they were not eligible to participate, and the survey was automatically terminated.

The eligibility screening form included the same questions for all participants recruited in stages 1 and 2 of the study, as the expert inclusion criteria remained the same across all 3 stages to allow for a flexible approach to recruitment.

3.3.2 Delphi procedure and analysis

3.3.2.1 Ethical approval

This study was approved by the Faculty of Medical Sciences Research Ethics Committee (REC) at Newcastle University in the UK (reference: 1507/3843/2018; see Appendix O).

3.3.2.2 *The 3-stage process*

The Delphi method consisted of a predetermined 3-stage process as outlined by Langlands et al. (2008). As mentioned earlier, this approach to the Delphi method has also been adopted for use in other psychosis research (see Byrne & Morrison, 2014; Law & Morrison, 2014; Morrison & Barratt, 2010).

3.3.2.3 *Stage 1*

Core components (pertinent to the use of CF for voices, and persecutory delusions) were identified via a scoping literature review - an extensive electronic and manual search of the literature; including CF components outlined in CBTp models, journals, manuals, and book chapters, together with the BPS's good practice guidelines on the use of psychological formulation (DCP, 2011). A list of CF components was compiled by H.M.S. until there was saturation, and no new CBTp components were identified from the literature.

The components were reviewed and finalised into a list of 190 items ($n = 95$ items for auditory hallucinations, and $n = 95$ items for persecutory delusions; see Appendix M, pgs. 6-40 of the survey, for the CF items/components in stage 1). The items were organised under 5 headings, as outlined in the five Ps formulation framework (Dudley & Kuyken, 2014). Furthermore, under each of the five Ps there was an overarching question which asked about the perceived importance of each of the five Ps (see Appendix M, Q's 22-23; Q's 42-43; Q's 74-75; Q's 100-101; Q's 128-129).

A panel of international CBTp experts that participated in stage 1 ($n = 20^{15}$) were asked to further refine the list of items by providing written qualitative feedback

¹⁵ The 20 experts that participated in stage 1, agreed to waive their anonymity to be acknowledged in the publication: Alison Brabban (UK), Faye Doell (Canada), Paul French (UK), Andrew Gumley (UK), Gillian Haddock (UK), Roger Hagen (Norway), Mark Hayward (UK), Elizabeth Kuipers (UK), Tania Lecomte (Canada), Zhanjiang Li (China), Tania Lincoln (Germany), Tony Morrison (UK), Roger Ng

via an online link to the survey in Qualtrics. The qualitative data was collectively reviewed by H.M.S., D.T., and R.D. It was decided *a priori*, that if two or more experts recommended any item(s) should be removed, or any new item(s) should be added, then these recommendations would be incorporated in stage 2 of the Qualtrics survey. In addition, if there was a strong rationale/argument put forward by one of the experts that any new items should be removed, or any new items should be added, then these suggestions were discussed by H.M.S., D.T., and R.D., and a collective decision was made. Furthermore, H.M.S. decided to remove a couple of items from the survey which appeared to be duplicated and/or incorrectly categorised (in terms of the five Ps framework).

This resulted in amendments to the formatting of the survey, amendments to the categorisation of certain items, the re-wording of several items, the addition of 20 new items, and the removal of 32 items (see Appendix P for those items that were added and removed in stage 1). The finalised survey from stage 1 (to be used in stage 2) contained 178 items ($n = 89$ items for auditory hallucinations, and $n = 89$ items for persecutory delusions). This included 16 presenting issues ($n = 8$ items for auditory hallucinations, and $n = 8$ items for persecutory delusions), 36 predisposing factors ($n = 18$ items for auditory hallucinations, and $n = 18$ items for persecutory delusions), 16 precipitating factors ($n = 8$ items for auditory hallucinations, and $n = 8$ items for persecutory delusions), 90 perpetuating factors ($n = 45$ items for auditory hallucinations, and $n = 45$ items for persecutory delusions), and 20 protective factors ($n = 10$ items for auditory hallucinations, and $n = 10$ items for persecutory delusions). See Appendix N, pgs. 6-32 of the survey, for the CF items/components of the survey, that were used in stage 2.

3.3.2.4 Stage 2

The stage 2 survey was initially piloted by D.T., to determine the time frame taken to complete the survey, as well as its readability.

A total of 78 international experts from 12 different countries completed the survey in stage 2 (including 15 of the original experts that participated in stage 1). Prior to completing the survey, the experts were provided with a brief set of instructions

(China), Sara Tai (UK), Nick Tarrier (UK), Neil Thomas (Australia), Mark van der Gaag (The Netherlands), Dawn Velligan (USA), Jesse Wright (USA), and Nicola Wright (Canada).

which asked them to think about a complete CF that had focused on events leading to the emergence, and maintenance of the psychosis, which had been developed by the end of therapy (see Appendix N, pgs. 3-4 in the survey). The experts were then asked to rate the importance of each formulation component in relation to the treatment of auditory hallucinations, then persecutory delusions. Each statement was rated on a Likert scale of 1-5 (1, essential; 2, important; 3, neither important nor unimportant; 4, unimportant; 5, should not be included). The results from the survey in stage 2, were downloaded from Qualtrics into SPSS for analysis, and the data was analysed by obtaining group percentages.

In accordance with the methods used by Langlands et al. (2008), the following criteria were used in relation to the inclusion, exclusion, and rerating of items:

1. If >80% panel members rated an item as ‘essential’ or ‘important’ it was included.
2. If 70-79% of panel members rated an item as ‘essential’ or ‘important’, then all panel members were asked to rerate that item.
3. Any items that did not meet the 2 conditions above were excluded.

This resulted in the inclusion of 156 items (out of a total number of 178 items), the exclusion of 6 items, with 16 items to be rerated in stage 3. In total, 87.6% of the items were endorsed by >80% of the experts as either ‘essential’ or ‘important’, including all 5 of the overarching five P items, for both auditory hallucinations and persecutory delusions (see Appendix Q for the items endorsed as ‘essential’ or ‘important’). It is also worth noting (as indicated in Appendix Q) that a total number of 25 items were endorsed by 100% of the expert panel as ‘essential’ or ‘important’ ($n = 12$ items for auditory hallucinations, and $n = 13$ items for persecutory delusions), and this included 2 of the overarching five P items (perpetuating and protective factors). As such, there was a high level of consensus regarding the items seen as both ‘essential’ and ‘important’.

These findings indicate that the items elicited from the literature review, (and stage 1 expert review) were viewed as highly relevant when they were rated in stage 2. Nevertheless, given the desire for simplicity and parsimony (Butler, 1998; Kuyken et al., 2009; Kuyken et al., 2016; Padesky et al., 2011) a decision was taken to shift the analysis to focus on the items endorsed as ‘essential’ only (whereas the analysis

procedure undertaken by Langlands et al. (2008) included items endorsed as either ‘essential’ or ‘important’).

The following criteria were then used in relation to the inclusion, exclusion, and rerating of items:

1. If >80% panel members rated an item as ‘essential’ it was included.
2. If 70-79% of panel members rated an item as ‘essential’, then all panel members were asked to rerate that item.
3. Any items that did not meet the 2 conditions above were excluded.

This resulted in the inclusion of 8 items (including 1 of the overarching five P items for both auditory hallucinations and persecutory delusions), the exclusion of 157 items¹⁶, with 13 items¹⁷ to be rerated in stage 3.

It is worth noting that a total number of 46 items in stage 2 ($n = 23$ for persecutory delusions, $n = 23$ for auditory hallucinations) achieved the same consensus ratings for both symptoms.

3.3.2.5 Stage 3

All 78 international experts that participated in the previous stage were emailed a summary of the findings so far. This contained a list of the formulation components that >80% of the panel had rated as essential in stage 2 ($n = 8$ items).

In stage 3, each expert was also sent a final survey (see Appendix R for an example of a stage 3 survey), which provided anonymous group feedback in terms of overall group percentages and frequencies, as well as individual scores, with regards to the items that 70-79% of panel members had rated as essential in stage 2 ($n = 13$ items).

¹⁶ $n = 20$ of the excluded items were rated by some experts as ‘should not be included’. For these 20 items, the frequency range for ‘should not be included’ was between 1-3 experts, and the consensus range for ‘should not be included’ was endorsed by 1.3%-3.8% of the expert sample.

¹⁷ $n = 4$ of the 13 items were rated by 79.5% of the panel as ‘essential’ in stage 2 (and so almost reached consensus for endorsement at the level of >80%). When these 4 items were rerated again in stage 3, they were all endorsed as ‘essential’ by >80% of the expert panel in the final Delphi stage (and therefore included).

The expert panel were then asked to rerate these 13 items ($n = 5$ items for auditory hallucinations, and $n = 8$ items for persecutory delusions), taking into consideration their original responses, as well as the overall group ratings from the previous round. The experts were then provided with a brief set of instructions which explained that they could keep their original response(s) from stage 2, or they could change their response(s) and provide new rating(s).

70 CBTp experts responded in stage 3, giving a response rate of 89.7% (including 14 of the original experts that took part in stage 1). This resulted in an additional 8 items being included, and 5 items being excluded¹⁸. As in stage 2, items were included if they were rated by >80% of the expert panel as 'essential' and items which did not reach this level of consensus were excluded.

It is worth noting that a total number of 2 items in stage 3 ($n = 1$ for persecutory delusions, $n = 1$ for auditory hallucinations) achieved the same consensus rating for both symptoms.

¹⁸ No items were rated as 'should not be included'.

3.4 Results

The 78 experts that participated in the main part of this study came from 12 different countries (43, UK; 10, USA; 5, Canada; 4, Japan; 3, China; 3, Germany; 3, Norway; 2, Australia; 2, The Netherlands; 1, Pakistan; 1, Russia; 1, Spain). The 78 participants also met a number of expert criteria (in stage 2, $n = 33$ had written a CBTp treatment manual, $n = 23$ had conducted a RCT or treatment trial of CBTp as a Chief or Principal Investigator, $n = 36$ had been involved in a RCT or treatment trial of CBTp by providing expert supervision or training, $n = 50$ had been involved in a RCT or treatment trial of CBTp as a trial therapist, $n = 70$ had supervised clinicians on how to construct CBTp CFs in clinical practice, $n = 66$ had led one or more workshops/teaching sessions on how to construct CBTp CFs). A demographics form was included in Qualtrics at stage 1 (see Appendix M pgs. 3-5 in the survey) and stage 2 (see Appendix N pgs. 4-6 in the survey) to collect additional information about the expert participant sample. Table 3.1 outlines the characteristics of the experts at all 3 stages. Figure 3.1 illustrates the number of items that were included, rerated, and excluded at each stage of the study.

A total of 16 items were endorsed as essential by >80% of the expert panel members, with 8 items endorsed as essential for the formulation of auditory hallucinations ($n = 1$ presenting issue; $n = 7$ perpetuating factors) and 8 items endorsed as essential for the formulation of persecutory delusions ($n = 1$ presenting issue; $n = 7$ perpetuating factors). This included 1 of the overarching five P items (perpetuating factors), which was endorsed as essential for the CF of both auditory hallucinations, and persecutory delusions.

The exact same items were endorsed for both symptoms, however some of the items were endorsed at different stages. Table 3.2 below, shows the items that reached consensus as ‘essential’ in stages 2 and 3.

Table 3.1 Participant characteristics at each stage of the study

	Stage 1 (<i>n</i> = 20)	Stage 2 (<i>n</i> = 78)	Stage 3 (<i>n</i> = 70)
Gender			
Male	11 (55%)	29 (37.2%)	25 (35.7%)
Female	9 (45%)	49 (62.8%)	45 (64.3%)
Country currently working¹⁹			
UK	9 (45%)	43 (55.1%)	38 (54.3%)
USA	2 (10%)	10 (12.8%)	9 (12.9%)
Canada	3 (15%)	5 (6.4%)	4 (5.7%)
Japan	0 (0%)	4 (5.1%)	4 (5.7%)
China	2 (10%)	3 (3.8%)	3 (4.3%)
Germany	1 (5%)	3 (3.8%)	3 (4.3%)
Norway	1 (5%)	3 (3.8%)	3 (4.3%)
Australia	1 (5%)	2 (2.6%)	2 (2.9%)
The Netherlands	1 (5%)	2 (2.6%)	2 (2.9%)
Pakistan	0 (0%)	1 (1.3%)	0 (0%)
Russia	0 (0%)	1 (1.3%)	1 (1.4%)
Spain	0 (0%)	1 (1.3%)	1 (1.4%)
Level(s) of expertise			
Written a CBTp treatment manual	19 (95%)	33 (42.3%)	31 (41.3%)

¹⁹ Or if retired, the country where they had previously worked.

Conducted a RCT or treatment trial of CBTp as a Chief or Principal Investigator	17 (85%)	23 (29.5%)	21 (30%)
Involved in a RCT or treatment trial of CBTp by providing expert supervision or training	16 (80%)	36 (46.2%)	31 (44.3%)
Involved in a RCT or treatment trial of CBTp as a trial therapist	14 (70%)	50 (64.1%)	44 (62.9%)
Supervised clinicians on how to construct CBTp CFs in clinical practice	20 (100%)	70 (89.7%)	63 (90%)
Led one or more workshops/teaching sessions for clinicians on how to construct CBTp CFs	19 (95%)	66 (84.6%)	58 (82.9%)
Profession			
Psychiatrist	3 (15%)	8 (10.3%)	8 (11.4%)
Clinical Psychologist	16 (80%)	54 (69.2%)	47 (67.1%)
Counselling Psychologist	0 (0%)	1 (1.3%)	1 (1.4%)
Psychologist (post doc)	0 (0%)	1 (1.3%)	1 (1.4%)
Cognitive Therapist	0 (0%)	7 (9%)	6 (8.6%)
Nursing	1 (5%)	5 (6.4%)	5 (7.1%)
Other (1 Chartered Psychologist/Psychological Therapist. 1 PhD level Licensed Professional Clinical Counselor with Supervision Designation)	0 (0%)	2 (2.5%)	2 (2.9%)
Years worked in a clinical and/or research capacity implementing CBTp			
Less than 5 years	0 (0%)	13 (16.7%)	12 (17.1%)
5-9 years	0 (0%)	13 (16.7%)	12 (17.1%)
10-14 years	4 (20%)	19 (24.4)	16 (22.9%)
15-19 years	4 (20%)	15 (19.2%)	15 (21.4%)
20-24 years	6 (30%)	10 (12.8%)	9 (12.9%)
25-29 years	4 (20%)	4 (5.1%)	3 (4.3%)

30+ years	2 (10%)	4 (5.1%)	3 (4.3%)
-----------	---------	----------	----------

Figure 3.1 Number of items included, rerated, and excluded at each stage of the study

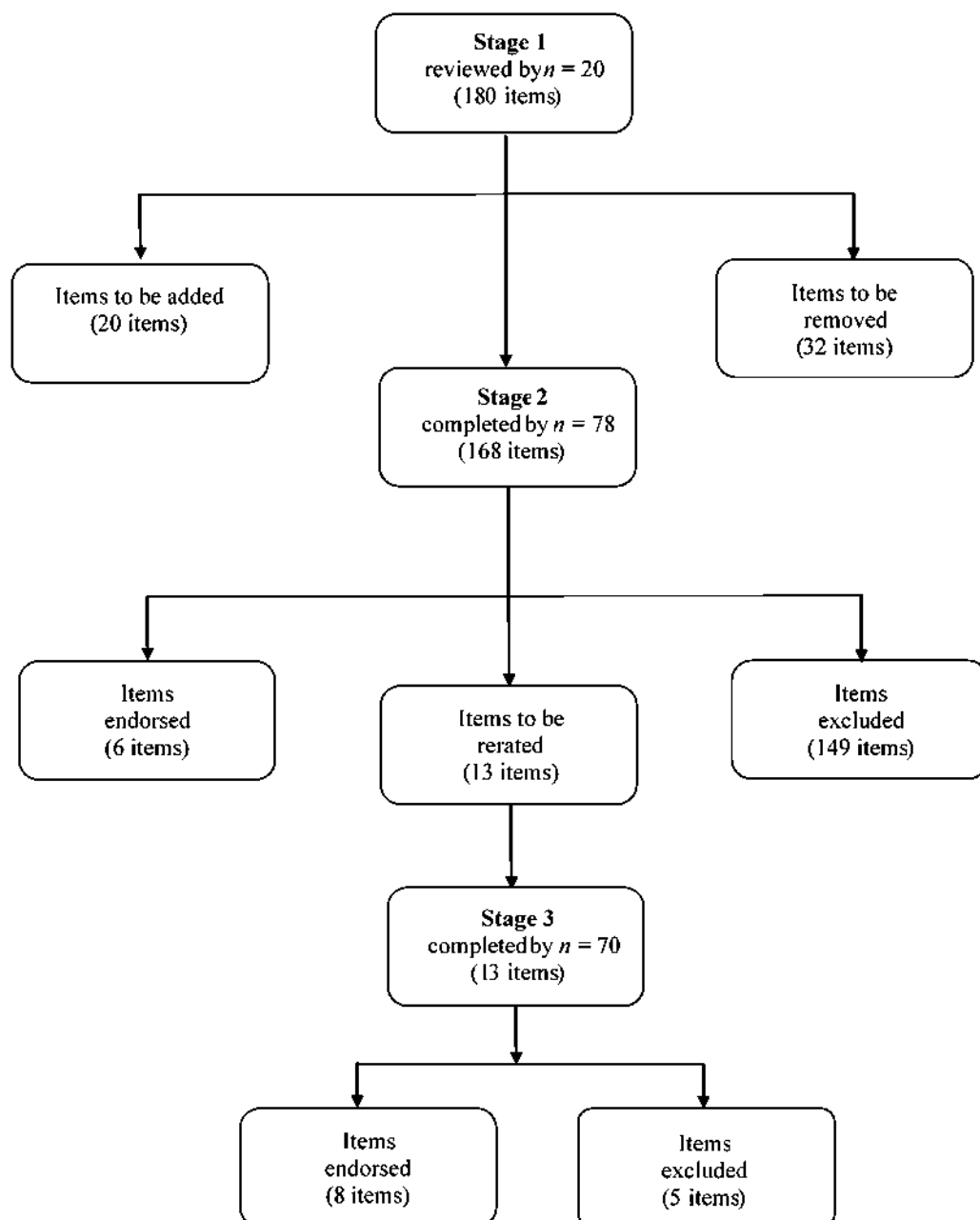


Table 3.2 Components endorsed as ‘essential’ in the CF of auditory hallucinations, and persecutory delusions

	Auditory Hallucinations		Persecutory Delusions	
	Stage	Percentage	Stage	Percentage
	Included	Agreement (n)	Included	Agreement (n)
<i>Presenting Issues: What are the problems?</i>				
<i>The presenting issues list in a formulation for auditory hallucinations, or persecutory delusions should include:</i>				
Identification of risk to self and/or others (e.g. acting on command hallucinations, acting on the persecutory beliefs, self-neglect, or self-harm/suicidal preparation).	3	90 (63)	3	90 (63)
<i>Perpetuating Factors: What keeps the problems going?</i>				
How important do you believe perpetuating factors to be, when working to treat auditory hallucinations, or persecutory delusions?	2	82 (64)	2	85 (66)
<i>The perpetuating factors in a formulation for auditory hallucinations, or persecutory delusions should include:</i>				
Current appraisals such as: Appraisal of presenting symptoms (e.g. that the identity of the voices or persecutors are dangerous and evil).	2	83 (65)	3	81 (57)

Current appraisals such as: Appraisals about others (e.g. that other people will treat the individual differently if they know they hear voices'; that other people are dangerous and want to harm them).	2	82 (64)	2	81 (63)
Current appraisals such as: Appraisals about the self (e.g. that the individual deserves to be harmed; that the voices or persecution are a sign of losing control, badness, or deserved punishment).	2	82 (64)	3	87 (61)
Behavioural factors such as: Ineffective (potentially harmful) coping strategies which are used with intent, to diminish the severity of the psychotic symptoms (e.g. drinking alcohol to sleep at night, rather than lying awake listening to voices, or ruminating about 'persecutors').	2	82 (64)	3	90 (63)
Behavioural factors such as: Safety seeking behaviours which are an automatic way of responding, to diminish the severity of the psychotic symptoms (e.g. direct, or subtle avoidance, escape, within-situation behaviours, or symbolic neutralising behaviours (e.g. cutting up paper to appease the voices rather than cutting the self; or punching a wall rather than hitting a suspected 'persecutor'))).	3	89 (62)	2	81 (63)

Emotions which may be directly linked to the content of the voices or persecution such as: High levels of arousal, fear, anger, shame, guilt etc.	3	83 (58)	3	81 (57)
---	---	---------	---	---------

None of the predisposing, precipitating, or protective factors were endorsed by >80% of the panel as essential for the CF of either auditory hallucinations, or persecutory delusions. Furthermore, 4 of the 5 overarching five P items were not endorsed as essential. Therefore, no consensus was reached amongst the experts regarding these components (Appendix S shows the items that were excluded).

The items considered less than important (rated <80% as ‘essential or ‘important’) in stages 2 and 3 ($n = 22$ items) are outlined in Appendix T.

3.5 Delphi secondary analysis

3.5.1 Secondary research question

As previously mentioned in section 4.3.2.5, 13 of the items achieved 70%-79% consensus as 'essential' in stage 2, which the experts were then asked to rerate in stage 3. Of those 13 items, 8 items were endorsed by >80% of the expert panel as 'essential' in stage 3. This meant that 5 items were rerated in stage 3 but were *not* endorsed as 'essential' (Table 3.3 below, shows the 5 components *not* endorsed as 'essential'). However, 4 of these 5 items achieved >70% consensus as 'essential', and 1 of the 5 items achieved >60% consensus as 'essential', in the final Delphi stage, meaning they fell just below the threshold for inclusion in the study.

A secondary research question is whether these 5 items *may* have been endorsed by a subgroup of experts with 'high esteem indicators', had H.M.S. just asked them for their consensus ratings in stage 3 (and disregarded the ratings of the wider sample of experts).

3.5.2 Rationale

This is an interesting question to investigate because a subgroup of experts within the sample had undertaken a higher proportion (and higher level) of the 'expert criteria', in comparison to the wider sample of experts that took part in this study.

3.5.3 Expert groups

Table 3.4 below, shows the different level(s) of expertise in stage 3, between a group of experts with 'higher esteem indicators' in comparison to the wider sample of expert participants. The group of experts with 'high esteem indicators' ($n = 26$), had a higher proportion of those who had 1) written a CBTp treatment manual, 2) conducted a RCT or treatment trial of CBTp as a Chief or Principal Investigator, and 3) had been involved in a RCT or treatment trial of CBTp by providing expert supervision or training. Therefore, the 'high esteem indicator' experts represented a subgroup of the total sample that were manual developers and RCT/treatment trial leaders, as well as having

high levels of experience in supervision, and training. As such, the experts with ‘high esteem indicators’ appeared to be pioneers/developers/innovators and implementers of CBTp and were defined as having met *at least* 5 of the 6 expert study inclusion criteria.

The wider sample of experts in stage 3 had primarily ‘applied indicators’ ($n = 44$) meaning that they appeared to be a subgroup of experts that were mainly RCT/treatment trial therapists, and supervisors/trainers. As such, the experts with high levels of ‘applied indicators’ appeared to be implementers/deliverers/followers of CBTp and were defined as having met *less than* 5 of the expert study inclusion criteria.

The subsequent tables below, show the 5 components *not* endorsed as ‘essential’ by >80% of the experts in stage 3 (Table 3.3), and the different level(s) of expertise in stage 3, between a subgroup of experts with ‘higher esteem indicators’ in comparison to the wider sample of experts with ‘applied indicators’ (Table 3.4).

Table 3.3 Components *not* endorsed in stage 3 as ‘essential’ in the CF of auditory hallucinations, and persecutory delusions

	Auditory Hallucinations	Persecutory Delusions
	Percentage Agreement (n)	Percentage Agreement (n)
<i>Perpetuating Factors: What keeps the problems going?</i>		
<i>The perpetuating factors in a formulation for auditory hallucinations, or persecutory delusions should include:</i>		
Current appraisals such as: Appraisals about the world (e.g. that the world is unpredictable).	-	74 (52)
Current appraisals such as: Negative appraisals of life events (e.g. bad things happen to me, so this means I am a bad person).	64 (45)	-
Behavioural factors such as: Schema maintenance safety behaviours (e.g. Praying or carrying out religious rituals to appease the voices, both serve to reinforce the belief that the voices are powerful and	-	76 (53)

maintain the core schema that the individual is weak and vulnerable.

Acting in a way that elicits hostility from others by being aggressive

or acting in a way that elicits social isolation by treating others

suspiciously, both serve to maintain the core schema that other

people are bad).

Protective Factors: What are the person's strengths?

The protective factors in a formulation for auditory hallucinations, or persecutory delusions should include:

Helpful coping strategies (e.g. acceptance, mindfulness, exercise,

77 (54)

76 (53)

eating well, capacity to self-soothe/self-nurture, meditation).

Table 3.4

Level(s) of expertise in stage 3, between a subgroup of experts with ‘higher esteem indicators’ versus the wider sample of experts with ‘applied indicators’

	Experts with ‘high esteem indicators’ (<i>n</i> = 26)	Experts with ‘applied indicators’ (<i>n</i> = 44)
Level(s) of expertise		
Written a CBTp treatment manual	24 (92%)	7 (16%)
Conducted a RCT or treatment trial of CBTp as a Chief or Principal Investigator	21 (81%)	0 (0%)
Involved in a RCT or treatment trial of CBTp by providing expert supervision or training	22 (85%)	9 (20.5%)
Involved in a RCT or treatment trial of CBTp as a trial therapist	22 (85%)	22 (50%)
Supervised clinicians on how to construct CBTp CFs in clinical practise	26 (100%)	37 (84%)
Led one or more workshops/teaching sessions for clinicians on how to construct CBTp CFs	25 (96%)	33 (75%)

3.5.4 *Secondary analysis*

To investigate whether there were any differences in the consensus ratings between the experts with ‘high esteem indicators’ vs the experts with ‘applied indicators’ in stage 3, H.M.S. examined whether the 5 items differed (in terms of their level of endorsement as ‘essential’) between the 2 groups.

The Shapiro-Wilk test of normality was .00 for all 5 items, therefore a Mann-Whitney U test was calculated at a two-tailed significance level of .01 (a Bonferroni correction was applied for the 5 different tests, and so the p value reduced $.05/5 = .01$).

3.5.4.1 *Perpetuating factors: What keeps the problems going?*

1. For item 33.iii in relation to persecutory delusions, a Mann-Whitney U test indicated that ‘current appraisals such as: appraisals about the world (e.g. that the world is unpredictable)’ was rated the same for the subgroup of experts with ‘high esteem indicators’ ($Mdn = 1.0, n = 26$) in comparison to the wider sample of experts with ‘applied indicators’ ($Mdn = 1.0, n = 44$), $U = 460.00, p = .08, IQR = 1$.
2. For item 34.v in relation to auditory hallucinations, a Mann-Whitney U test indicated that ‘current appraisals such as: negative appraisals of life events (e.g. bad things happen to me, so this means I am a bad person)’ was rated the same for the subgroup of experts with ‘high esteem indicators’ ($Mdn = 1.0, n = 26$) in comparison to the wider sample of experts with ‘applied indicators’ ($Mdn = 1.0, n = 44$), $U = 513.00, p = .37, IQR = 1$.
3. For item 44.vii in relation to persecutory delusions, a Mann-Whitney U test indicated that ‘behavioural factors such as: schema maintenance safety behaviours (e.g. Acting in a way that elicits hostility from others by being aggressive or acting in a way that elicits social isolation by treating others suspiciously, both serve to maintain the core schema that other people are bad)’ was rated the same for the subgroup of experts with ‘high esteem indicators’ ($Mdn = 1.0, n = 26$) in comparison to the wider sample of experts with ‘applied indicators’ ($Mdn = 1.0, n = 44$), $U = 439.00, p = .04, IQR = 1$.

3.5.4.2 *Protective factors: What are the person's strengths?*

4. For item 52 in relation to persecutory delusions, a Mann-Whitney U test indicated that 'helpful coping strategies (e.g. acceptance, mindfulness, exercise, eating well, capacity to self-soothe/self-nurture, meditation)' was rated the same for the subgroup of experts with 'high esteem indicators' ($Mdn = 1.0, n = 26$) in comparison to the wider sample of experts with 'applied indicators' ($Mdn = 1.0, n = 44$), $U = 526.00, p = .04$, $IQR = 1$.

5. For item 52 in relation to auditory hallucinations, a Mann-Whitney U test indicated that 'helpful coping strategies (e.g. acceptance, mindfulness, exercise, eating well, capacity to self-soothe/self-nurture, meditation)' was rated the same for the subgroup of experts with 'high esteem indicators' ($Mdn = 1.0, n = 26$) in comparison to the wider sample of experts with 'applied indicators' ($Mdn = 1.0, n = 44$), $U = 526.00, p = .04$, $IQR = 1$.

3.5.5 **Results**

The results for the secondary analysis showed no significant differences for each of the 5 CF items that were *not endorsed* in the final Delphi stage (stage 3). This indicated that the subgroup of experts with 'high esteem indicators' *would not* have endorsed any of the 5 items that did not achieve consensus as 'essential' components of a CF in stage 3, had H.M.S. just asked *them*. We can therefore be confident in the reliability of the original findings, that the total expert sample (as a whole), were rating the items similarly in the final Delphi stage, despite apparent differences in their level(s) of expertise.

3.6 Discussion

This is the first Delphi study to establish an expert consensus of the essential CBT ingredients of a CF for auditory hallucinations, and persecutory delusions in schizophrenia spectrum disorders. The study involved recruitment of a large panel of international experts from Europe, North America, and Asia, helping to ensure diverse cultural representation.

The results showed that 1 presenting issue was endorsed as essential for the CF of both symptoms, and this related to the identification of risk to self and/or others. The same 7 perpetuating factors were also endorsed as essential for both symptoms, including 1 of the overarching five P items - perpetuating factors, which were generally viewed as essential. More specifically, 3 of the perpetuating factors related to current appraisals about the presenting symptoms, the self, and others. Two further perpetuating factors related to behaviours - ineffective coping strategies, and safety seeking behaviours. A final perpetuating factor related to emotions, linked to the content of the voices or persecution. Figures 3.2 and 3.3 below, provide illustrative content to show the components that were endorsed by the experts as essential²⁰.

Presenting issues are the specific problems or difficulties a SU describes when they first seek help via therapy. A CF can only include the information that is available, so the process of outlining presenting issues is essential groundwork for the subsequent CF (Dudley & Kuyken, 2014). The only presenting issue endorsed as an essential component for the CF of both auditory hallucinations, and persecutory delusions was that of risk to self and/or others (for example, acting on command auditory hallucinations or persecutory beliefs self-neglect, or self-harm/suicidal preparation).

Acting on command hallucinations, or perceived persecution, is a significant clinical issue. Although only a small minority of SUs with psychosis act violently (Walsh et al., 2002), previous research has found that command hallucinations and suspicious/persecutory beliefs (which imply threat to the individual) may be associated with violence towards others, suicidal behaviour, or self-harm (Braham et al., 2004; Coid et al., 2013; Ullrich et al., 2018). As such, the assessment of risk in SUs diagnosed with schizophrenia spectrum disorders is a core clinical skill, and the responsibility of mental health professionals' worldwide (Ullrich et al., 2018). Indeed, the endorsement

²⁰ Figures 3.2 and 3.3 are provided here as examples for the thesis chapter only. These CF diagrams were not shown to the experts as part of the study.

of this component may be more to do with the essential requirements of services to manage risk, albeit for the person's benefit, than to form a component of CF per se.

Perpetuating factors indicate how difficulties or problems are maintained (Dudley & Kuyken, 2014), and their importance in the CF of voices and persecutory delusions have been well documented (Freeman et al., 2002; Morrison, 1998). Appraisals are the interpretation and meaning attributed to experiences, which are central to cognitive models of psychosis (Morrison, 2001). Appraisals identified in relation to presenting symptoms such as voices, may include beliefs about the nature and identity of the experience itself. For example, a person may hear a voice and think 'the devil' is speaking to them. Or, they may think 'that was a strange experience, I must be overtired' (Kingdon & Turkington, 2008). If the former appraisal is adopted, this may lead to further appraisals about others, such as the perceived power and omniscience of 'the devil', or the appraisal that 'other people will treat me differently if they know I hear voices'. Appraisals about others in relation to persecutory delusions, tend to be that others (namely the persecutors) are dangerous, bad, evil etc. and that they wish the individual harm in some way. This perceived threat, may then activate appraisals about the self, such as 'it is safer (for oneself) to be paranoid' (Morrison et al., 2005), or that the persecution is 'deserved' ('bad me'), or 'undeserved' ('poor me'; Trower & Chadwick, 1995). Appraisals about the self in relation to auditory hallucinations, may include beliefs that voice hearing is a sign of 'weakness', or 'loss of control' (Chadwick et al., 2000).

Given that a person's ability to cope with their voices may vary according to how they appraise them (Romme & Escher, 1989) it is presupposed that if negative beliefs about voices can be reduced, this might ease emotional distress, change problematic behaviours, and facilitate the use of adaptive coping strategies (Chadwick & Birchwood, 1994). It is therefore fitting, that two of the perpetuating factors endorsed as essential in the CF of voices and delusions in this study, related to ineffective coping, and reliance on safety seeking behaviours.

Ineffective (potentially harmful) coping strategies, are used with intent to diminish the severity of the psychotic symptoms and/or the associated distress (such as drinking alcohol to sleep at night rather than lying awake listening to voices or

Figure 3.2 Essential ingredients of a CF for auditory hallucinations

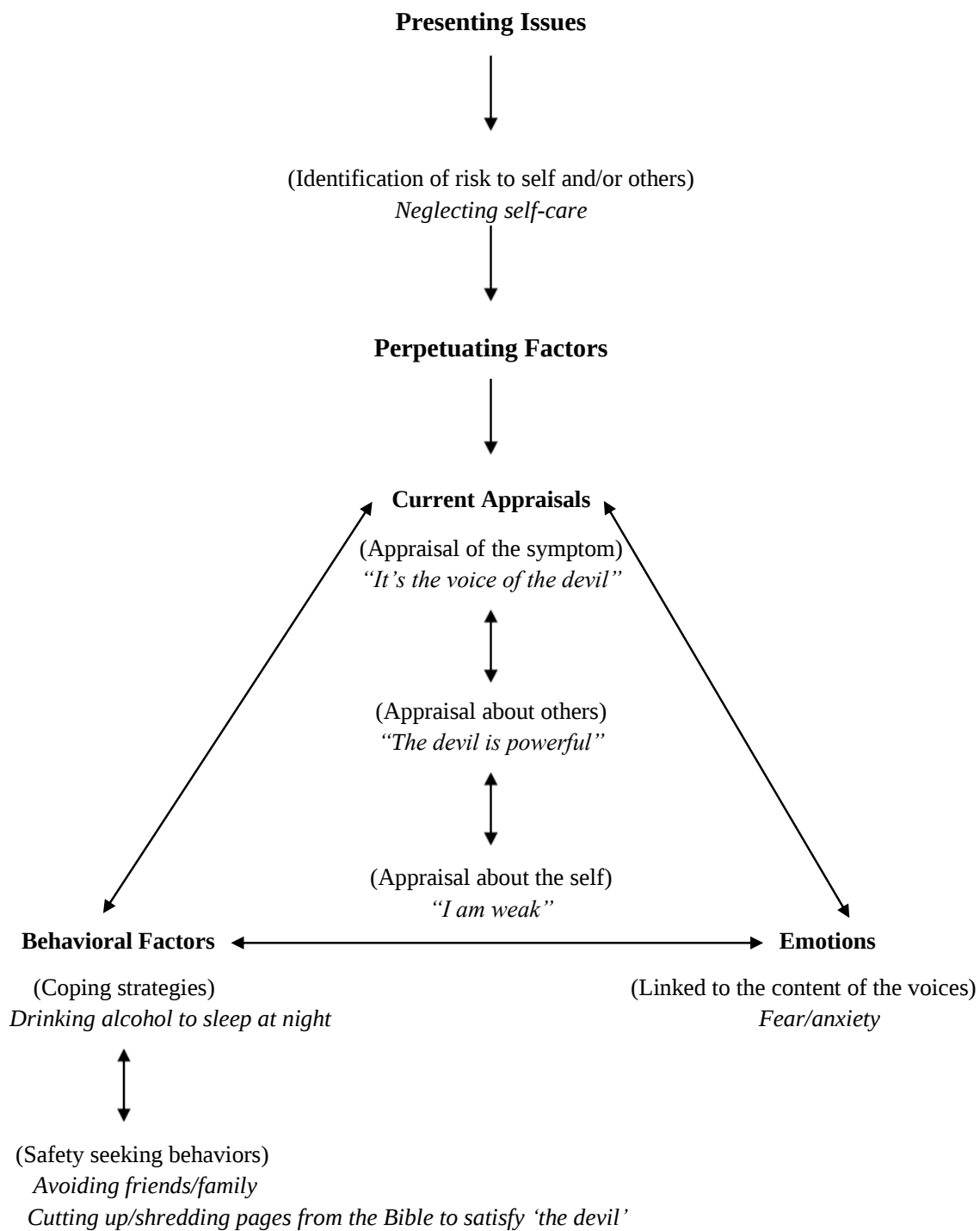
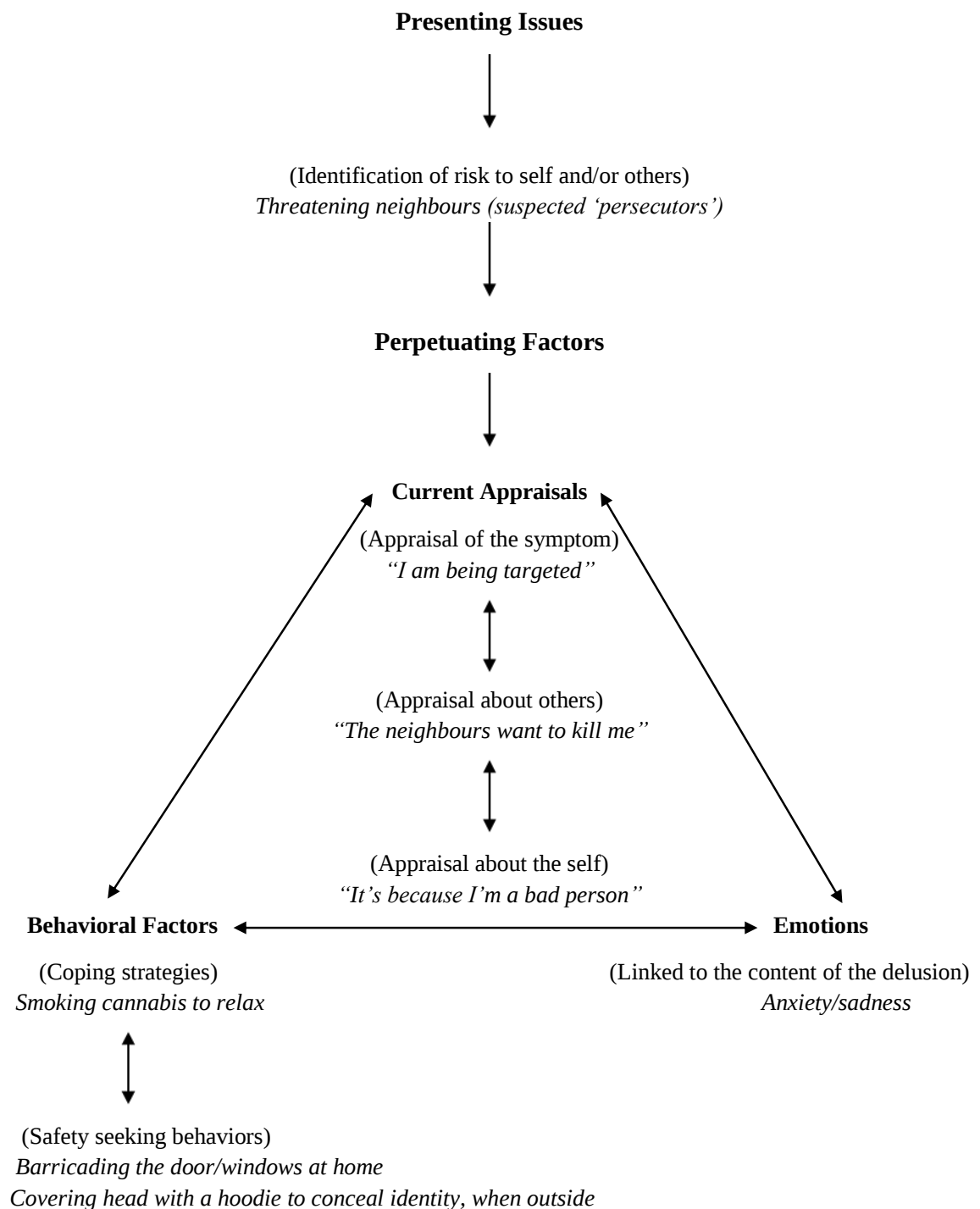


Figure 3.3 Essential ingredients of a CF for persecutory delusions



ruminating about ‘persecutors’). Indeed, one of our previous research studies has shown that the coping strategies naturally adopted by voice hearers, tend to focus on ineffective distraction techniques that are of limited benefit, as well as the use of safety behaviours (Howard et al., 2012). Safety seeking behaviours are a frequently used method of responding to the appraisals, which may include direct or subtle avoidance, escape, within-situation behaviours, or symbolic neutralising behaviours (such as cutting up paper to appease the voices rather than cutting oneself; or punching a wall rather than hitting a suspected ‘persecutor’). There are subtle differences between the two, however both ineffective coping strategies, and safety seeking behaviours may provide temporary relief, but tend to maintain the problem (Thwaites & Freeston, 2005).

That cognitive appraisals elicit (and maintain) certain emotional responses, is also a link that has been well established (Beck et al., 1979). For example, individuals that appraise their voices as malevolent tend to fear them, whereas those who appraise their voices as benevolent tend to feel reassured by them (Chadwick & Birchwood, 1994). Similarly, cognitive models of delusions have emphasised the role of anxiety and depression in the maintenance of persecutory delusions (Freeman et al., 2002), and we have previously argued that escalating levels of anxiety pre-date the emergence of delusional systems (Turkington et al., 2015). As such, it is apt that emotions which may be directly linked to the content of the voices or persecution (such as high levels of arousal, fear, anger, shame, guilt etc.) was also a component that was endorsed as essential.

In terms of the current research, there are many implications of the findings. This is the first study of expert consensus to suggest that two of the most common symptoms in schizophrenia spectrum disorders, should be formulated in the same way. Given that only a small number of essential ingredients were endorsed as essential for auditory hallucinations, and persecutory delusions, the findings also suggest that CFs could be parsimonious, with a specific focus on conceptualising maintenance cycles, to facilitate change. This reflects findings from a similar consensus-seeking study (using Q-sort methodology) that asked therapists to rate the important features of a CF in CBT for depression. The statement ‘how problems are maintained’ was ranked as most essential, leading the authors to suggest that CFs should primarily focus on the ‘here and now’ (Flitcroft et al., 2007). Similarly, a recent qualitative study by Newman-Taylor et al. (2022; published after the current Delphi study), found that many current (or recent) trainees engaging in postgraduate training in CBTp “*expressed a preference*

for simple maintenance cycles” (p. 5) when formulating psychosis, further noting that: *“more complex formulations can be unhelpful in practice*” (p. 7).

In relation to the current study, there was also a high degree of specificity in that all the perpetuating components endorsed, directly related to the appraisal and experience of the symptoms themselves. Such specific CFs that home in on a small number of essential components, may help to filter out large amounts of complex information, that have the potential to create ‘noise’ within a formulation. Simplified CFs can be easily taught, meaning they should be more understandable and memorable for SUs to recall. If they are easier to recall, this may encourage an earlier focus on change processes to facilitate recovery (Spencer et al., 2023; see also Chapter 5).

As mentioned earlier in the Delphi secondary analysis (see section 4.5.1.3) there were several components that were *not* endorsed as essential by >80% of the expert panel in the final Delphi stage, that nevertheless received a high level of endorsement by the expert panel. These items *may* have been endorsed had H.M.S. included a fourth Delphi stage, as they fell just below the threshold for inclusion in the study (therefore they could be considered as posteriori considerations). Table 3.3 shows that 4 of the 5 components were endorsed as essential by >70% of the expert panel, in the final Delphi stage (stage 3). These findings show that 2 perpetuating factors were also highly endorsed, for the CF of persecutory delusions. Specifically, 1 perpetuating factor related to appraisals about the world, and the other related to schema maintenance safety behaviours. Furthermore, 1 protective factor was highly endorsed as essential for the CF of both symptoms, and this related to the use of helpful coping strategies.

Appraisals about the world might include thoughts that ‘the world is unpredictable’, ‘unsafe’, or ‘uncontrollable’, and that the individual should be fearful of the outside world (Freeman et al., 2002). This persecutory ‘world view’ aligns with suspicious appraisals concerning the malevolent intent of others (Peters et al., 2017).

Freeman and colleagues have argued that CBT should target other maintenance factors (such as negative appraisals about the self) before the individual is supported to increase their level of activity in the local community, to relearn that they are safe (Freeman, Bradley, Waite, et al., 2016; Freeman, Waite, Emsley, et al. 2016). Perhaps this explains why ‘appraisals about the world’ did not reach the >80% level of endorsement as an ‘essential’ component, but this component followed behind as a close contender. It is also interesting that it was not highly endorsed as essential for the CF of auditory hallucinations, given that the experts provided similar consensus ratings

for the formulation of both symptoms. Nevertheless, this component *was* highly endorsed by >90% of the panel as ‘essential’ or ‘important’ for the formulation of auditory hallucinations (see Appendix Q).

As previously discussed, safety seeking behaviours were endorsed as an essential component for the CF of both symptoms. Related to this, schema maintenance safety behaviours were also highly endorsed for the CF of persecutory delusions. Schema maintenance safety behaviours may include for example, acting in way that elicits hostility from others by being aggressive, or acting in a way that elicits social isolation by treating others suspiciously (which both serve to maintain the core schema that ‘others are bad’ or that ‘others are hostile’; Fowler et al., 2006).

Segal (1988) defined ‘schema’ as a cohesive and enduring body of knowledge formed in relation to past reactions and experience, that drive how the self, others, and the world are appraised. Once a core schema (or core belief) is formed, the mind does not willingly attend to evidence that might disconfirm it (Gilhooly, 1983). Consequently, when two appraisals are inconsistent with one another the underlying core schema is threatened, and this produces a state of discomfort and ‘not knowing’ referred to as cognitive dissonance (Brabban & Turkington, 2002). Schema maintenance safety behaviours operate to reduce dissonate states and maintain a ‘cognitive equilibrium’ whereby people and events can be reliably predicted (even if this serves to maintain core schema that are distressing e.g. ‘others are bad’).

It has also been argued that persecutory delusions reflect an attributional defence against negative core beliefs about the self, and subsequent low self-esteem (Bentall et al., 1994). Blaming others for negative events (rather than blaming the self, or the situation) prevent negative thoughts about the self from reaching conscious awareness (Bentall et al., 1994). As such, the main function of schema maintenance safety behaviours in this sense, are to defend against underlying (preconscious) core schema vulnerability, leading to the formation and maintenance of persecutory delusional systems (Turkington et al., 2015). As Bentall and Kaney (1996) stated:

When these negative self-representations are primed by threatening events, leading to discrepancies between the self-representations and self-ideals, external (other-blaming) attributions for the threatening events are elicited. These attributions are self-protective in the sense that they reduce the patient’s awareness of discrepancies between the self and self-ideals, but carry the penalty of activating schemata that represent threats from others (p. 1231).

Once again, it is interesting that this component was not highly endorsed as ‘essential’ for the CF of auditory hallucinations, although it was highly endorsed by >90% of the panel as ‘essential’ or ‘important’ (see Appendix Q).

One of the components that was highly endorsed as essential for the CF of *both* auditory hallucinations, and persecutory delusions, related to protective factors and the use of helpful coping strategies. Protective factors incorporate SU strengths into the CF, and it is presupposed that everyone has areas of strength and resource (Kuyken et al., 2009). Protective factors have been described as “*all that is right with the person*” (Dudley & Kuyken, 2014, p. 19), and Kuyken et al. (2009) have argued that a strengths-focused approach should be imbedded within each stage of the formulation process.

Findings from one of our previous studies has shown that voice hearers have a natural tendency to use avoidance-based coping, to escape from their voices (Howard et al., 2012). However, some SUs may in fact be using some methods of coping that are helpful, and these strategies should be highlighted as strengths within earlier formulations. For those SUs not utilising helpful coping strategies, it is possible to incorporate SU values into the problem area to use as a springboard to implement helpful ways of coping (Kuyken et al., 2009). For example, the SU that values their physical health is likely to identify with the importance of self-nurture (such as physical exercise, eating well, and getting a good night’s sleep) as a means of coping with troublesome voices, or ruminative thoughts about persecution. Formulation diagrams collaboratively put together mid, and/or at the end of therapy, can serve to remind individuals of their newly adopted coping strategies, and we have previously argued that focused-based (metacognitive) coping strategies such as acceptance, mindfulness, and meditation are the most helpful (Turkington et al., 2016).

Accepting hallucinatory experiences rather than trying to suppress voices, or push them away, is aided by normalising and educating the SU about famous/successful voice hearers that chose to accept their voices, such as Anthony Hopkins, and Beethoven (Turkington et al., 2016). The person with persecutory beliefs, might learn to ‘accept’ that the persecutors have been ‘after them for years’ but they face their fears and decide to go out anyways. Mindfulness-based interventions can also be implemented safely if taught by an experienced clinician (Böge et al., 2021) whereby the SU learns how to observe the voices, or paranoid thoughts, without judging, escaping, or trying to engage with them (Turkington et al., 2016).

Giving up ineffective coping strategies to practise new, more effective strategies, can be highlighted in later ‘post therapy’ formulation diagrams. This can serve as a reminder to the person of ‘how far they have come’ by visually demonstrating change (see Spencer et al., 2023), and this point is revisited later, in Chapter 5. Indeed, building an awareness of coping abilities in the CF, particularly in the later stages of therapy when SUs are preparing to manage independently, can help to increase self-esteem, as well as reduce distress associated with the psychotic symptoms (Kuyken et al., 2009).

The findings from this Delphi study, also shed light on which components were considered to be less than ‘important’ in the CF of voices and persecutory delusions (such as the identification of current medical conditions/illnesses, genetic factors, and lack of adherence to antipsychotic medication; see Appendix T). However, it may be argued that attention to these particular components are more aligned to the role of the medic/prescriber, than to the role of the cognitive behavioural therapist.

3.6.1 *Strengths and limitations*

There are a number of strengths associated with this research. Perhaps the most obvious are the high levels of participant recruitment and retention. This was largely attributed to the intrinsic motivation of the experts, and H.M.S. working hard to increase engagement via frequent email correspondence. As mentioned earlier, a 70% response rate is thought to be indicative of the preservation of rigour in utilising the Delphi method, and confidence in the validity of the findings (Sumsion, 1998), and this response rate was far exceeded. We can also be confident in the quality of the experts recruited into the study as the clear definition of a ‘CBTp expert’ resulted in the recruitment of highly esteemed academics and experienced clinicians within the field that had a) written manuals of CBTp, b) conducted treatment trials of CBTp, c) taught CBTp (including formulations of CBTp), d) supervised CBTp, and e) delivered CBTp.

Another main strength of the study (which has been previously noted), related to the endorsement of the survey items (elicited via the scoping literature review), that were reviewed by the experts in stage 1, and highly endorsed as either ‘essential’ or ‘important’ in stage 2. This demonstrates high levels of face and construct validity, meaning we can be confident in the content of the survey items themselves.

Use of the five Ps framework may be seen as a limitation as it is largely atheoretical however, its strength is that it is an organisational framework that is

compatible with CBT models of auditory hallucinations, and persecutory delusions. Therefore, the five Ps framework enabled us to capture and incorporate all theoretical and empirically supported formulation components that were identified from the literature.

Given a small number of components were endorsed as essential for both auditory hallucinations, and persecutory delusions, the overall findings may be seen as the best of both worlds. We have shown elegance, clarity, and simplicity, by identifying a specific model of CF that applies to both voices *and* persecutory delusions, that also incorporates the bare essence/key ingredients that are necessary to (and compatible with) the CBT model, in terms of thoughts, feelings, and behaviours. Furthermore, parsimony was achieved by identifying the simplest and most straightforward ingredients of a CF, that are both easy to for the therapist to co-develop and communicate (Butler, 1998). Indeed, parsimonious CFs should be more easily remembered by SUs.

Overall, the proposed recommendations may lead to core guidance for the process of developing CFs, and improvements in training for clinicians that conceptualise voices, and persecutory delusions in CBT for schizophrenia spectrum disorders. These recommendations differ from information provided in existing training courses, and treatment manuals, and they are unique in that a large panel of international experts have endorsed them. Nevertheless, there are several limitations.

The most obvious limitation is that CF is a collaborative process, which is not wholly expert driven. Consensus may not have been established for many components, as it may have been difficult to gauge applicability for many of the items, without speaking to the SU themselves, and so, the essential items that were endorsed, represent the views of the CBTp experts only. Whether CBTp experts agree with one another, is perhaps less meaningful than whether the SU and therapist can agree together on the content of a formulation (Dudley et al., 2010) to explore the possibility for change (Butler, 1998).

Furthermore, as the SU was not part of the process, this may go some way to explaining why the causal (precipitating) factors were not endorsed as ‘essential’, which is an omission that *may* be seen to limit the clinical utility of the findings. Similarly, the experts did not endorse a typical list of presenting issues (for example, low mood, low self-esteem, change in educational or vocational performance) which, in clinical practice, would usually be identified and formulated in collaboration with the SU.

Instead, H.M.S. semi-predefined the presenting issues, by asking the experts to think about the essential ingredients of a CF for auditory hallucinations, or persecutory delusions. This was seen a purposeful compromise to enable the research to focus in on a CBTp CF (rather than a generic formulation, or a cognitive formulation of a transdiagnostic set of issues). Indeed, this trade off may have artificially constrained and reduced the ecological validity of the findings. Nevertheless, there are challenges when using experimental methodology, and in attempting to standardise an idiosyncratic process for the purposes of clinical research. Perhaps this provides some explanation as to why the only presenting issue endorsed as an essential component of the CF for both auditory hallucinations, and persecutory delusions was that of risk to self and/or others. Without speaking to the SU, the therapist could not have known what other presenting issues might be ‘essential’ to include. Whereas the identification of risk is applicable to all individuals that access mental health services.

Another limitation *may* be seen as the focus on ‘essential’ ingredients only, which excluded many ‘important’ items, for example precipitating factors which include core aspects of the stress-vulnerability model which asserts that acute or chronic stress can precipitate or exacerbate psychosis (Clements & Turpin, 1992). Also, predisposing factors, including childhood trauma which did not reach consensus as an essential component, despite a causal (dose-response) relationship to voice hearing (Daalman et al., 2012; Read et al., 2005). As such, H.M.S. recognises the need for flexibility. Indeed, the key message conveyed here, is that therapists should focus on the absolute ‘essential’ ingredients that were identified by the experts (therapists should not start out with over-complicated CFs) however, this does not mean that these essential ingredients are the *only* ingredients that should be formulated. If the essential components do not adequately explain the SU’s formulation, or help to facilitate change, then therapists can widen their net to include other ‘important’ five P ingredients. CF is a dynamic, evolving, and personable process that can be adapted over time, and so this model should not be seen as something that is seen to constrain and limit what the formulation process can do over the course of treatment. As such, these findings do not preclude additional ‘important’ components being included in the CF (see Appendix Q) when they are salient (or essential) to the individual.

Furthermore, findings from the qualitative study (see Chapter 5) suggest that SUs often struggle to remember their CFs. So, what is conveyed to the SU may be fewer of the additional ‘important’ components (depending on the level of

simplicity/complexity the individual can assimilate), and what is formulated ‘in the therapists head’ might include more of the important components (in addition to what is viewed as ‘essential’ to incorporate).

Another limitation was that H.M.S. asked the experts to think about the essential elements that would have been formulated by the end of therapy. This does not take into consideration the dynamic nature of therapy, in that the needs of the formulation may shift from figuring out what is keeping the person stuck in a perpetuating cycle, to what components of a CF might help to keep them well in the longer term.

In addition, the survey did not differentiate highly systematised persecutory delusions, from circumscribed persecutory delusions. Whether a core maintenance CF is adequate to deal with a persecutory delusional system is questionable, and we have previously argued the need for longitudinal CFs to look back along the timeline and address underlying core schema vulnerability, when working to treat delusional systems (Turkington et al., 2015).

Finally, it may have been useful to look for any differences in consensus between European, North American, and Asian countries (and even between regions), as this would have confirmed whether formulations are likely to be developed differently and whether agreements differed or remained the same in different parts of the world; so the bias towards a UK-based sample may be seen as another limitation, but also a strength in that the UK is a leader in the field of CBTp, and so has disproportionately higher numbers of CBTp experts in comparison to other countries.

There are also several methodological limitations to consider. Firstly, the Delphi method does not offer indisputable fact, it merely represents a snapshot of cross-sectional opinion for a particular group of experts, at that moment in time (Hasson & Keeney, 2011; Thompson, 2009). As such, the ‘consensus view’, does not necessarily mean the *correct* answer or opinion has been found (Hasson et al., 2000). Similarly, the results of the Delphi method are presented on the basis of a certain level of consensus, so, depending upon how the level of consensus was defined (i.e. in this study it was predefined at >80%), a certain proportion of experts do not agree with the findings (hence in this study up to 20% of the experts did not agree with the consensus view; Niederberger & Spranger, 2020). Finally, the limited exchange of information fed back in the ‘evaluation phase’ (stage 3), suggests the convergence of opinion resulted from a normative, rather than informational influence. The experts may therefore have been swayed by others’ positions, as no individual arguments were put forward, and no new

information was introduced (Murphy et al., 1998). Hence although the Delphi method may remove many of the problems associated with counterproductive group dynamics, it may also remove several benefits associated with group interaction (such as the exchange of information, which helps to provide a rationale and understanding as to why individual decisions were made; Murphy et al., 1998).

3.6.2 *Future research*

There is a clear need for future research. Testing this CF model against observable data, would help to improve confidence in the external validity of the findings (Hasson & Keeney, 2011). As such, attempts could be made to investigate whether the Delphi method has produced the ‘right’ answer (Jones & Hunter, 1995), by applying the findings in real-world clinical settings (Powell, 2003). This could help to evaluate whether the core ‘essential’ ingredients translate to improvements in outcomes for our SUs. This is an important question, because consensus about CF content, does not convey anything about the efficacy of the formulation.

Butler stated *“it goes without saying...that a formulation which is simple, clear, and easy to understand, and therefore easy to explain, is more readily testable than one which is overly complex”* (Butler, 1998, p. 20). Therefore, one way to test out these findings, might be to randomly assign SUs to receive one of two types of CFs: 1) the CF model described here, based on the core ‘essential’ ingredients (plus any additional ‘important’ components deemed to be ‘essential’ to the individual, but with a clear focus on parsimony), versus 2) idiosyncratic CBTp formulations, whereby therapists are encouraged to formulate any of the five P ingredients they feel are necessary or appropriate to include. The primary outcome of this study could look to see whether an ‘essential’ ingredients CF model, helps to mediate better treatment effects for those with auditory hallucinations, and/or persecutory delusions. Secondary qualitative outcomes might look to see whether an ‘essential’ ingredients CF model is comprehensible, understandable, and memorable for SUs, and whether it engenders hope, and optimism for change.

Additional research could carry out a similar consensus seeking process with SUs that have received CBTp for auditory hallucinations, or persecutory delusions, which would allow for true collaboration, by highlighting what is ‘essential’ or ‘important’ in a CF for them. Future studies could also use the Delphi method to

develop expert consensus for the formulation of other psychotic symptoms, such as visual hallucinations, grandiose delusions, or negative symptoms.

Chapter 4. Examining the treatment effects of case formulation in CBTp for service users in the absence of antipsychotics: A therapeutic process evaluation of the ACTION trial

“Though being a firm believer of law of causality, I still have thousands of events I don’t have reasons for”.

– Vivek Shah (2018).

The main findings outlined in the following thesis chapter have been published in Schizophrenia Research²¹ and the reference for the publication is:

Spencer, H. M., McMenamin, M., Emsley, R., Turkington, D., Dunn, G., Morrison, A. P., Brabban, A., Hutton, P., & Dudley, R. (2018). Cognitive behavioural therapy for antipsychotic free schizophrenia spectrum disorders: Does therapy dose influence outcome? *Schizophrenia Research*, 202, 385-386.
<https://doi.org/10.1016/j.schres.2018.07.016>

This research was also presented at a conference symposium and the reference for the abstract is:

Spencer, H. M., & Turkington D. (2014, July 22 - 25). *Components of therapy as mechanisms of change in cognitive therapy for people with psychosis who choose not to take antipsychotics*. [Paper presentation]. 42nd Annual Conference and workshops, British Association for Behavioural and Cognitive Psychotherapies (BABCP), Birmingham, UK.

²¹ Impact Factor at the time of publication was 4.56.

4.1 Abstract

The ACTION (Assessment of Cognitive Therapy Instead of Neuroleptics) trial investigated CBT for people with schizophrenia spectrum disorders that chose not to take antipsychotic medication. Seventy-four participants were randomised to either CBT plus treatment as usual (TAU; $n = 37$), or TAU ($n = 37$). This secondary analysis of ACTION used principal stratification to examine whether process variables (e.g. maintenance CF, longitudinal CF) modified treatment effects, as measured by the Positive and Negative Syndrome Scale (PANSS). Instrumental variable (IV) methods were also used to estimate the treatment effects of each additional session of therapy attended on the PANSS total score. Process variables were recorded by therapists after each session. No individual process variables impacted on the treatment effects. However, it was noted that participants who received a longitudinal formulation in the first 4 sessions of therapy, had poorer treatment effects than those who did not. However, this finding was not statistically significant, $p = .17$, 95% CI [-37.24, 6.68]. A ‘dose-response’ was also found, meaning that each CBT session attended, reduced the PANSS total score by approximately 0.6 points, $p = .03$, 95% CI [-1.20, -0.06]. These findings provide a tentative signal that longitudinal CFs in CBTp may lead to poorer treatment outcomes, if explored too early on in therapy. Length of therapy also appears to be important for those that receive CBTp in the absence of antipsychotic medication. The findings were revealed in an exploratory analysis with a small sample. Larger scale studies are needed in the future to understand the therapeutic processes/components of therapy, that lead to effective treatment.

4.2 Introduction

CBTp has been shown to be effective for people diagnosed with schizophrenia and other psychoses, when combined with antipsychotic medication (NICE 2014, 2016). Several meta-analyses (and network meta-analyses) have shown robust evidence for this approach (Bighelli et al., 2018; Turner et al., 2014; Wykes et al., 2008). Furthermore, when combined with social skills training, CBT has been shown to be particularly effective in comparison to other psychological interventions (Mc Glanaghy et al., 2021).

The benefits of CBTp for schizophrenia spectrum disorders have also been demonstrated in the first pragmatic open study (Morrison, Hutton, Wardle et al., 2012), and RCT (Morrison et al., 2014) for people that chose not to take antipsychotic

medication. The RCT, known as ACTION (Assessment of Cognitive Therapy Instead of Neuroleptics), reported an estimated Cohen's *d* effect size of 0.46 (common to all follow-up time periods) on the primary outcome measure, the Positive and Negative Syndrome Scale (PANSS; Kay et al., 1987). This provided a signal for the safety, acceptability, and effectiveness of CBTp as a stand-alone intervention, in the absence of antipsychotic medication (Morrison et al., 2014).

Four years later, the COMPARE (Cognitive behaviour therapy Or Medication for Psychosis - A Randomised Evaluation) trial (Morrison et al., 2018) delivered a head-to-head three-arm comparison of CBTp only, versus antipsychotic medication (AP) only, versus a combination of both (CBT plus APs) for the treatment of first episode psychosis (FEP). Morrison et al. (2018) found that the PANSS total scores were significantly lower²² in the combined group, in comparison to the CBT-only group. However, efficacy did not differ significantly between the two monotherapies (CBT *versus* antipsychotics), or between the combined group and the AP-only group. This led Morrison et al. to conclude that whilst CBT is more effective when delivered in conjunction with antipsychotic medication, CBT delivered in the absence of antipsychotics may be just as effective when compared to antipsychotics alone, with CBT having the added benefit of fewer adverse side-effects (Morrison et al., 2018).

More recently, the MAPS (Managing Adolescent first episode Psychosis Study; Morrison et al., 2020) delivered a three-arm RCT comparison of psychological intervention (i.e. CBT, and optional family intervention) only, versus AP only, versus a combination of both (psychological intervention plus APs) for the treatment of FEP (Pyle et al., 2019). Morrison et al. (2020) found that all three treatments provided benefit in terms of lowering PANSS total scores. Again, this added further credibility to the original claims from the ACTION trial, that psychological interventions in the absence of antipsychotics do not appear to be detrimental (Morrison et al., 2020; Morrison et al., 2014).

Whilst RCTs have demonstrated the benefits of CBTp, little is known about the factors associated with a good clinical outcome (i.e. the features of treatment that are responsible for change; Kazdin, 2023). Indeed, there is a paucity of understanding about

²² The higher the PANSS total score, the greater the overall psychotic symptomatology.

the processes through which therapeutic change occurs (Holmes et al., 2018). As Emsley & Dunn (2010) once stated:

Good trials evaluating complex interventions should be able to answer both pragmatic and explanatory questions. As well as asking ‘Does it work?’, we should also be asking ‘How does it work?’, ‘What components are responsible for efficacy?’ and ‘Can it be tailored to work more effectively with particular types of patient?’ (p. 237).

Eight years later, the Lancet Psychiatry Commission for psychological treatments outlined a recommendation to better understand *how* and *why* existing psychological therapies work (Holmes et al., 2018). They argued that: “*beyond knowing that an intervention is efficacious, research initiatives are needed that clarify the key mechanisms through which interventions work*” (Holmes et al., 2018, p. 237). Indeed, this understanding may strengthen treatments such as CBT by enabling them to become more targeted and precise (Emsley & Dunn, 2010; Holmes et al., 2018). This would seem apt, particularly for complex and severe mental health difficulties, such as psychosis.

CBTp is typically offered for around nine months to augment standard treatment (Beck et al., 1979). During this time, pharmacological interventions (such as antipsychotics) may be switched or adjusted, which risks masking the impact of psychological treatment on symptoms (Morrison, Hutton, Wardle et al., 2012). Hence, exploring CBT process variables and ‘dose’ of therapy in people not taking antipsychotic medication, provided an opportunity to examine these factors, in the absence of an interaction with such medication.

‘CBT process variables’ is a term first described by Dunn et al. (2015) in the statistical literature. The term was also referred to by Flach et al. (2015) in their reporting of a secondary analysis of the Early Detection and Intervention Evaluation for people at high-risk of psychosis-2 (EDIE-2) trial. Process variables are defined as “*active intervention strategies*” (Flach et al., 2015, p. 124) as well as “*characteristics of a therapeutic intervention (...) that might influence or be associated with the efficacy of the intervention*” (Dunn et al., 2015, p. 27). Examples of process variables include a) the strength of the therapeutic alliance and b) the incorporation of pre-specified components of therapy (e.g. formulation, cognitive techniques, behavioural strategies,

the completion of homework tasks in-between therapy) that, if present, may lead to changes in clinical outcomes (Dunn et al., 2015). As Dunn et al. explained: “*The assumption is that there exists treatment effect heterogeneity and that some of this heterogeneity might be explained by these process measures*” (Dunn et al., 2015, p. 27).

Indeed, CBT is a complex intervention and its impact has been associated with a number of different factors, including: a) the therapist’s clinical experience (Steel et al., 2012), and competence (Strunk et al., 2010), b) the type of process variables utilised (such as behavioural experiments, or cognitive change techniques; Chadwick et al., 1994), c) SU characteristics (Currell et al., 2016; Freeman et al., 2002), d) use of homework (Dunn et al., 2006; Flach et al., 2015), and e) the quality of the therapeutic alliance (Gilbert & Leahy, 2007). Indeed, therapeutic alliance is thought to be an important feature of effective treatment for people with psychosis (Goldsmith et al., 2015). This, in part, may help to explain why manualised befriending as a control intervention has been found to be therapeutic, despite its apparent lack of durability (Sensky et al., 2000). A recent network meta-analysis (see Mc Glanaghy et al., 2021) also found that befriending was more effective than TAU (defined as medication with ongoing case management) for schizophrenia and psychosis. Hence, this supports an argument we have previously made - that manualised befriending is potentially an active psychological intervention, owing to the non-specific elements of the therapist-SU interaction (see Turkington et al., 2017).

Collaborative CF is also considered an important process, from which cognitive therapists undertake their whole treatment approach (Kuyken et al., 2009; Routledge Mental Health’s Podcast, 2010), and, as referred to in Chapter 2, Kuyken et al. (2005) found that greater therapists’ competence was associated with more reliable and higher quality CFs (Kuyken et al., 2005). Similarly, SUs who reported ‘sudden gains’ in CBT for depression, received therapy from clinicians who demonstrated greater competence in formulation (Abel et al., 2016). Owing to this, CF *may* be seen as a potential mechanism linked to outcome.

As explained in Chapter 1, CF is considered a fundamental process in the treatment of complex cases, such as psychosis (Dudley et al., 2011; Spencer, 2019). Whilst there are many models that serve as the basis for a CF in CBTp (e.g. Freeman, 2016; Morrison, 2001, 2017), there continues to be a notable lack of research that has

investigated whether the presence of formulation leads to a direct (or indirect²³) improvement in treatment, and treatment outcomes (Bieling & Kuyken, 2003; Kuyken, 2006). As Bieling and Kuyken (2003) once stated: “*there is limited evidence linking case formulation with outcome (...). We agree that, together with reliability and validity, this criterion should be a primary one whereby case formulation stands or falls*” (p. 60).

A recent study by Gates et al. (2021) did, however, find that the presence of a written CF in CBT for mood and anxiety disorders, was associated with a significant reduction in attrition rates, in a clinical setting. In terms of psychosis, a previous study by Chadwick et al. (2003) published more than twenty years ago, failed to show a relationship between formulation and treatment outcomes in CBTp. However, Dudley et al. (2007) found that the frequency of normalising techniques (i.e. providing an alternative, less distressing explanation for the person’s experiences) when grouped together within a broader context of formulation, predicted a good clinical outcome in a RCT of CBT for treatment-resistant schizophrenia (see Sensky et al. 2000). This led Dudley et al. to conclude that “*formulation [in CBTp] is a significant predictor of a response to treatment*” (Dudley et al., 2007, p. 568). However, the authors acknowledged that their methodology was unable to “*address questions of cause or consequence*” (Dudley et al., 2007, p. 569).

Flach et al. (2015) were the first to successfully demonstrate an empirical effect of CF on outcome (as measured by the Comprehensive Assessment for At-Risk Mental State; CAARMS; Yung et al., 2005) for people at risk of psychosis in the EDIE-2 trial. The effect of CF on outcome, was estimated by an instrumental variable (IV) analysis, and findings showed that the presence of CF (for 55% of SUs that received a CF, at some point in therapy) reduced the CAARMS total score by an estimated 23-points, $p = .03$, 95% CI [-44, -1.7]. This is an interesting finding, particularly given that the novel statistics accounted for: “*potential unmeasured confounding between the mechanism and outcome, which is an important advance on previous research*” (Flach et al., 2015, p. 126).

Prior to publishing the findings from this thesis chapter, the treatment effects of CF on outcome had never been investigated in a randomised trial of CBTp, in the

²³ For example, via the selection of appropriate interventions (Bieling & Kuyken, 2003).

absence of antipsychotic medication. Clearly, if CF does not influence and support a good clinical outcome, then its place within CBT may be overstated.

In addition to the limited research surrounding process variables (e.g. CF) as potential mechanisms of change, there has also been limited empirical support in relation to the number of CBT sessions needed to ensure a good clinical outcome. Current NICE guidelines for schizophrenia in the UK recommend that SUs should be offered at least 16 sessions of manualised CBTp (NICE, 2014). However, in many RCTs the ‘dose’ of therapy far exceeds this figure (see Morrison et al., 2014; Morrison et al., 2018). Dunn et al. (2012) found that CBTp was only effective if SUs received a full ‘dose’ of therapy to allow for the use of a wide range of active change strategies, such as cognitive and behavioural techniques. Whereas SUs that received partial therapy (consisting of only engagement and assessment), did not benefit (Dunn et al., 2012).

Given the majority of RCTs for psychosis involve participants taking antipsychotic medication, a secondary analysis of the ACTION trial provided a unique opportunity to investigate mechanisms of change in the absence of an interaction with antipsychotics, using novel analytic methods.

4.2.2 Aims

The primary aim of this study was to examine the treatment effects of CF (longitudinal, maintenance, or both), as well as a range of other process variables (e.g. homework, cognitive techniques), on the primary outcome measure, the PANSS total score (Kay et al., 1987). A second aim was to estimate the treatment effects of each additional session of therapy, on the PANSS total.

4.2.3 Hypotheses

Firstly, it was hypothesised that greater use of CF (longitudinal, maintenance, or both) would lead to larger treatment effects, as measured by the PANSS total score. Secondly, it was hypothesised that a greater number of CBT sessions attended, would lead to larger treatment effects, as measured by the PANSS total score.

4.3 Method

4.3.1 Procedure

4.3.1.1 Ethical approval

The ACTION trial protocol (see Morrison et al., 2013) was approved by the NHS National Research Ethics Service (NRES) Greater Manchester REC (reference: 09/H1014/53; see Appendix U). Local NHS Trust approvals were obtained from Research and Development (R&D) departments for Greater Manchester Mental Health NHS Foundation Trust, Northumberland, Tyne and Wear NHS Foundation Trust²⁴ (NTW), and Tees, Esk and Wear Valleys NHS Foundation Trust (TEWV).

4.3.1.2 ACTION trial design and participants

All participants recruited into the ACTION trial read the participant information sheet (PIS; see Appendix V for an example), then provided written informed consent (see Appendix W for an example). Eligible participants (aged 16-65 years) were in contact with mental health services. They either met International Classification of Diseases, Tenth Revision (ICD-10; World Health Organization, 1992) criteria for schizophrenia, schizoaffective disorder, or delusional disorder; or they met entry criteria for an Early Intervention in Psychosis (EIP) service (operationally defined by the PANSS) to allow for diagnostic uncertainty in the early phases of psychosis. At baseline entry to the trial²⁵, all participants scored at least a 4 on PANSS delusions, or hallucinations; or at least a 5 on suspiciousness, persecution, or conceptual disorganisation. In addition, participants had either discontinued antipsychotic medication for at least 6 months (whilst still experiencing psychotic symptoms); or they had never taken antipsychotic medication (and were currently choosing not to take it).

The ACTION trial participants (total $n = 74$) were randomised into one of two arms: 1) CBT plus TAU ($n = 37$), or 2) TAU ($n = 37$). Participants in both arms were similar

²⁴ Now known as Cumbria, Northumberland, Tyne and Wear NHS Foundation Trust (CNTW).

²⁵ All ACTION trial participants at the Newcastle site, were screened, recruited, and baseline assessed by H.M.S.

in age (CBT $M = 32.95$, $SD = 13.11$; TAU $M = 29.68$, $SD = 11.95$), and had similar baseline PANSS total scores (CBT $M = 70.24$, $SD = 13.75$; TAU $M = 73.27$, $SD = 13.42$). The mean duration of untreated psychosis across both arms was 92.17 months ($SD = 126.83$), and the main source of referral was from EIP services ($n = 43$; 59%). Further information regarding participant characteristics can be found elsewhere (see Morrison et al., 2013, 2014).

4.3.1.3 Intervention

Participants randomised to therapy were treated as per the Morrison (2001) cognitive model. Up to 26 sessions of CBTp were offered over a 9-month period, plus up to 4 booster sessions. Individual process variables were documented by the therapist after each CBT session²⁶ via the use of structured written ‘session records’, designed for the RCT (see Appendix X). Therapists also received training in the use of the session records; for example, they documented components of therapy whilst rating tapes of ACTION trial therapy, using the Cognitive Therapy Scale-Revised (CTS-R; Blackburn et al., 2001) to ensure fidelity to the treatment protocol (Morrison et al, 2014).

The process variables included: 1) CF (maintenance and/or longitudinal), 2) homework, 3) agreed goals, 4) cognitive techniques (e.g. normalisation, examining advantages/disadvantages, and generating alternative explanations), 5) behavioural strategies (e.g. behavioural experiments, addressing safety behaviours, reducing social isolation, and improving coping skills), and 6) metacognitive strategies (e.g. attentional strategies, and working on positive/negative meta-beliefs). Therapists were expected to use the process variables, as and when it was clinically appropriate to do so. A detailed description of the process variables can be found elsewhere in two core treatment manuals, used by therapists throughout the trial (see Kingdon and Turkington, 2005, and Morrison et al., 2004).

²⁶ A process variable was recorded by the therapist if it had been a focus of the therapy session.

4.3.1.4 *The Positive and Negative Syndrome Scale (PANSS)*

The primary outcome measure on the ACTION trial was the PANSS total score. PANSS assessments were administered by the trial RAs (e.g. H.M.S.) at baseline, then, every 3-months, up to 18-months (e.g. 3, 6, 9-, 12-, 15-, and 18-months). The end of therapy PANSS assessment was offered to all participants at 9-months. Participants recruited in the first 18-months of the study (from February 2010 to August 2011) were offered all three of the post-therapy PANSS assessments (i.e. the follow-up assessments at 12-, 15-, and 18-months). However, participants recruited after August 2011, were offered fewer/steadily reduced follow-up assessments (at the 12-, 15- and 18-months' time-points). This was owing to limited funding and resources. The ACTION trial reported an estimated Cohen's *d* effect size of 0.46. This demonstrated a small to medium effect, common to all follow-up time periods (Morrison et al., 2014).

4.4 Data analysis

Analyses for the current study were carried out using Stata version 14.0 (StataCorp, 2015).

4.4.1 *Statistical analysis 1: Principal stratification*

A novel statistical approach, based on principal stratification (Dunn et al., 2012, 2015; Frangakis & Rubin, 2002) was used to investigate the potential impact on treatment effects for each of the process variables (e.g. longitudinal CF, maintenance CF, homework, cognitive techniques), as measured by the PANSS total score. Each process variable can be thought of as a 'post-randomisation treatment effect modifier' (e.g. there may be a larger treatment effect, in the presence or absence of the process variable; Landau & Emsley, 2022). Assuming that each of the process variables can be dichotomised (which was possible for the ACTION trial, see Table 4.1 below), then principal stratification (Frangakis & Rubin, 2002) can be used to identify two latent

classes amongst the ACTION trial participants: ‘compliers’ and ‘non-compliers’²⁷, (Dunn et al. 2012, 2015).

‘Compliers’ are defined as a subgroup of participants randomised to the treatment group that received the process variable (i.e. categorised for the purposes of this analysis as having received a ‘high’ dose of the process variable, as determined by the dichotomisation; see Table 4.1 below), and who *would* have received the process variable had they, possibly counter to fact, been randomised to receive treatment (but they were randomised to the control group; Dunn et al., 2015; methodsMcr, 2013).

‘Non-compliers’ are defined as a subgroup of participants randomised to the treatment group that *did not* receive the process variable (categorised for the purposes of this analysis as having received a ‘low’ dose, as determined by the dichotomisation), and who *would not* have received the process variable had they, possibly counter to fact, been randomised to receive treatment (Dunn et al., 2015; methodsMcr, 2013).

Since the process variables are only measured in the treatment group, then class membership can only be known for those participants randomised to therapy. Class membership is *unknown* (i.e. latent) in the control group (Dunn et al., 2012, 2015). Statistical modelling was therefore used to *identify* predictors of class membership for those in the treatment group, to *predict* class membership (i.e. the ‘compliers’ and ‘non-compliers’) in the control group (see Dunn et al., 2015 for a more detailed explanation). The rationale for predicting class membership in the control group was to enable a within-class comparison of outcomes between randomised groups. ACTION was a pilot RCT with a small sample size, and so there were power limitations (Morrison et al., 2014).

Owing to randomisation, the statistical model assumes that the proportion of ‘compliers’ and ‘non-compliers’ will be the same, within the two arms of the trial. This is because, at entry to the trial, both groups on average have equal distribution of measured baseline characteristics, including class membership; the only difference between the two groups is randomisation allocation (methodsMcr, 2013). However, *“for the patients in the control group, we needed to predict whether they had high/low*

²⁷ It is acknowledged by H.M.S. that the terms ‘complier’ and ‘non-complier’ are not user-friendly. Referring to ‘non-compliers’ suggests that some participants were non-compliant with treatment; when in fact, it may be that they were not actually *offered* the process variable. Nevertheless, these terms are used throughout the statistical literature and so they are adopted here for the purposes of consistency. However, alternative terms such as ‘recipient’ and ‘non-recipient’ would have been preferable.

levels [of the process variable], based on their largest probability” (R. Emsley, personal communication, April 12, 2024).

The next step was to estimate an intention to treat (ITT) effect at 18-months (i.e. 6-months post-therapy) for the subgroup of participants who were defined as having a particular process variable present in their therapy sessions, as per the dichotomisation (i.e. those defined as having a ‘high’ dose of the process variable). An ITT effect at 18-months was also estimated for the subgroup of participants who *did not* have a particular process variable present in their sessions, as per the dichotomisation (i.e. those defined as having a ‘low’ dose of the process variable). This allowed for comparison of ITT effects to see which subgroup of participants had a better outcome (Dunn et al., 2012, 2015).

Analytical decisions in relation to the dichotomisation of each process variable was based on consultation with the Chief Investigator (Prof. Anthony Morrison) in relation to how the ACTION trial therapists had been directed, as per the study protocol. Table 4.1 below, shows how each process variable was dichotomised.

Logistic regression was then used to evaluate the relationship between the following baseline predictor variables: 1) site (Manchester or Newcastle), 2) age, 3) gender, 4) baseline PANSS total score, 5) education, and 6) medication (naïve or discontinued). No baseline variables were specified *a priori* as likely to predict the presence (i.e. ‘high’, or ‘low’ levels) of a particular process variable (as per the dichotomisation). Owing to this, we used a least absolute shrinkage and selection operator (lasso) variable selection method (Tibshirani, 1996).

Table 4.1 Dichotomisation of each process variable

Process variable	How was it dichotomised?
Longitudinal formulation	Present if it featured in at least one of the first four sessions (here, we also included those who had less than four sessions of therapy).
Maintenance formulation	Present if it featured in at least one of the first four sessions (here, we also included those who had less than four sessions of therapy).
Homework	Present if homework was completed for >50% of sessions.

Agreed goals	Present if goals were agreed for >50% of sessions.
Cognitive techniques	Cognitive techniques were made up of normalisation, examining advantages/disadvantages, evidential analysis, generating alternative explanation, and core beliefs. Present if at least 50% of sessions had at least one cognitive technique.
Behavioural strategies	Behavioural strategies were made up of coping, role play/skills, survey planning/review, safety behaviours, reducing social isolation, and relapse prevention. Present if at least 50% of sessions had at least one behavioural strategy.
Metacognitive strategies	Metacognitive techniques were made up of positive/negative meta beliefs, metacognitive strategies, and attentional strategies. Present if at least 50% of sessions had at least one metacognitive strategy.

This involved introducing a restriction which meant that as the number of predictors increased, the values of their coefficients must decrease. The limit forces the values of some coefficients to zero and therefore excludes these as predictors.

Once the baseline predictors had been identified (see results section 4.5.1 below), all the treatment group ‘compliers’ (i.e. those that received a ‘high’ dose of the process variable) were included in the ‘compliers’ analysis, and the ‘non-compliers’ (i.e. those that received a ‘low’ dose of the process variable) were included in the ‘non-compliers’ analysis. The control group was included in both analyses, with each control participant assigned a different weight, based on the class membership prediction model (Dunn et al., 2015). The model allowed for the assumption that missing outcome data was missing at random (MAR; Little & Rubin, 2002) i.e. assumed ignorable, conditional on the baseline variables in the model (Dunn et al., 2015).

4.4.2 Statistical analyses 2: Instrumental variable (IV) methods

This secondary analysis of ACTION also employed the use of IV regression methods (Dunn et al., 2015; Emsley et al., 2010) to estimate the treatment effects of each additional session of therapy attended, using Stata’s ivregress command (StataCorp, 2015). IV methods can be used to assess the efficacy of a treatment in the presence of

‘hidden confounding’ (Dunn & Bentall, 2007; Landau & Emsley, 2022). In causal inference, a confounder (or ‘lurking’ variable) is a variable that is not measured, that influences both the dependent variable and independent variable, causing a spurious association (Emsley & Dunn, 2010). Therefore, the IVs approach allowed for an unbiased estimation of the linear effect of sessions, on the PANSS total score at 18-months, even in the presence of hidden confounders that might account for the number of sessions attended (e.g. therapeutic alliance, which we did not measure).

Assuming an exclusion restriction (i.e. no direct effect of treatment assignment on outcome), allowed the use of treatment assignment (e.g. CBT plus TAU; or TAU) as an instrumental variable for the number of sessions attended. The participants in the control group were incorporated by setting their number of sessions at zero, as dictated by the design (Dunn et al., 2015).

The IV analysis uses a two-stage least squares (2SLS) estimation procedure (Baltagi, 2011). The first stage fitted a linear model for number of sessions on treatment assignment and the baseline PANSS score, then saved the predicted number of sessions attended for each participant. The second stage regressed the PANSS score at 18-months on the predicted number of sessions from the first stage and baseline PANSS score. Missing outcome data for the total PANSS score at 18-months was the same in both arms (20/37), therefore the IV 2SLS regression was undertaken with 34 observations ($n = 17$ from the treatment group, and $n = 17$ from the control group). Furthermore, as recommended by Dunn et al. (2015), bootstrapping the entire 2SLS procedure enabled us to provide valid standard errors, confidence intervals, and corresponding p values.

4.5 Results

4.5.1 Process variables

Table 4.2 below, shows the number of ‘compliers’ and ‘non-compliers’ in the treatment arm, in relation to each process variable.

Table 4.2 Number of ‘compliers’ and ‘non-compliers’ in the treatment arm for each process variable

Process variable	Compliers	Non-compliers
Longitudinal formulation	12 (32%)	25 (68%)
Maintenance formulation	33 (89%)	4 (11%)
Homework	15 (40.5%)	22 (59.5%)
Agreed goals	16 (43%)	21 (57%)
Cognitive techniques	14 (38%)	23 (62%)
Behavioural strategies	17 (46%)	20 (54%)
Metacognitive strategies	1 (3%)	36 (97%)

Maintenance formulation, homework, agreed goals, cognitive techniques, and behavioural strategies had no baseline predictors of class membership, and so unfortunately, they could not be investigated. The dichotomisation for metacognitive strategies did not vary sufficiently, as only one participant was classified as having metacognitive strategies present in at least 50% of their therapy sessions. As such, metacognitive strategies were also unable to be investigated. Longitudinal formulation was the only process variable that could be investigated.

In relation to the dichotomisation of longitudinal formulation, 12 participants in the treatment group *did* receive a longitudinal formulation in at least one of their first 4 sessions of therapy, and the remaining 25 participants in the treatment arm *did not* receive a longitudinal formulation in the first 4 sessions of therapy (see Table 4.2 above). The results of the principal stratification (used to examine whether longitudinal formulation modified treatment effects at 18-months) are also shown below, in Table 4.3.

Table 4.3 Results from the principal stratification analysis for longitudinal formulation

	Coefficient	Standard Error	P value	95% CI	N
ITT for participants with longitudinal formulation present in at least 1 of the first 4 sessions ('compliers')	- 6.54	11.20*	.17	(-37.24, 6.68)	26
ITT for participants with no longitudinal formulation present in the first 4 sessions ('non-compliers')	-15.28	9.78*	.50	(-25.72, 12.63)	25

* Bootstrapped standard errors.

The ITT effect was estimated by including the number of 'compliers' in the treatment group with 18-months outcome data ($n = 9$), *plus* all of the control group with 18-months outcome data ($n = 17$); total $n = 26$ for the 'compliers' subgroup. The model also included the number of 'non-compliers' in the treatment group with 18-months outcome data ($n = 8$), *plus* all of the control group with 18-months outcome data ($n = 17$); total $n = 25$ for the 'non-compliers' subgroup. As explained in section 4.4.1, the control group were used in both analytic comparisons, with each participant assigned a different weight, based on the class membership prediction model (i.e. the probability of them being either a 'complier' or a 'non-complier'; R. Emsley, personal communication, April 24, 2024).

Baseline predictors of longitudinal formulation were site, $p = .05$, 95% CI [-4.16, -0.00], gender, $p = .04$, 95% CI [-4.26, 0.13], and medication, $p = .05$, 95% CI [-4.04, 0.15]. In the 'compliers' subgroup (i.e. those that *did* receive a longitudinal formulation in at least one of their first 4 sessions; or those that *would have* received a longitudinal formulation in at least one of their first 4 sessions, had they, possibly counter to fact, been randomised to receive treatment), longitudinal formulation reduced

the PANSS total score at 18-months by 6.5 points, after adjusting for the baseline PANSS total score. However, this effect was not statistically significant, $p = .17$, 95% CI [-37.24, 6.68]. In the ‘non-compliers’ subgroup (i.e. those that *did not* receive a longitudinal formulation in the first 4 sessions of therapy; or those that *would not* have received a longitudinal formulation in at least one of their first 4 sessions, had they, possibly counter to fact, been randomised to receive treatment), longitudinal formulation reduced the 18-month PANSS total score by 15.3 points, after adjusting for baseline PANSS total score. However, this effect was also not statistically significant, $p = .50$, 95% CI [-25.72, 12.63]. These findings indicate that the subgroup of participants who received longitudinal formulation in the first 4 sessions of therapy, experienced smaller treatment effects than the subgroup of participants who did not receive any longitudinal formulation in the first 4 sessions of therapy. However, as already stated, these findings were not statistically significant.

4.5.2 Number of therapy sessions

Descriptive statistics in relation to baseline and 18-month PANSS total scores, and number of sessions, are shown below in Table 4.4. The mean number of sessions attended in the treatment arm was similar across both sites (i.e. Newcastle and Manchester). The effect of each additional session attended on outcome, estimated through the IV regression was -0.63, $p = .03$, 95% CI [-1.20, -0.06]. This meant that on average, each additional session of CBT reduced the PANSS total score by approximately 0.6 points (i.e. the ITT effect increased, with the increasing ‘dose’ of therapy). Moreover, an assumption made by a linear dose response model is that there is no ceiling level for the dose response (i.e. there would be no point at which each CBTp session would no longer reduce the PANSS total score by 0.6 points; R. Emsley, personal communication, April 12, 2024). Therefore, hypothetically speaking, a participant could score the maximum number of points on a baseline PANSS assessment (i.e. 210 points) and receive approximately 300 sessions of CBTp to reduce their PANSS to the minimum total score (i.e. 30 points²⁸).

²⁸ The PANSS has 30-items and is scored from 1 (not present) to 7 (severe).

Table 4.4 Descriptive statistics from the ACTION trial: Baseline and 18-month PANSS total scores, and number of sessions (organised by site, and randomisation allocation)

	Treatment group					Control group				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
Newcastle site										
Baseline PANSS total score	17	72.47	13.79	57	100	16	81.63	13.29	61	106
Number of sessions attended	17	13.47	9.11	2	34	16	0	0	0	0
18-month PANSS total score	7	55.71	20.01	40	93	9	80.67	17.54	43	105
Manchester site										
Baseline PANSS total score	20	68.35	13.78	36	101	21	66.90	9.67	54	93
Number of sessions attended	20	13.90	9.38	3	27	21	0	0	0	0
18-month PANSS total score	10	57	17.96	31	86	8	60.63	18.75	34	83

4.6 Discussion

This secondary analysis is the first to have investigated the therapeutic processes/components of therapy, as well as ‘dose’ of therapy, to estimate the potential treatment effects on outcome, in a randomised trial of CBTp for people that chose not to take antipsychotic medication (Spencer et al., 2018). Calculating the effect of randomised treatment allocation (i.e. the ITT effect) on outcome via a RCT, is widely considered to be the ‘gold standard’ in determining the effectiveness of an intervention (Hariton & Locascio, 2018).

Findings from the principal stratification suggest that SUs who received longitudinal formulation in the first four sessions of CBTp, experienced smaller treatment effects than SUs who did not receive any longitudinal formulation in their first four sessions. However, the findings were not statistically significant. As such, this can only be regarded as a tentative ‘signal’.

Given the high prevalence of childhood trauma we have since reported in relation to a sub-sample of ACTION trial participants (see Stevens et al., 2019), this points to the existence of a ‘traumatic psychosis’ subgroup (for definitions, see Kingdon & Turkington, 2008, and Stevens et al., 2017). The clinical implications of undertaking a longitudinal formulation in the early stages of therapy, with those experiencing traumatic psychosis, may have unearthed painful material and affect too soon. Moreover, if the therapist and SU became ‘stuck’ at this early stage (i.e. exploring the SU’s trauma history in detail) then the therapist may not have had time to focus on developing an understanding of the maintaining factors, or to generate a rationale with the SU for the use of cognitive and behavioural change strategies. Whilst CBT acknowledges and draws on the importance of the past, it is very much focused on the ‘here and now’ (Kuyken et al., 2009). Focusing too much on a person’s life history, suggests that therapy may not be effectively treating the presenting issues, to enable change to ensue. Indeed, this reflects the findings from Chapter 5, in relation to SUs experiences of longitudinal formulation in CBTp, which suggested that a focus on predisposing and precipitating factors appeared to result in SU disempowerment, which may inhibit change. Whereas, maintenance formulations were experienced by SUs as self-empowering, which may facilitate change (Spencer et al., 2023).

Secondly, using IV methods, the second part of this analysis examined whether ‘dose’ of therapy modified treatment effects. Results showed a statistical and clinically significant impact of number of sessions on outcome, meaning, that length of therapy

predicted a better treatment response. This finding makes sense given the (often) complex and comorbid nature of psychosis; in particular, when it comes to working with SUs experiencing paranoia where it takes time to build engagement (Lawlor et al., 2015). This is also an important finding, given recent trends in the literature that promote the use of brief CBTp owing to its cost saving implications, and usefulness in providing greater access to therapy (Hazell et al., 2016; Naeem et al., 2016).

4.6.1. *Strengths, limitations, and recommendations for future research*

In treatment trials for psychological disorders, non-adherence may manifest in the form of not receiving some, or all, of the prescribed dose of medication. Similarly, it may refer to the SU not attending all, or any, of the therapy sessions, or not receiving some components of therapy. In most of the literature, this is investigated by simply correlating the effect of sessions (or process variables) on outcome, which is subject to hidden confounding (Dunn et al., 2015; Landau & Emsley, 2022). Indeed, this was a main limitation of the secondary analysis of our open pilot study - with ‘hypothesised’ mechanisms of change being correlated with various outcome measures (see Morrison, Turkington et al., 2012). In contrast, the novel (and more advanced) statistical analyses presented here, offer a real strength as they prevented bias being introduced into our estimates. For example, the IV analysis accounted for the fact that those participants who attended their CBTp sessions, may have differed from those that only attended a small number of sessions²⁹.

Psychological research that aims to examine the treatment effects of components of therapy as mechanisms of change, has considerable scope to build better treatments (Brown et al., 2019). Identifying components of therapy that are responsible for the effect sizes found in RCTs such as ACTION, could lead to targeted treatments to help improve clinical outcomes for people with psychosis (Dunn et al., 2015). Nevertheless, the current research has several limitations.

Firstly, ACTION was a pilot RCT, therefore H.M.S. acknowledges the exploratory nature of the results presented here. Our decision to use the lasso method (Tibshirani, 1996) was based on work by Flach et al. (2015) who found that this performed best in the presence of potentially weak predictors of latent class

²⁹ Participants allocated to CBT received a mean number of 13.3 sessions, and the range was between 2–27 sessions (Morrison et al., 2014).

membership. Despite this, the small sample size meant that problems still arose in identifying predictors of latent class membership. Indeed, this accounts for why we were unable to explore anticipated treatment effects between maintenance formulation and outcome, for instance (Flach et al., 2015). This was disappointing, given the perceived impact of maintenance formulation as ‘a vehicle for change’ in CBTp, as highlighted by SUs in Chapter 5 (Spencer et al., 2023). Pre-specifying baseline variables that are likely to predict class membership would also have been a better alternative, than data-driven decisions. Moreover, it is argued by Dunn et al. (2015) that attempting to understand why therapies work should be built into the initial design stage of RCTs, not examined *post-hoc*. Arguably, this is a recommendation for future research.

Another limitation is the reliance on therapist self-report. Instead, H.M.S. could have independently rated audio recordings of the therapy sessions, to assess therapists’ use of the individual process variables (i.e. that the component of therapy was present, as a focus of the session). However, this is likely to have been time consuming, and beyond the limits of this thesis. In addition, other potential explanatory variables were not investigated, such as the therapeutic relationship, training and experience of the therapists, or duration of illness, as explored in our earlier pragmatic open study, and found to be significant predictors of outcome (i.e. via the use of correlation; Morrison, Turkington et al., 2012).

In addition, decisions based around the dichotomisation of the process variables may be considered arbitrary, and defining whether a process variable was ‘present’ as a focus of a therapy session, may be viewed as subjective. Moreover, the presence of a formulation within the first four sessions of therapy, does not tell us anything about its *quality*. Whether a CF is present or not, is perhaps less meaningful than whether it is accessible and understandable to the SU (Thrower et al., 2024), whether it helps to create change (Spencer et al., 2023), and whether it is simple and “*free from unnecessary details*” (McMurran & Bruford, 2016, p. 38) as recommended by the Case Formulation Quality Checklist (McMurran & Bruford, 2016).

Another limitation refers to the tightly defined Morrison longitudinal/maintenance model/CF template (Morrison, 2017), used in the ACTION trial. Whilst Morrison’s longitudinal/maintenance model/CF template can only make sense if therapists apply it to the individual, it may be criticised for not being wholly idiosyncratic (i.e. as the same [nomothetic] CF template, was applied to all SUs).

Indeed, RCTs may be criticised for evaluating the treatment utility of CF (Bruch, 2015) by applying a “*generic ‘one size fits all’ formulation that might (...) not fit the person well...*” (Eells, 2009, p. 294).

In relation to the ‘dose’ response finding, currently there appears to be no literature directly comparing brief CBTP (i.e. 6-10 sessions, delivered within a period of less than 4 months), with standard CBTP (i.e. 12-20 sessions, provided within a period of 4 to 6 months). In 2015, Naeem et al. attempted to review RCTs comparing brief CBTP, versus CBTP of a standard duration, but a systematic search (from 2013 to 2015) produced no studies (Naeem et al., 2015). Almost ten years later, this is still a barren area of investigation. Despite this, a systematic review and meta-analysis of ten RCTs investigating the efficacy of brief (<16 sessions) CBTP, found significant reductions in psychotic symptomatology at end of treatment ($d = -0.46$), and at follow-up ($d = -0.40$; Hazell et al., 2016). This is an interesting finding and aligns with participants having received a mean number of 13 sessions in the ACTION trial. Future research would do well to investigate further, the fundamental issue of a ‘dose-response’ relationship. For example, via the use of a manipulationist trial (Brown et al., 2019).

In addition, ‘dose’ of therapy may not only depend on the number of sessions, but also on the length of time allocated to each session (e.g. 90- or 60-minutes), as well as the frequency of the therapy sessions (e.g. weekly, or twice weekly). This is also a potential area for future research.

As previously discussed by Dunn and Bentall (2007), a final limitation of the IV analysis is the assumption of a linear effect of number of sessions on outcome. However, our finding does not consider the *quality* of the sessions, nor whether some sessions had greater value/impact than others. Instead, we are implicitly assuming that an average person who attends all sessions will receive a greater ‘dose’ of therapy than someone who attends fewer sessions. Nevertheless, it is plausible that each session did not reduce the PANSS total score by an equal amount. Most often, in psychotherapy, change does not occur in a linear fashion but rather suddenly, in discrete moments, as people experience marked shifts (i.e. ‘sudden gains’; Abel et al., 2016). Indeed, a meta-analysis by Shalom & Aderka (2020) found that (across psychological modalities and disorders) number of sessions was *negatively* associated with the effect sizes of sudden gains. Nevertheless, the consequence of this limitation does not affect our overall conclusion that more sessions led to an improved outcome. However, it does leave the magnitude of the recovery attributable to each session, open to challenge.

On balance, this exploratory secondary analysis of the ACTION trial must only be interpreted as suggesting potential hypotheses to be tested in a larger RCT with adequate power. For example, that randomises participants to receive differing amounts of therapy. Future larger scale studies are also needed to understand mechanisms of change in CBT to help improve outcomes for people with psychosis.

Chapter 5. Case formulation – A vehicle for change? Exploring the impact of CBT formulation in early psychosis: A reflexive thematic analysis

“It gave me direction, and then I used that to change”.

– Gary

The main findings outlined in the following thesis chapter have been published in *Psychology and Psychotherapy: Theory, Research and Practice*³⁰ and the reference for this publication is:

Spencer, H. M., Dudley, R., Johnston, L., Freeston, M. H., Turkington, D., & Tully, S. (2023). Case formulation – A vehicle for change? Exploring the impact of Cognitive Behavioural Therapy formulation in first episode psychosis: A reflexive thematic analysis. *Psychology and Psychotherapy: Theory, Research and Practice*, 96(2), 328-346. <https://doi.org/10.1111/papt.12442>

This research was also accepted for a conference symposium presentation and the reference for the abstract is:

Spencer, H., Johnston, L., Dudley, R., Freeston, M., & Turkington D. (2013, September 25 - 28). *A qualitative study exploring the personal impact of case formulation for service users, in Cognitive Behavioural Therapy for psychosis (CBTp)* [Paper presentation]. 43rd Annual Congress, European Association for Behavioural and Cognitive Therapies (EABCT) Conference 2013, Marrakech, Morocco.

³⁰ Impact Factor at the time of publication was 3.96.

5.1 Abstract

Formulation is considered a fundamental process of Cognitive Behavioural Therapy for psychosis (CBTp). However, an exploration into the personal impact of different levels of case formulation (CF) from a service user (SU) perspective is lacking; particularly for those in the early stages of psychosis. This Big Q qualitative design used semi-structured interviews with 11 participants. A reflexive form of thematic analysis (TA) was adopted. NVivo 12 computer-assisted qualitative data analysis software aided data organisation and analysis. One overarching theme ‘CF - A vehicle for change?’ was developed as a pattern of shared meaning across the dataset. Four main themes related to the overarching theme: (1) Vicious cycles: *‘I never really thought about it being me maintaining the problems’* (including one subtheme - Self-empowerment: *‘I can improve things (...) I could do something about it’*), (2) Early life experiences: *‘My experiences have shaped the person that I am, therefore, it's not my fault’* (including one subtheme - Disempowerment: *‘[My] core beliefs have been damaged’*), (3) Keep it simple: *‘Don’t push it too far over the top in case it becomes like spaghetti’*, and (4) A collaborative process?: *‘We both did it together (...) she did all, and I just watched’*. Maintenance formulations may be experienced as self-blaming, but also self-empowering, which may help to facilitate change. Longitudinal formulations may be experienced as non-blaming, but also disempowering, which may inhibit change. Simple, parsimonious, and collaborative CFs may also facilitate change. Whereas overly complex, too inclusive, and less-collaborative CFs may inhibit change. How CBTp therapists might look to improve the impact of different levels of CF for SUs in early psychosis are described.

5.2 Introduction

Early Intervention and At-Risk Mental State (ARMS) services are committed to ensuring that those experiencing a FEP, and those at high-risk of developing psychosis, have rapid access to evidence-based interventions for early treatment and prevention (NHS England, 2019; NICE, 2020).

CBTp is one such treatment (NICE, 2014). Collaborative CFs take centre stage as a daily feature of clinical practice within EIP and ARMS services (despite being under-researched), with SUs seen as active co-collaborators that are central to the process that unfolds over time (Routledge Mental Health's Podcast, 2010).

The early phase of psychosis is a period where the onset of symptoms is relatively recent, with SUs often feeling confused or puzzled by the nature of their experiences. A key function of formulation is to help create understanding of these experiences, and direct treatment, with CFs attempting to make sense of the *meaning* and *mechanisms* of psychosis. Consequently, SU reactions to the experience of CF in first episode psychosis (FEP) and in an ARMS, seem particularly important/timely to understand. This represents a shift away from viewing individuals with psychosis as 'un-understandable' (Jaspers, 1963/1997), towards listening to the voices of SUs via the use of qualitative enquiry (Hodgetts & Wright, 2007).

Over recent years, there has been a real drive towards understanding SU experiences of treatment, with the value of qualitative research more widely embraced (Clarke, 2022). Indeed, there is high demand for qualitative research (Braun & Clarke, 2021a) particularly alongside RCTs (O'Cathain et al., 2013). Qualitative research continues to be recognised as a "*prerequisite of good quantitative research, particularly in areas that have received little previous investigation*" (Pope & Mayes, 1995, p. 42). However, many qualitative researchers believe it holds value within its own right (Clarke, 2021a). Nevertheless, the impact of CF from a SU perspective is lacking, especially in relation to psychosis, and this question has not yet been explored in relation to early psychosis.

SUs have been asked about their experiences of CF in CBT for depression (Kahlon et al., 2014; Thew & Krohnert, 2015), depression/anxiety (Redhead et al., 2015), and other mental health disorders (Kannis-Dymand et al., 2021). Broadly speaking, CFs were found to be 'a helpful experience overall' (Thew & Krohnert, 2015), leading to an increase in 'understanding and/or acceptance' of their difficulties (Kannis-Dymand et al., 2021), whilst helping them to 'move forward' (Redhead et al., 2015). Nevertheless, the authors recommended that CFs should be undertaken sensitively and collaboratively, to help mitigate potential adverse reactions (Kahlon et al., 2014; Kannis-Dymand et al., 2021; Redhead et al., 2015).

In EIP, studies have suggested that SUs value psychosocial formulations as a way of making sense of the factors that contributed to the onset of their psychosis (Cairns et al., 2015; Dudley et al., 2009; Harris et al., 2012). Similarly, young people's reactions to CF in a CBT-based case-management intervention, for post-traumatic stress disorder (PTSD) in FEP, found that SUs valued making connections between their current symptoms, and past trauma. However, this was also experienced as challenging (Halpin et al., 2016).

Only two qualitative studies have explored the experiences of CF in CBTp (Chadwick et al., 2003; Pain et al., 2008). An additional study explored SU experiences of 'CBT-oriented' formulations in psychosis (although the therapy itself was described as integrative; Gibbs et al., 2020). All three studies noted that SUs had ambivalent emotional reactions to the CF process.

Chadwick et al.'s (2003) study (which used an unspecified qualitative method), explored SUs' experiences in relation to a CF diagram and an accompanying letter. They found that some SUs felt 'reassured' and 'encouraged' by their CF, whilst simultaneously finding it 'saddening', 'upsetting', and 'worrying'. Similarly, Pain et al. (2008) used content analysis to explore SUs' reactions to CF in the form of a written diagram. They reported that the meaning, emotions, and behaviour evoked by individual CFs, were found to be complex and multifarious. Using grounded theory (GT), Gibbs et al. (2020) found that CFs helped make sense of past events, and their impact on the present, whilst providing a route to change. However, SUs also described 'an array of emotions' in relation to seeing their formulation written down, with some feeling 'understood' and 'relieved', whilst others felt 'vulnerable' and 'confused'.

Overall, research investigating the impact of CF has highlighted that to share a CF with a SU is potentially a powerful experience, that may be perceived as helpful and/or distressing.

One reason to account for these mixed reactions may be that CFs have not been enquired about as a process where evolving levels build progressively over the course of therapy. Where clinically indicated, CFs often move from descriptive, to maintenance, then longitudinal levels of explanation (Kinderman & Lobban, 2000; Kuyken et al., 2009). Descriptive formulations in CBTp, such as those devised from an A-B-C model

(Ellis, 1957; see Spencer, 2019, and van de Gaag et al., 2013 for examples), offer a basic explanation of how thoughts, feelings and behaviours are linked. Maintenance formulations focus on the individual in the ‘here and now’, identifying perpetuating factors that are maintaining the problem(s) to generate a hypothesis for change (Dudley & Kuyken, 2014). In contrast, longitudinal formulations ‘look back’ over the individuals’ timeline, to develop a shared understanding about the origins and development of the psychosis, through identification of precipitating and predisposing factors (Dudley & Kuyken, 2014). These distinct levels of CF therefore raise several important questions concerning their impact on SUs. Indeed, the CFs outlined in studies by Chadwick et al. (2003), and Pain et al. (2008) were described as ‘developmental’ (longitudinal), which also included the use of maintenance cycles. Enquiring about CF as if it is one and the same, makes it harder for researchers and clinicians to pinpoint (and differentiate between) which aspects of the CF evoke adverse emotional reactions, versus which aspects of the CF are experienced positively.

5.2.1 Aims

In contrast to previous study samples (e.g. Chadwick et al., 2003; Pain et al., 2008, and Gibbs et al., 2020), the principle aim of this study was to explore the personal impact of CF for SUs that engaged with CBT for the treatment of early psychosis. A secondary aim was to explore the impact of different formulations (maintenance, and longitudinal³¹) in-keeping with the literature which proposes that CF should be researched as evolving levels (Kinderman & Lobban, 2000; Kuyken et al., 2009).

5.2.2 Research questions

The primary research question is:

³¹ We asked participants about their experiences of different levels of formulation (maintenance focused vs longitudinally focused formulations). However, some of our participants’ longitudinal formulations contained maintenance cycles, and so when asking about the impact of longitudinal formulation, we focused on predisposing, and precipitating factors. When asking about the impact of maintenance formulations, we focused on perpetuating factors.

What is the personal impact of CF for SUs that engaged with CBT for the treatment of early psychosis?

A secondary research question is:

What is the personal impact of different levels of formulation for SUs that engaged with CBT for the treatment of early psychosis?

5.3 Methodology³²

5.3.1 Design

A Big Q qualitative design was adopted (Kidder & Fine, 1987), alongside the use of semi-structured interviews. Big Q qualitative research recognises the strength and value of researcher subjectivity to ‘sculpt’, ‘shape’ and ‘co-create’ meaning (Clarke, 2021a).

5.3.2 Procedure

5.3.2.1 Ethical approval

This study was approved by the NHS Health Research Authority (HRA) Newcastle and North Tyneside 1 REC (reference: 12/NE/0219; see Appendix Y). Local NHS Trust approvals were obtained from R&D departments for NTW, and TEWV. Caldicott approvals were also obtained from NTW and TEWV for the purposes of releasing patient identifiable information from the NHS to Newcastle University, to enable participant interviews to be analysed using NVivo computer-assisted qualitative data analysis software³³.

³² Braun and Clarke explain that TA is closer to a ‘method’, but how we use TA - the design and conceptual choices that we make, turn it into a methodology. So, the correct heading is ‘methodology’ (Clarke, 2021g; Sage, 2023).

³³ NTW and TEWV Trusts did not have a licence for NVivo software, nor did they wish to purchase one.

5.3.2.2 Participant sampling and recruitment

A criterion-based purposeful sampling strategy was used to identify participants who had experience of the topic under investigation (Palinkas et al., 2015). The following inclusion criteria provided some level of homogeneity and focus, whilst also being flexible so that heterogeneity could be explored to capture a wide range of experiences: a) aged 16-65 years, b) had engaged with CBTp (delivered on a 1:1 basis by a cognitive therapist) for an ARMS or FEP, where therapy was formulation driven and psychosis was addressed on the problem list, c) had received (or had been offered) a written copy of their CF diagram, and were willing to discuss it, d) were willing to have their interview audio recorded, e) were able to provide informed consent, and f) had sufficient command of the English language.

Exclusion criteria included: a) moderate to severe learning difficulties, b) organic impairment, c) primary diagnosis of substance misuse (as the above factors may have impeded understanding of the formulation, and/or the ability to have engaged with the CF process), d) had dropped out of therapy, *prior* to experiencing the formulation process.

Guidance around sampling in reflexive TA suggest there will never be ‘completeness of understanding’ to justify an end point for stopping (Braun & Clarke, 2021b)³⁴. In reflexive TA, sample sizes reflect a) the quality of the data (e.g. richness, depth, complexity, diversity), b) the practicalities of time and resources, and c) the need to achieve balance between breadth and depth of understanding (Braun & Clarke, 2021a, 2021b). Therefore, judgements surrounding how many participants to interview, and when to stop data collection are subjective, and context-dependent (Braun & Clarke, 2022). Owing to the above, sample sizes in reflexive TA “*cannot be determined [wholly] in advance of analysis*” (Braun & Clarke, 2021b, p. 201).

³⁴ Braun and Clarke reject ‘rhetorical’ notions of the ‘saturation of themes’ having occurred to justify the termination of sampling (Braun & Clarke, 2021f; Sage, 2023), as advocated for in other qualitative methods such as GT (see Birks & Mills, 2015). Braun and Clarke argue that the concept of saturation mistakenly “*suggests completeness of understanding and a determinable, fixed point for stopping data collection*” (Braun & Clarke, 2021b, p. 202). Whereas “*there are always new theoretical insights to be made as long as data continues to be collected and analysed*” (Low, 2019, p. 131).

Points a-c listed above, reflect the rationale for the final sample size being appropriate in this study (data reporting the final sample size is outlined in section 5.4.1 below). Firstly, H.M.S. was able to gauge the richness and quality of the data whilst interviewing participants, and when transcribing the audio files verbatim. Based also on the length of each interview (for details see section 5.3.2.5 below) it was noted in the reflexive diary (see section 5.3.3.4 below) that a range of participants with a diverse set of characteristics (see Table 5.1) had been willing to discuss their personal experiences of CF at a deep level. Second, owing to the thesis time constraints and project timeline for the anticipated phase of data collection, it was agreed in supervision that recruitment for the reflexive TA study would be terminated within a certain timeframe. Hence the decision to end recruitment was primarily based on a balanced judgement of the richness of the data, and the constraints of this research being conducted as part of an unfunded doctoral research degree. Nevertheless, this was a flexible arrangement in that H.M.S. could opt to re-visit participant recruitment with a view to collecting more data if it later transpired (i.e. whilst developing the qualitative analysis) that the data were too ‘thin’ to develop rich patterns of shared meaning across the dataset (Clark, 2022). Finally, as the analysis progressed it became apparent that a good breadth and depth of understanding had been achieved in relation to participants’ experiences. Hence the size of the sample was considered appropriate to meet the aims of the study. This process aligns with methodological guidance in relation to reflexive TA that:

researchers should make an in-situ decision about the final sample size, shaped by the adequacy (richness, complexity) of the data for addressing the research question (...). Such decisions could and should be made within the process of data collection, reviewing data quality during the process – and recognising that sample size alone is not the only factor at play (Braun & Clarke, 2021b, p. 211).

Indeed “...data collection stops when the researcher has reached a sufficient or adequate depth of understanding to build a theory” (Braun & Clarke, 2021b, p. 202).

In terms of recruitment, H.M.S. presented at monthly North-East psychosis research meetings to promote the study to cognitive therapists working with early psychosis. Participant information sheets (see Appendix Z for an example) were then shared by cognitive therapists with potential participants who were invited to take part

in the study at the end of therapy (excluding any scheduled booster sessions). Written consent was obtained (see Appendix AA for an example), and all participants consented to the use of anonymised quotes in the write-up of this study.

5.3.2.3 Interview topic guide

The interview topic guide (see Appendix AB) was developed via the use of a) existing literature, b) consultation with the supervisory team (that have clinical and research expertise in the area), c) and SU consultation with an expert by experience.

The topic guide was exploratory, whilst being centred around participants' written formulation diagram(s) to help anchor the interview to the CF process e.g. '*Can you explain to me how your formulation diagram(s) were drawn up and put together?*'; '*Can you talk me through the content of this formulation?*'. Questions in the topic guide were also spontaneous and responsive to the participants developing account, combined with prompts to drill down to richer, deeper levels of meaning (Braun & Clarke, 2013) e.g. '*Tell me more about that*'; '*In what way(s)?*'.

5.3.2.4 Interview pilot and SU consultation

The interview topic guide was piloted with a SU researcher who had lived experience of CF in CBTp. Feedback led to minor amendments (such as the sequencing and wording of some of the questions). Therefore, data collected in the pilot study was included in the final analysis. PIS, and consent forms were also reviewed by the Manchester PRU SU Reference Group (SURG) and revised in response to their feedback.

5.3.2.5 Qualitative interviews

Semi-structured interviews were conducted by H.M.S. at the participants home, or mental health service. Interviews were audio-recorded using an encrypted Dictaphone and conducted in one session that lasted an average of 41 mins (range 23-55 mins). Participants were offered £20 as a thank you for their time.

5.3.3 *Thematic analysis*

5.3.3.1 *Reflexive, codebook, and coding reliability approaches to thematic analysis*

A reflexive form of TA was adopted, which uses qualitative techniques underpinned by qualitative research values within a Big Q qualitative paradigm (Braun & Clarke, 2021c, 2023; Kidder & Fine, 1987). This differs from coding reliability (e.g. Boyatzis, 1998; Guest et al., 2012) which uses a ‘mash-up’ of qualitative techniques underpinned by quantitative research values, within a small q qualitative paradigm (Braun & Clarke, 2021c; Kidder & Fine, 1987). Codebook TA (e.g. King & Brooks 2017; Ritchie & Spencer, 1994) sits somewhere in the middle, combining Big Q qualitative research values with a structured approach to coding, within a ‘MEDIUM Q’ qualitative paradigm (Braun & Clarke, 2021c).

In Big Q qualitative research, participant accounts are seen as partial, situated, contextual, and provisional (Braun & Clarke, 2023), that merely reflect (relational) moments in time (Braun & Clarke, 2021a; Clarke 2021a). Owing to this, qualitative researchers are seen as “*contextually and temporally situated practitioners*” (Braun & Clarke, 2021a, p. 23) with meaning co-created within a certain time and context. Moreover, Braun & Clarke (2022) argue that if we ourselves came back to our own reflexive TA several years later, the conclusions we might draw, would likely shift and change. Hence the knowledge that is generated in Big Q qualitative research is not designed to be replicable (Clarke, 2021b, 2021c). Replicability is a hallmark of realist positivism, and the ‘closed system’ of psychological laboratories (Pilgrim, 2020). Whereas critical realism views ontology as an ‘open system’, which does not buy into notions of the “*putative fixity of human functioning*” (Pilgrim, 2020, p. 165). As such, we can only attempt to understand the ever-changing, largely unobservable, complexities of human experience, by identifying “*tendencies, demi-regularities or patterns that connect through time*” (Pilgrim, 2020, p. 172).

In comparison, small q qualitative research incorporates aspects of the positivist paradigm whereby accuracy, reliability, and replicability of coding are seen as essential (Kidder & Fine, 1987). Use of multiple independent coders, and consensus between coders (via the use of inter-rater reliability checks) are seen as measures of quality, to

control for researcher subjectivity and bias (Braun & Clarke, 2021c, 2021d, 2023; Clarke, 2021b). However, in reflexive TA, coding is viewed as an organic, and fluid process that can change and is not fixed (Sage, 2023). Moreover, ‘accuracy’ of interpretation is viewed as a futile process that is impossible to achieve (Clarke, 2021b).

Big Q also rejects the notion of an objective, un-biased scientist, recognising the strength and value of researcher subjectivity to ‘sculpt’, ‘shape’ and ‘co-create’ meaning (Braun & Clarke, 2023; Clarke, 2021a). As Braun et al. (2022) stated: *“You’re telling a story as an analyst, and you’re making an argument in relation to the data, and if you see yourself as doing those two things, it positions you in a very different way from just simply capturing or describing what was in your data”* (p. 434).

Indeed, the assumption that qualitative researcher’s play a role in shaping the analysis, is not viewed as inherently problematic. This active self-agency is embraced as an inevitable part of the process (Sage, 2023). As such, themes are reported as having been ‘constructed’, ‘generated’, ‘crafted’ or ‘developed’, to acknowledge the active role of the researcher in theme development³⁵ (Braun & Clarke, 2023; Braun et al., 2022; Sage, 2023). Consequently, there is a need for reflexive TA researchers to ‘state their position’ and to ‘situate’ their findings (Braun & Clarke, 2023; Clarke, 2021b; see section 5.3.3.1 below).

In contrast, themes ‘emerge’, are ‘found’, ‘identified’, or ‘discovered’ in small q coding reliability approaches to TA (Sage, 2023). The researcher is positioned as someone who is passive, and somehow detached from the process of knowledge production (Braun et al., 2022). Their role is to search for, and ‘scoop out’ the themes that reside within the data (Braun & Clarke, 2020a; Clarke, 2022).

In Big Q qualitative research, the analysis of words is also not seen as reducible to statistical data (such as frequencies or percentages) as seen in small q qualitative research. For example, coding reliability TA report the number of participants that talked about a particular theme, or the proportion of the dataset that referred to a particular theme, which is not aligned with the values of reflexive TA (Braun & Clarke, 2020b; Clarke, 2021d). This is because there is an acknowledgement in reflexive TA, that the interview is participant-led; the researcher is guided by *participant* responses,

³⁵ In reflexive TA, themes do not ‘emerge’ from the data (Braun & Clarke, 2023) as suggested in other qualitative methods such as GT methods (see Birks & Mills, 2015), and Interpretative Phenomenological Analysis (IPA; see Smith et al., 2009).

and the things *they* wish to discuss. As such, the flexibility and openness of the topic guide enables participants to shape the interview and take it off in a different direction (albeit within the circumscribed limits of the research question), meaning that some topics may never be discussed, whereas new topics may come to light (Braun & Clarke, 2020a). This allows for unanticipated insights, and the development of new knowledge (Clarke, 2021a), but makes the reporting of numerical data somewhat meaningless (Braun & Clarke, 2020a). So, reflexive TA is interested in a ‘patterned-based’ response, but “*frequency does not determine value*” (Braun & Clarke, 2020a, p. 21). Themes may appear more in some interviews, but less (or not at all) in others. As such, the themes reported do not need to be the most common/recurrent themes (Clarke, 2021d). Instead, the themes reported should be the ones that best answer the research question, and the ones that ‘hang together’ to tell an interpretive story about the data (Braun & Clarke, 2013, 2021c, 2023).

In contrast, both coding reliability, and codebook approaches to TA involve the use of structured coding via the use of a fixed codebook. Themes are decided early on, prior to analysing the data, and coding is seen as a process of identifying *evidence* for the themes already outlined in the codebook (Sage, 2022). The coding frame is *applied* to the data, and multiple coders then code to the codebook. A major criticism of this approach is that themes are decided ahead of the analysis, or after some initial data familiarisation, and so tend to be superficial, and descriptive (Braun & Clarke, 2021c). In coding reliability TA, themes are also conceptualised as ‘analytic inputs’, based on ‘topic summaries’ that directly relate to the researcher’s pre-determined interview questions (Braun & Clarke, 2023). Whereas in reflexive TA, themes are conceptualised as ‘analytic outputs’ (Braun & Clarke, 2020a, 2020b). Coding reliability, and codebook TA may be also criticised for prematurely shutting down the possibility of ‘unanticipated insights’, as coding does not evolve over the course of the analysis to capture the researcher’s deepening understanding of the data, which is an important part of reflexive TA (Braun & Clarke, 2021c).

For a comprehensive discussion of the similarities and differences between reflexive TA and other patterned-based (across-case) qualitative analytic approaches (e.g. qualitative content analysis, IPA, GT, and discourse analysis) see Braun and Clarke (2020b, 2021c).

5.3.3.2 *The six-phase approach to reflexive TA*

Braun and Clarke's (2006) six-phases were flexibly engaged with (Sage, 2022). Owing to the recursive nature of the method, researchers can move backwards and forwards between phases (Clarke, 2021d). The six phases were as follows: 1) familiarisation, 2) coding, 3) generating initial themes, 4) reviewing and developing themes, 5) refining, defining, and naming themes, and 6) writing up (Braun & Clarke, 2006, 2020b; Clarke, 2021d).

Familiarisation involved H.M.S. immersing herself in the data by listening back to the interview audio recordings and transcribing the interviews verbatim in Microsoft Word. During this process, identifiable information was removed from the transcripts, and pseudonyms were assigned.

Semantic (explicit/overt) and latent (implicit/underlying) coding of the interview transcripts was a fluid process. This involved reading through the transcripts line-by-line and assigning pithy labels with linked memos to excerpts, to capture what was of interest or relevance in the data³⁶ (Clarke, 2021b; see Appendix AC for an example of a coded transcript). Coding was primarily inductive - grounded in participants' accounts. However, it was also deductive - informed by theory and existing literature. Deduction provided an 'interpretative lens' for making sense of (and orientating H.M.S. to) the data (Braun & Clarke, 2006; Clarke, 2022; Sage, 2022).

NVivo 12 (QSR International Pty Ltd, 2018) was then used to ensure a systematic approach to the analysis via: defining, memoing, searching, visualising, and the collation (clustering) of similar codes, across the data items. Codes were organised under theme headings (referred to as 'nodes' in NVivo). Nodes were then organised hierarchically into tree structures to identify patterns and relationships in the data (Bazeley & Jackson, 2013). Initially, many nodes were generated in NVivo (see Appendix AD). A process of reviewing, developing, and refining, meant that nodes were then merged/amalgamated to form rich, multi-faceted patterns of shared meaning (see Appendix AE). Other nodes were abandoned/discarded from the analysis if a) it transpired they were duplicates of the same theme, b) they did not directly relate to the

³⁶ Initially, coding took place in Microsoft Word by highlighting the data and using the 'review' and 'new comment' functions.

impact of CF, or c) they did not ‘hang together’ to tell an overall story about the analysis (Braun & Clarke, 2013, 2021c; Sage, 2023). Qualitative research produces many ‘stories’ about the data, and so this required H.M.S. to make active and deliberate choices about the data, such as reporting the themes of most interest, including those that best answer the research question (s; Braun et al., 2022; Clarke, 2021b; Sage, 2023).

Phases three to five also involved the creation of initial thematic maps. This captured the provisional themes H.M.S. was starting to develop in the preliminary stages of the analysis (see Appendices AF and AG for examples). The process of thematic mapping provides *“a tool to visualise the overall story the analysis tells, identifying boundaries around themes and connections across themes, as well as the overall structure of the analysis”* (Braun & Clarke, 2021c, p. 102). One of the latter thematic maps was developed by H.M.S. in supervision with R.D. (see Appendix AH). This provided an outline of the themes that H.M.S. anticipated would be included the final analysis, within an overarching theme of ‘change’.

Braun & Clarke suggest that when organising themes, they may be structured laterally (between three-four themes) and hierarchically (no more than three layers; Sage, 2022), but only if this helps to present the findings in this way as: *“that’s about as complex as you would ever want to go”* (Braun & Clarke, 2013, p. 231). A reflexive TA more complex than this, is indicative of an analysis that is too fragmented, or under-developed (Clarke, 2021d; 2022).

‘Overarching themes’ tend not to contain codes or data. Their purpose is to provide organisation and thematic structure, tying several themes together to form the overall analysis (Braun & Clarke, 2013; Clarke, 2021d). In contrast, individual ‘themes’ are distinctive (Clarke, 2021d) and defined as: *“Interpretative stories – rich and multifaceted patterns of shared meaning organised around a central concept or idea, created by the researcher through intense analytic engagement”* (Clarke, 2022, 12:11).

A theme has also been described as akin to ‘a multifaceted crystal’ - containing lots of observations that tell a story about a particular aspect of the data (Clarke, 2021d). Finally, ‘subthemes’ may be used to capture (and develop) an important facet of a theme, but this is not a requirement (Clarke, 2021d).

In the write-up phase, latent, rich, and multifaceted interpretations of meaning and experience were generated, developed, and then further refined (i.e. by going back to the data) to capture the ‘essence’ of each theme (Braun & Clarke, 2006, 2020a). During this final phase, a final thematic map was also created (see Figure 5.1 below).

5.3.3.3 *Ontology and epistemology*

The theoretical flexibility of the reflexive TA method means the ontological and epistemological positions³⁷ of the researcher should be explicitly stated (Braun & Clarke, 2020b).

For this study, an experiential (interpretative) orientation was adopted, underpinned by critical realist ontology (Bhaskar, 2020; Pilgrim, 2020) and contextualist epistemology (focusing on the person-in-context; Clarke, 2021e; Ushioda, 2009). These positions are all compatible with one another, and with reflexive TA (Braun & Clarke, 2021c; Clarke, 2021f).

An experiential orientation was deemed to be appropriate as it focuses on participants’ experiences, and views language as reflective of their underlying experience. The researcher attempts to come alongside participants’, to step into their shoes and see the world as they see it - ‘giving voice’ to their concerns (Clarke, 2021f).

Critical realism allowed H.M.S. to explore participants’ experiences, acknowledging the existence of an objective, external world, as well as a socially constructed world (Pilgrim, 2020). Critical realism posits that there is a ‘true’ reality, but we can only partially access this ‘truth’ because it is obscured (mediated) by social context, language, culture, social positioning etc. As such, we can never come to know reality in its ‘true’ and purest form (Clarke, 2021c; Danermark et al., 2019). People can have their own personal beliefs and understandings. However, these do not change the actual state of reality (Sturgiss & Clark, 2020). This fits with the view of formulation as a shared narrative, or story that is ‘constructed’ rather than ‘discovered’ (Harper & Spellman, 2014), and with people *creating* a narrative about their life experiences and

³⁷ Ontology refers to theories of ‘reality’, whereas epistemology refers what counts as valid ‘knowledge’ (Clarke, 2021c).

difficulties (DCP, 2011). Indeed “*while human perspectives are important, these are always ‘accounts of reality’ (...). Reality remains mind independent*” (Sturgiss & Clark, 2020, p. 143). In reflexive TA, there should be no assumption that ‘truth’ (in its purest form) resides within the data, owing to researcher subjectivity and interpretation (Clarke, 2021g). Consequently, critical realism posits that all knowledge is “*fallible and open to adjustment*” (Danermark et al., 2019, p. 19).

Contextualism views knowledge as emerging from context, and so knowledge is always seen as local, situational, and provisional. However, it does retain an interest in understanding ‘truth’ and so it has a realist dimension (Braun & Clarke, 2013; Clarke, 2021c). Contextualism argues that while no single method can ever get to the truth (in its purest form), some knowledge will be valid in certain contexts (so is context specific; Clarke, 2021c). Reflexivity can help render visible some of the contexts for knowledge production – so researchers can share some of the contexts that have shaped their research, and the knowledge that has been produced (Clarke, 2021c; Sage, 2023). A central metaphor of contextualism is the idea of ‘the person-in-context’ – we can’t study humans in isolation from the contexts that render their lives meaningful. To understand humans, we need to understand the context in which they live. So, the focus for the researcher is to stay close to people’s ‘lived experience’ (Braun & Clarke, 2021c; Clarke, 2021c).

5.3.3.4 *Reflexivity and other indicators of quality*

Guidance for the conduct of good quality TA (see Braun & Clarke, 2006) and good quality reflexive TA (see Braun & Clarke, 2021d, 2023; Clarke, 2022) have been followed in the engagement with the method. This helped to ensure methodological coherence/integrity. For example, ‘generic’ quality criteria and practices in qualitative research (infused with realist assumptions about improving accuracy) that are theoretically and methodologically incompatible with reflexive TA were avoided, such as saturation, member checking, and triangulation (Braun & Clarke, 2021a, 2021b, 2021c, 2023; Clarke, 2022; Sage, 2023). Similarly (as mentioned in section 5.3.2.1 above), quality practices associated with positivist (small q) approaches to TA were

avoided, such as inter-coder reliability, and consensus between coders (Braun & Clarke, 2021d, 2023; Clarke, 2022).

Reflexivity has been described as *the* hallmark of ‘quality assurance’ in reflexive TA (Braun et al., 2022). This is defined as the rigorous self-reflection, questioning, and interrogation of one’s role as a researcher (Braun & Clarke, 2021c). As such, Braun and Clarke (2021c, 2023) advocate for ‘knowing practice’, that researchers should *strive* to ‘own their perspectives’ (Elliot et al., 1999) to avoid unacknowledged assumptions creeping into their research (Braun & Clarke, 2023; Clarke, 2021h). Therefore, H.M.S. kept a reflexive journal to articulate her positioning, theoretical assumptions, and values. This positioning is stated below in the *first* person, as recommended in the write up of reflexivity, in reflexive TA³⁸ (Braun & Clarke, 2019a; Clarke, 2022; Sage, 2022).

Myself and my supervisors acknowledge that we think favorably about the use of CF in CBTp. We all work (or have worked) as CBT clinicians and/or researchers in the NHS. Several of us (H.M.S. and R.D.) have also worked within EIP services, and so we are positioned as clinicians and researchers that recognise the value that CBTp can bring to individuals experiencing a FEP. Our familiarity with the wider literature leads us to understand that SUs report mixed feelings about the CF process. Therefore, our own prior assumptions will have likely permeated the current analysis³⁹ (Braun & Clarke, 2021c). For example, we remained open to hearing both positive and negative experiences.

This study is part of a wider body of research that I have conducted within the area of CF in CBTp. I have not personally engaged in CBT for psychosis (or any other disorder) and so I have not experienced the process of a CF. In this sense, I am an ‘outsider researcher’ (Braun & Clarke, 2021c). Nevertheless, there were times when I identified as an ‘insider researcher’ (Braun & Clarke, 2021c) when participants spoke about difficult life experiences that I could relate to. This led me to reflect on how the impact of a longitudinal CF *might* be experienced, as I contemplated how it might have felt for me personally.

Consequently, my own positioning as a researcher *will* have also shaped the analysis (Sage, 2021, 2023). However, this was managed in several ways. Keeping a

³⁸ First person pronouns are recommended as reflexive TA researchers should engage in reflexivity in a visible and deliberative way (Clarke, 2022; Sage, 2023).

³⁹ This is viewed as adding richness and depth to our analysis, understanding where our findings sit in relation to the ‘bigger picture’ (Braun et al., 2022).

reflexive journal and engaging in regular high quality reflexive supervision, helped me to attempt to ‘bracket off’ my own preconceptions, in relation to the data (Clarke, 2022).

Nevertheless, researcher subjectivity is viewed as an inevitable part of the analytic process, and as a resource that drives the research (Braun & Clarke, 2021c). Indeed, ‘accuracy’ of interpretation is viewed as a futile process in reflexive TA, whereas *depth* of interpretation is viewed as a skill that resides within the researcher (Braun & Clarke, 2021d). Owing to this, two interview transcripts were independently coded by one of the supervisors (S.T.) for the purposes of ‘analytic enhancement’ (Sage, 2021, 2022). This brought different insights/interpretations into the analytic process (e.g. to question some of the assumptions I was making, and to highlight data I may have overlooked; Braun & Clarke, 2020b).

At times, I found the flexibility of the method challenging. I noted in my reflexive diary that *‘having worked as an RA on several RCTs, I am used to clear processes, procedures, and strict adherence to research protocols. As such, I am familiar with ‘certainty’ and having clear rules/steps to follow’*. However, Braun & Clarke emphasise that reflexive TA *“is not about perfect rule following”* (Braun & Clarke, 2021c, p. 12). Instead, the method has a fluid, recursive nature (Braun et al., 2022) and so it was a learning curve for me to tolerate uncertainty, and to trust in the process. Braun & Clarke explain that for novice Big Q qualitative researchers: *“being tolerant of uncertainty is an important skill to practise for good reflexive TA”* (Braun & Clarke, 2021c, p. 11).

Having also been trained in a positivist approach to research - where subjectivity is perceived as a threat to objectivity - the method provided a real shift in how I learnt to think about subjectivity (Clarke, 2021c). It may be argued that a small q approach to TA may have been a better ‘fit’ in terms of bridging the gap between the use of both quantitative and qualitative approaches in this mixed methods thesis. Nevertheless, I was happy to embrace a qualitative method in its purest form. Furthermore, my primary reason for choosing reflexive TA was a pragmatic one – *“most [psychology] supervisors aren’t trained in a cornucopia of qualitative methods”* (Clarke, 2022, 26:03; if indeed any qualitative methods), and I found a CBTp supervisor, who was experienced in the use of reflexive TA. As Braun & Clarke explain: *“Researchers select analytic approaches for all sorts of reasons, sometimes conceptual, sometimes*

pragmatic and sometimes because an approach is familiar and comfortable - to themselves or to their research supervisor, mentor or collaborator” (Braun & Clarke, 2020b, p. 38).

It is also argued that *“good qualitative research needs time”* (Braun & Clarke, 2021a, p. 31). The global pandemic afforded me time to spend analysing my data, whilst working from home. This (coupled with a COVID-19 university extension) gave me the reflective space that was needed, for my analysis to flourish. Braun & Clarke refer to this as the *“slow wheel of interpretation”* (Braun & Clarke, 2021a, p. 23), as opposed to producing a shallow and descriptive analysis that they (and other qualitative researchers) are highly critical of (Braun & Clarke, 2021a). Indeed, they argue: *“it can be all too easy to collect qualitative data, but far harder to find or make the time to analyse them in anything other than a superficial way”* (Braun & Clarke, 2021a, p. 32). They further argue, that as qualitative researchers *“we can often notice and understand more beyond our initial engagements with our data if we can allow ourselves time and space to sit with them”* (Braun & Clarke, 2021a, p. 33). Indeed, time is seen as a resource for reflexive TA, whereas scarcity of time is viewed as a real threat to quality (Braun & Clarke, 2021a).

Time also enabled me to ‘craft’ my use of the method (Braun & Clarke, 2021a) and to develop my ‘qualitative sensibility’ (Sage, 2022). Braun & Clarke acknowledge that their thinking around (reflexive) TA has evolved/changed over time (Braun & Clarke, 2021a), and so, the extra time afforded to me, meant that I could immerse myself in the reflexive TA literature. Indeed, it was timely (and fortuitous!) that I undertook this research at a time when Braun & Clarke were very active in terms of publishing papers, publishing a new book (see Braun & Clarke, 2021c), uploading online lectures to YouTube, holding interactive webinars online, and regularly interacting with people on social media. I valued the personal communication I had with Braun and Clarke on Twitter – being able to ask questions which they responded to, helped me to develop a clearer understanding of the method. This ensured that I did not misappropriate and misuse the method (as is often the case with researchers that have used (and published) reflexive TA research over recent years; Braun & Clarke, 2021a; Clarke, 2022; Sage, 2022).

5.4 Analysis

5.4.1 Contextualising the data

In total, 11 participants (gender, $n = 6$, 54.5% male; $n = 5$, 45.5% female; age, $M = 29$ years; range 16-41 years) were recruited. Ten participants were recruited from EIP services in the North-East of England, and one participant (who piloted the interview) had been involved in an ARMS service in the North-West of England. Collectively the participants were treated by six cognitive therapists - all accredited by the BABCP. The CBTp (number of therapy sessions completed, $M = 23$ sessions; range 3-40 sessions) was delivered face-to face. Timing of the formulation was naturalistic, with therapists formulating as and when it felt clinically appropriate to do so. The formulations themselves were based on maintenance focused cognitive models of depression (e.g. Beck et al., 1979), and/or longitudinal models of psychosis (e.g. Morrison, 2017). All participants were White British. Table 5.1 outlines the participant characteristics.

5.4.2 Overview of analysis

One overarching theme entitled: ‘CF - A vehicle for change?’ was developed as a shared pattern of meaning across the dataset. Four key themes relating to the overarching theme were generated and reported here. Themes 1 and 2 were associated with the content of the CF, and themes 3 and 4 were associated with the CF process.

The first theme related to maintenance formulations and aligned with the concept of *why now?* This theme was entitled: ‘vicious circles’, which had one subtheme: ‘self-empowerment’. The second theme related to longitudinal formulations and aligned with the concept of *why me?* This theme was entitled: ‘early life experiences’, which had one subtheme: ‘disempowerment’. The third theme related to both maintenance, and longitudinal formulations and aligned with the concept of *how to?* This theme was entitled: ‘keep it simple’. The fourth theme also related to both maintenance, and longitudinal formulations and aligned with the concept of *how to?* This theme was entitled: ‘a collaborative process?’. All four themes are outlined below and presented in Figure 5.1.

Table 5.1 Participant characteristics

Pseudonym	Age	Gender	Clinical presentation	Type of formulation discussed in interview	Hard copy of their written formulation diagram(s) physically referred to in interview?	Presenting issues described in the formulation	Number of therapy sessions received at the time of interview	When did therapy end?
Ron	37	Male	At-risk mental state	Maintenance, and longitudinal	No ⁴⁰	Persecutory beliefs	8	3 years previously
Tia	17	Female	First episode of psychosis	Maintenance (pre, mid, and post therapy)	Yes	Voices, obsessive-compulsive thoughts/behaviours	40	Receiving 1 booster session every 2 months
Ciara	34	Female	First episode of psychosis	Maintenance (pre and post therapy), and longitudinal	Yes	Rumination, frustration, persecutory beliefs	39	3 weeks previously
Katelyn	31	Female	First episode of psychosis	Maintenance, and longitudinal	Yes	Auditory hallucinations, childhood trauma, self-criticism, perfectionism, low mood	14	3 months previously

⁴⁰ Had received a copy of his formulation diagram, but it had been mislaid.

Michael	16	Male	First episode of psychosis	Maintenance, and longitudinal	No ⁴¹	Paranoid "irrational thoughts", anxiety, anger	20	4 months previously
Gary	24	Male	First episode of psychosis	Maintenance, and longitudinal	Yes	Persecutory beliefs	16	Due 1 final booster session
Dominic	26	Male	First episode of psychosis	Longitudinal	Yes	Flashbacks, childhood trauma, auditory, visual, and olfactory hallucinations, persecutory beliefs, anxiety, depression	38	Due 1 final booster session
Neil	41	Male	First episode of psychosis	Maintenance	No ⁴²	Persecutory beliefs	3 (dropout)	1 month previously
Chris	26	Male	First episode of psychosis	Maintenance	Yes	Persecutory beliefs, social anxiety	36	1 week previously
Julie	33	Female	First episode of psychosis	Longitudinal	Yes	Sexual/physical/emotional abuse, auditory hallucinations, persecutory beliefs, panic	17	Due 1 final booster session

⁴¹ Did not wish to keep a copy of the diagram (owing to concerns about what his formulation diagram contained, and whether his family might discover it at home).

⁴² Reported that the staff on the hospital ward disposed of it whilst cleaning his room.

Lucy	35	Female	First episode of psychosis	Maintenance, and longitudinal	Yes	Auditory and visual hallucinations, persecutory beliefs, anxiety	28	2 weeks previously
------	----	--------	----------------------------	-------------------------------	-----	--	----	--------------------

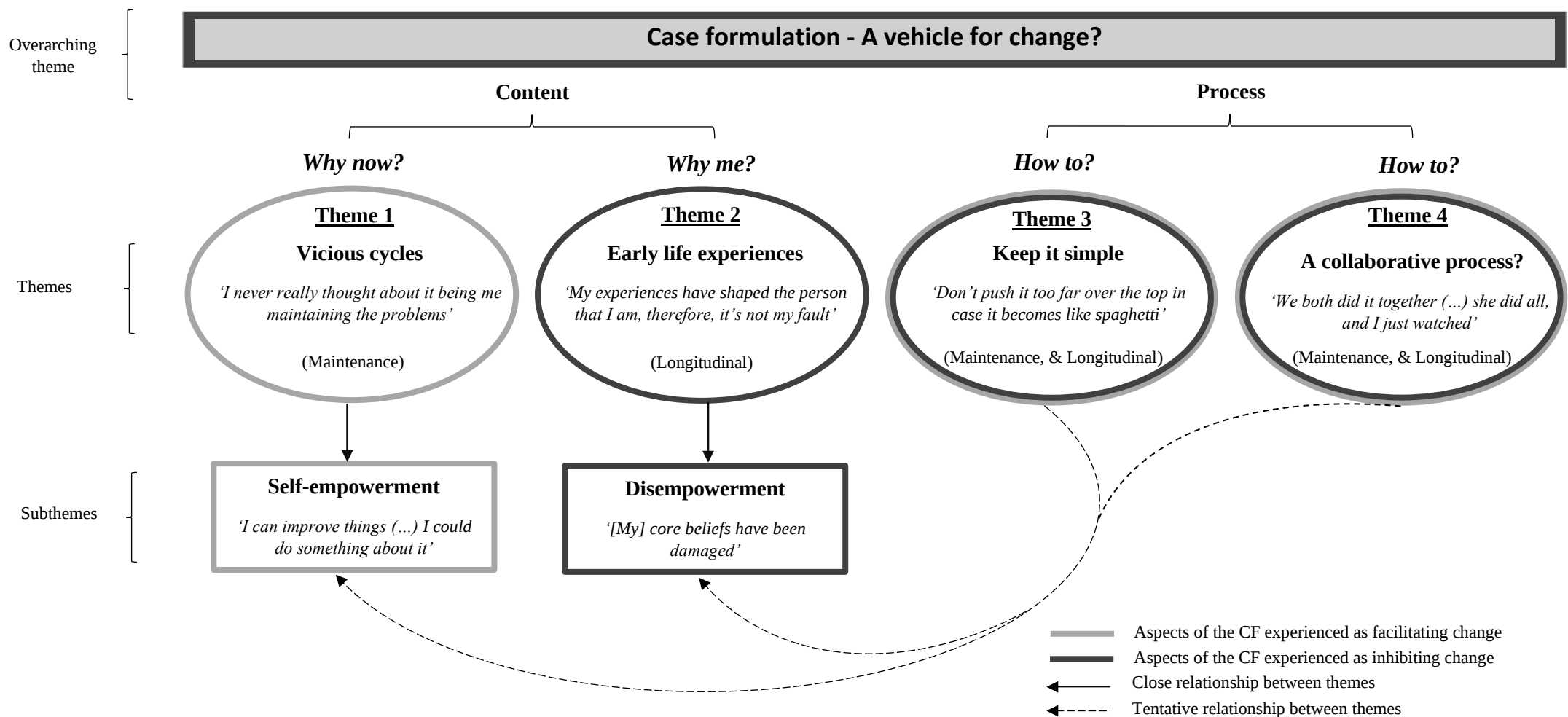


Figure 5.1 Thematic map of the overarching theme, themes, and subthemes

5.4.3 *Overarching Theme – CF - A vehicle for change?*

One overarching theme was developed to provide organisation, and thematic structure. The title of the overarching theme captures the analysis overall, by suggesting that certain levels of a CF in CBTp led SUs to experience the process as ‘a vehicle for change’, whereas other levels of the CF did not.

5.4.4 *Theme 1*

Vicious circles: ‘*I never really thought about it being me maintaining the problems*’ (Gary).

This theme was developed to capture the personal impact of the maintenance formulation which involved taking ownership, affirming self-agency (Etelämäki et al., 2021), and having an internal locus of control (LoC; Rotter, 1966).

Participants realised that certain things they were doing *themselves*, contributed to the perpetuation of their difficulties in an ongoing maintenance cycle. Often this related to behaviours such as avoidance or escape:

‘I would just leave the situation, a lot of the time (...) it was just a vicious circle. I’d keep doing that’ (Chris).

‘I’ll tell myself things [voices/visual hallucinations] are not real, but I would still run away from them, hyperventilate, have panic attacks, and it [maintenance formulation] was showing how it’s all just a vicious circle. Everything I do then makes [embarrassed laugh] more things happen’ (Lucy).

‘Vicious circle’ was a CBT phrase often used by participants. It evoked a sense of being caught/trapped in an endless negative feedback loop, with safety behaviours creating new problems that aggravated the original problem. The person-centred

language also helps us to contextualise Lucy's embarrassed laugh, in that participants were guided by the maintenance formulation to position themselves at the centre of the problem.

This was often alluded to as an 'aha!' moment of sudden insight or discovery, an understanding which the maintenance formulation afforded. For example, learning '*to be more polite, more sociable*', instead of '*avoiding people*' (which perpetuated his suspiciousness) led Michael to further state: '*I don't think I would've been able to come to that sort of epiphany by myself*'.

However, for the most part, the impact of this revelation reflected an appraisal of self-blame; a sense that the maintenance of the psychosis was *their* fault. This evoked feelings of shame, frustration, and sadness in relation to the coping strategies participants had been using: '*I made things distressing for myself (...) I was giving her [female voice] all the power and all the control*' (Lucy). '*I know that it is, it's all me (...) what I do [rumination about persecutors] it frustrates me. I know it makes me worse!*' (Ciara). '*I never really thought about it being me maintaining the problems [sounds sad]*' (Gary).

Similarly, perpetuating factors were often described as a 'bad habit'. This language appeared to be self-deprecating, as the definition of 'bad habit' refers to recurrent negative behaviour patterns, often associated with a lack of self-control (Segen's Medical Dictionary, 2012):

'It's kind of a bad thing to do actually but [sighs] my bad habit was to smoke cannabis when I was on my own, so I was kind of doubling how isolated I felt, and just increasing how often I felt strange and like the world wasn't, well I wasn't real or whatever' (Ron).

'Checking up there [the loft] every time I walk past [to check for intruders], just makes things worse (...). You know, cos you check once and you hear another noise, and you check again. And then that becomes a bad habit, doesn't it?' (Ciara).

Despite this, recognising one's own maintenance factors, was the first step towards implementing change.

5.4.4.1 *Subtheme 1*

Self-empowerment: '*I can improve things (...) I could do something about it*' (Ron).

This subtheme was generated in response to participants experiencing a cognitive shift from self-blame, towards a re-appraisal of self-empowerment.

Participants talked about the realisation that if they were maintaining their own difficulties, then they could '*improve things*' and '*do something about it*' (Ron). This involved affirming self-agency to direct one's actions toward an achievement of goals: '*I have to do it myself*' (Katelyn). '*Only **you** can make the changes for yourself*' (Ciara). Consequently, change involved thinking about things differently, and/or dropping safety behaviours, to put an end to vicious circles:

'...it was up to me that if I wanted just to look on the bright side (...) I had that choice I didn't always have to go down the negative road. I could think about things differently or you know, try out different ways of interpreting what was happening' (Ron).

'I'm in control and I need to be in control of my own life and the only way I can do that, is if I stop these [safety] behaviours that I've got myself into...' (Lucy).

As such, there was an overarching pattern of shared meaning across the dataset, that participants could actively use the maintenance formulation as a vehicle for change; a vehicle *they* were driving:

*'It gave me direction, and then I used that to change how things were (...) one example was feeling like people were looking at me. By staying with my head down that was never really challenged, so when I stopped doing that, I realised people **weren't** looking at me' (Gary).*

Furthermore, enabling people to make changes via a quick success experience, helped to maximise the impact of the process:

*'I realised "oh yeah! Well of course I can" you know, I can improve things if I just do A, B and C, really simple, practical ways (...) not just have to sit and wait for years to get better, I could actually do it myself. And so, I'd seen things change **really** quickly in terms of how I felt, and what I could do, and get on with life so...' (Ron).*

Nevertheless, when maintenance formulations brought about a sense of self-empowerment to 'improve things' (in order to break 'vicious circles'), participants realised they needed to be ready and willing to make those changes:

'I started trying to actually sort of help myself, instead of trying to get other people to help me (...) and when I did, it worked!' (Michael).

Affirming self-agency to bring about change for themselves, was also experienced as hard work and difficult. This was sometimes *explicitly* stated: *'It is hard to put it into practice'* (Lucy). *'Everything was a hardship, everything was difficult'* (Katelyn). However, participants also *inferred* that making changes by affirming self-agency to halt maintenance cycles was hard work and difficult. For example, words such as 'try' or 'trying', were used to describe change, particularly in terms of dropping safety behaviours:

*'I **try** not to listen in to people's conversations as much (...) I **try** not to leave the situation'* (Chris).

*'...**trying** to not check [the loft for intruders]'* (Ciara).

Despite this, participants expressed an understanding that making changes was associated with progress and recovery, even though it required a real sense of commitment, determination, and bravery on their part:

'Being outside is difficult, but it's really good for me' (Lucy).

'After a while the anxiety does start to calm down. It's just battling through it when it's really high' (Chris).

5.4.5 Theme 2

Early life experiences: *'My experiences have shaped the person that I am, therefore, it's not my fault'* (Katelyn).

This theme was generated to capture the personal impact of the longitudinal formulation, specifically the realisation that participants' earlier life experiences had contributed to the development of their psychosis.

The longitudinal formulation helped participants make links between the past and present:

'...how my past had been, and how that had shaped my future (...) I understand where you get from your early experiences to how you feel now' (Katelyn).

Specifically, participants were able to make links and connections between adverse life events and their presenting psychotic symptoms, finding parallels and associations between the two:

'I thought there was people in the walls in the house, cameras watching me, people in the loft (...) we'd had a burglary and I still hadn't kind of recovered from that and that's I think part of it - having people in my house taking things, was maybe a bit related...' (Ciara).

'...things that I feel guilt and shame about from being a child, have affected the way perhaps my voices speak to me' (Lucy).

Furthermore, the impact of the CF enabled participants to understand that an accumulation of predisposing and precipitating factors, had contributed to the development of their psychosis:

'I've had some really shit life experiences. I was mugged and attacked on various occasions within a short space of time (...) I'd reached a kinda crisis point after so many stressful things, that I'd had a bit of a breakdown' (Ron).

Such ascriptions brought about perceptions of an external LoC, with the longitudinal formulation often affirming that participants had been victims of external forces. As such, there was a shared narrative of non-blame - a sense that the development of their psychosis was *not* their fault:

'I understand that my experiences have shaped the person that I am, therefore, it's not my fault' (Katelyn).

Nevertheless, these aspects of the CF did not appear to facilitate change; namely because ‘the past’ cannot be amended.

5.4.5.1 Subtheme 2

Disempowerment: ‘[My] core beliefs have been damaged’ (Dominic).

This subtheme was developed in response to the personal impact of the longitudinal formulation, specifically the realisation that historic (often traumatic) life experiences, had shaped, moulded, or harmed the individual in some way.

Formulations that focused solely on predisposing and precipitating factors, appeared to reduce one’s sense of self-agency, resulting in disempowerment.

Dominic’s longitudinal formulation led to an appraisal that his core beliefs were: ‘*built...like the main frame of a computer*’ (Dominic). Also, that his core beliefs had inherently shaped him as a person: ‘*a lot of them **become** you, every decision you make*’ (Dominic). This was particularly impactful because he developed an understanding that his core beliefs had not developed as they should, owing to his earlier life experiences:

‘[My] core beliefs have been damaged or manifested in a way that it's negative and nothing else (...) That thing that matters so much is because of core beliefs, them were all negative’ (Dominic).

Another computer analogy illustrated Dominic’s perception that his earlier life had predetermined his life as an adult. This sounded defeatist, with a strong sense of stuck-ness:

‘...all the [traumatic] experiences I’ve went under (...) and then a kid like me is programmed, and that’s locked in, and then it affects everything I will become, or what I am’ (Dominic).

As the excerpt above shows, the statement *'I've went under'* appeared to have negative connotations. Dominic didn't describe going 'through' these experiences, or 'rising above' them. Instead, he *'went under'*.

Looking back over negative aspects of one's life history also unearthed self-stigmatising thoughts, which discouraged hope of transformative change:

*'a normal person - my husband, doesn't have all of this going on in his head. He doesn't have bad experiences from childhood (...) he doesn't have **rules** [for living] like what I have'* (Katelyn).

Participants also repeatedly referred to their earlier life experiences as 'damaging'. This evoked a real sense of being broken, and feelings of sadness/hopelessness:

'It's [longitudinal formulation] making me understand that the way other people have treated me like say when they hit me, or I was sexually abused, or drowned. I started to understand (...) that's so bad and that's upset me so much and damaged me so much. Someone like me, all your history's bad' (Dominic).

'I've had some very erm what you might call damaging experiences through life (...) things that in different ways were really quite damaging to the way I'd grown up, and how I thought about myself in very negative ways' (Ron).

Furthermore, the 'quick' process of the maintenance formulation contrasted with the way participants talked about the longitudinal formulation. This sounded challenging, owing to the length of time taken to construct (and emotionally process) the content of the timeline:

‘...it’s more of a long-term thing, it takes a bit longer to work that out and work out how you can learn to feel differently about the past’ (Gary).

5.4.6 Theme 3

Keep it simple: ‘Don’t push it too far over the top in case it becomes like spaghetti’ (Dominic).

This theme applied to both longitudinal, and maintenance formulations. It was developed to highlight the personal impact of participants engaging in (and seeing) their CF written down in a simple, parsimonious diagram, which inadvertently brought about change. In contrast, complex formulations, too inclusive of detail/information were seen as inhibiting change.

Change primarily occurred in the context of maintenance diagrams. However, it was also described in the context of longitudinal diagrams if the CF incorporated a small number of perpetuating factors (or a simple maintenance cycle). For example, as seen in the longitudinal/maintenance model provided by Morrison (2017).

Participants talked about the contrast of their psychosis feeling complex, overwhelming, and disorganised in their mind, but clear, understandable, and organised when formulated on paper: *‘...putting them [thoughts/feelings/behaviours] on paper really clarified things for me’ (Ron).* *‘Once it was all down [on paper] it was easier to think about’ (Gary).* *‘That’s quite neat in comparison to how it was in my head’ (Katelyn).* Indeed, simple CF diagrams helped to make sense of (and normalise) the complexities of psychosis:

Interviewer: *‘You said that it helped, seeing it [written] down on paper?’*

Julie: *‘Knowing that I wasn’t crazy’.*

Interviewer: *‘And how did writing it on paper help you to see that you weren’t crazy?’*

Julie: *'Like, it all came together...why I'm hearing the voices'.*

Pre-printed formulation diagram templates (with headings to guide formulation content; e.g. see Beck et al., 1979, and Morrison, 2017), were particularly powerful in normalising psychosis, as the CF diagram/CBT model was seen as generalisable to others:

'Seeing it on paper...it makes me feel more norm [shy laugh] normal (...) because you can design a formulation that people have similar...cos he [therapist] gave me the titles, which means that other people must have had the same sort of titles. Like this [formulation template] is already thought out, before I've even told him anything' (Tia).

Furthermore, formulation diagram templates helped to provide simplicity, organisation, and structure. Moreover, the process of visualising one's 'thoughts', 'feelings', and 'behaviours' in a simple formulation diagram, appeared to be therapeutic in itself:

'...seeing this, you know, simple formulation in black and white really brought me down to earth' (Ron).

The phrase *'brought me down to earth'* is interesting, as it implies that Ron's CF diagram brought him back down to reality, enabling him to regain a more realistic view or attitude about his beliefs.

Dominic also suggested that CBTp therapists should tailor the complexity of formulation diagrams so they can be easily understood, working with individuals at their current level of capability then trying to gradually extend it (i.e. within their zone of proximal development [ZPD]; Vygotsky, 1978):

‘So, I would say at first start with simple, but making sure you know the individual, and then go up to a level what you think is understandable for them. But then don’t push it too far over the top in case it becomes like spaghetti’ (Dominic).

Indeed, Kuyken et al. (2009) has also suggested that if a CF is developed at the appropriate level for the SU, then it is likely to increase their SUs’ understanding and minimise the chances of them feeling overwhelmed (Kuyken et al., 2009).

Simple diagrams also enabled participants to make clear comparisons between pre-, mid-, and post-therapy maintenance diagrams, with therapists revisiting and revising maintenance cycles over time. This had the impact of visually demonstrating to participants that changes were occurring, providing a signal of progress and recovery: *‘well I used to be like this, and now I’m like this’* (Michael). *‘It’s a reminder of how I used to be (...) to see how far I had come’* (Julie). In turn, this helped reinforce self-empowerment (and ongoing motivation/readiness for change):

‘[they gave] me a little confidence boost to not let me give up...cos sometimes you can’t really tell the difference between whether the voice is getting better or not, but seeing it [pre- and mid-therapy formulation diagrams] on paper makes you think “oh actually I have made progress” and you want to keep doing it [therapy] to get better, so you don’t give up’ (Tia).

As such, this theme appeared to have a tentative relationship with the ‘self-empowerment’ subtheme of theme 1 (see Figure 5.1).

Participants also implied that pre-, mid-, and post-therapy formulation diagrams were akin to an audit trail – a way of evaluating the outcome of therapy: *‘just to be able to see that evidence if you like...over the weeks, that*

things **were** improving' (Ron). Even for Dominic (where change had been difficult, owing to the rigidity of his core beliefs) his pre-, mid-, and post-therapy formulation diagrams appeared to be valued, as they facilitated greater transparency between himself and his therapist:

'...if someone either gets a bit better or a bit worse, either way I think this is the best way to show them [via pre-, mid-, post therapy formulation diagrams]. You've got it there [written down], despite what happens, how someone goes' (Dominic).

The fact that the maintenance formulation was visually 'connected' (with arrows that linked perpetuating factors together in a circular diagram) also helped to provide understanding, for the possibility of change:

'It was just like, maybes if you could change one thing, other things would change along with it' (Chris).

In addition, participants advocated for simple, and parsimonious formulation diagrams because they were memorable: *'I can take this [formulation diagram] with me like in my mind, whatever happens' (Ciara)*. Indeed, this was particularly important in the context of psychosis, as participants often referred to cognitive difficulties, and memory impairments:

'I feel like my brain like forgets (...) and I'm trying to wake my brain up and trying to sort it out and it's - it keeps stopping all the time' (Neil).

As such, participants talked about the value in taking away a copy of their CF diagram, to be able to remind themselves of their formulation, and what they needed to do to change things:

‘...you’re trying to think “what did [therapist] say?” You’ve got it down on paper, you go back to it and think “well actually this is what I need to do”’ (Lucy).

As the excerpt above shows, Lucy’s sense of agency (*‘this is what I need to do’*) again indicates tentative links with the ‘self-empowerment’ subtheme of theme 1 (see Figure 5.1).

Nevertheless, not all participants opted to refer to (or keep) a copy of their simple formulation diagram, possibly owing to them having ‘sealed over’ (McGlashan et al., 1976). For those participants that did not wish to integrate the psychotic episode as part of their life experience, their simple CF diagram served as a painful reminder:

‘I didn’t really like doing the diagram and talking about...and looking at it written down (...) I’m hoping to get through it and then just forget it all ever happened. She [therapist] offered me a copy [of the formulation diagram]. I didn’t want one’ (Chris).

Similarly, Michael did not see the value in having/keeping a copy of his CF diagram, because he did not want to be reminded of it. He implied that his CF diagram merely captured a moment in time – a ‘snapshot’ of his life, that he had since moved on from:

‘...it wouldn’t be relevant for me now to read that [CF diagram] because that that sort of thing doesn’t bother me anymore. I couldn’t see any point in me looking back...’ (Michael).

Other participants implicitly advocated for keeping things simple and parsimonious. For example, the impact of seeing CF diagrams that were too inclusive of detail/information felt disorganised, chaotic, and overwhelming. The implication of this was not conducive to recall, understanding, or change:

'There's masses of it (...) I think there's more [of the diagram] on the back [of the page] (...) got lots of little bits scrolled all over the place' (Lucy).

*'...there's a **lot** on here [refers to her formulation diagram]. Too much to remember all of it' (Tia).*

Similarly for Neil:

Interviewer: *'What did you think of doing a diagram like that?'*

Neil: *'I thought it was a bit "I'm confused" at the time (...) I couldn't cope with it. Too much information, you know?'*

As can be seen in Neil's excerpt above, not understanding the written CF diagram can reduce cognitive confidence. This may be further reinforced by stigmatising messages conveyed by the traditional medical model, that those with serious mental illness have a 'broken brain' (Andreasen, 1985). Indeed, if the CF reduces cognitive confidence, then SUs may drop out of therapy (which Neil did, after only 3 sessions). Consequently, this facet of the theme indicated a tentative relationship with the 'disempowerment' subtheme of theme 2 and was associated with the CF inhibiting change (as shown in Figure 5.1).

5.4.7 Theme 4

A collaborative process? ‘We both did it together (...) she did all, and I just watched’ (Julie).

This theme also applied to both longitudinal and maintenance formulations. It was generated to capture the personal impact of the CF diagram, as both a collaborative and a less-collaborative process. Collaborative formulations were perceived as facilitating change, whereas less-collaborative formulations were seen as inhibiting change.

Participants often referred to the word ‘we’ when talking about the process of drawing the written formulation diagram with their therapist. This evoked a sense of partnership/togetherness: ‘**We did diagrams**’ (Michael). ‘**We both did it**’ (Dominic). ‘**We would sit together doing it**’ (Ciara); as well as a shared understanding: ‘*Me and [therapist] did it...that’s what **we** know about her [female voice]*’ (Lucy). ‘*I understand where **we** got all the bits of information from*’ (Katelyn).

Participants that played an agentic role in the formulation process, conveyed a real sense of ownership of their diagram(s), and of the changes that ensued:

‘It was good doing it together rather than him just giving me it, because it feels better when you do something to help yourself, and I can say “I did this”’ (Katelyn).

Collaboration also strongly underpinned participants’ talk around self-empowerment, signalling readiness for change:

*‘It’s like a two-way thing, isn’t it? You’ve got to both have your involvement in it to actually make it work (...) it couldn’t be just [therapist] telling me “Oh this is what it’s like for you” cos like you’re not interested, you’re not gonna take it in. You’ve got to want to get involved, and you’ve got to **want** to understand...’* (Ciara).

As such, this aspect of the theme indicated tentative links with the ‘self-empowerment’ subtheme of theme 1 (see Figure 5.1).

Collaboration was particularly evident when participants completed certain sections for homework (such as writing down their thoughts), and/or if they were agreeable to holding the pen and drawing the CF diagram for themselves:

‘He kind of prompted me on some of these things but I did most of it, like he would give me the titles and set me off with the first few points or something and then I would carry on for homework...’ (Tia).

‘...she’d [therapist] like mention “behaviours”, and then I’d write the behaviours down, and then she’d mention “what are the safety measures?” and then we’d just connect it all together’ (Chris).

As the excerpts above suggest, therapists often directed the formulation process, like a driving instructor who knows the road or holds the map - guiding and showing individuals the way with their knowledge of psychosis and the CBT model. As such, there was a sense that the participant and their therapist had *different roles* in terms of putting the CF diagram together. At times participants talked about the therapist as if they were the expert/teacher, and in charge of the process: *‘he did tend to lead’* (Ciara). *‘It was the plan he had in his head’* (Ron). *‘He’s the one that knows...he probably knew in his head what he wanted to get out...my thoughts, my feelings, my beliefs, the rules that I have’* (Katelyn). However, there were other times when participants talked about themselves as contributing more to the CF process which sounded empowering, perhaps because they saw themselves as the ‘expert’ in their own lives:

‘I did about eighty percent of it. She [therapist] added the input...helping me along, giving me suggestions, and then I’d fill in the blanks’ (Chris).

'The formulation? I would say that I did eighty percent and [therapist] wrote it down, but I was the one that was giving him the words'
(Katelyn).

Furthermore, for some participants, the formulation was not something that was 'revealed' at the end of therapy. It was a hypothesis that was continually revised and revisited, with the therapist checking things out with the individual along the way:

*'Aye cos I **did** have to do some stuff. Just testing out some of his like, theories an' that'* (Michael).

'He was kind of asking me as we'd go along "is this right for you?". You know he wasn't just taking what I said, putting it on a bit of paper for himself and then saying, "this is what I think". He'd check it with me step by step, and say "is this right? Do you agree with this? Does that make sense that these are all linked?"' (Ron).

As such, there was an overarching pattern of shared meaning across the dataset, that collaboration was a fluid, ever-changing process, rather than a fixed state – something that was constantly re-negotiated. This may account for why collaboration (in the context of the formulation) was spoken about in contradictory/ambivalent ways, even by the same participant: *'We both did it together (...) she did all, and I just watched'* (Julie).

Similarly for Gary:

'...ahh see I didn't write them ...I suppose I did a little bit. Although it's her [therapists'] writing, I think all of these are my words. Though

*maybe not...she probably gave me most of the things - like 'safety behaviour' wasn't my word, that's what **they** call it' (Gary).*

Moreover, if therapists took on the role of writing the CF diagram, participants could not always read their therapists handwriting. This could also lead to things being included in the diagram that participants felt uncomfortable with:

'I asked him to take it [CF diagram] back. It had something in it that I that I didn't want certain members of my family seeing' (Michael).

The CF diagram also felt less collaborative if the therapist took away the original hand-written version and typed it up - 'presenting' the typed version to the participant several weeks later. The impact of the therapist 'revealing' typed versions of her CF diagrams meant that Ciara hadn't looked at (and had difficulties recalling) some of her formulation diagrams. The implication of this, is that it impeded memory for change: *'I haven't actually read through these [formulation diagrams] yet err...[long pause]'* (Ciara). For Chris, this led to speculation that his handwriting wasn't good enough, that perhaps the formulation diagram was of more benefit to the therapist (and the EIP service), because it had been typed up as a formal document to go in his medical records, rather than his therapist keeping a copy of the original:

'I think it was just cos it will be going in a file it's a lot neater typed up. Well, I think that - I'm not sure. She [therapist] might have been calling my handwriting!' (Chris).

As such, this facet of the theme indicated tentative links with the 'disempowerment' subtheme of theme 2 (see Figure 5.1).

A good therapeutic alliance appeared to be important when it came to collaboration and engagement in the CF process. Indeed, feeling able to disclose

important information to the therapist for the purposes of a reliable formulation required trust: *'I value his opinion'* (Michael). *'I don't want to say bond 'cause that's really cliché but like, he is quite laid back (...) we just gelled'* (Katelyn). *'I just thought "he's a really calm guy, he's lovely to talk to, I'm happy to open up with him"'* (Ron).

Therefore, playing an active role in putting the CF diagram together with the therapist, relied on participants being open and honest, so that crucial aspects of the formulation could be included:

'He would ask me a question like "what was your childhood like?" And I would then go through everything that I felt was appropriate at the time, and then he would write down some things. But he would ask me to pick out the most important things' (Katelyn).

However, a good therapeutic alliance was sometimes impeded by the symptoms of psychosis, with some participants finding collaboration challenging (or at times impossible), which made the CF process untenable:

'I felt paranoid talking to him, like I was scared of him [therapist]. And errr I didn't wana do it' (Neil).

Indeed, SU openness and honesty appeared to be essential for a formulation to ensue, but this may be compromised by the very nature of psychosis.

5.5 Discussion

This is the first study to explore the personal impact of different levels of CF, for SUs that engaged with CBT for early psychosis. In summary, there appeared to be key differences in the ways in which the *content* of maintenance, and longitudinal formulations were experienced. Whereas the written *process* of these different levels were experienced similarly. In addition, certain factors of the CF led SUs to experience the process as ‘a vehicle for change’, whereas other factors did not.

The impact of CBTp maintenance formulations appeared to evoke self-agency, and an internal LoC. Self-agency (defined within the context of social *cognitive* theory) refers to one’s ability to take initiative and responsibility for one’s own actions in everyday life (Etelämäki et al., 2021). Similarly, internal LoC (defined within the context of social *learning* theory) refers to the belief that outcomes in life are attributed to one’s own behaviour, or personal characteristics. Conversely, external LoC is the belief that outcomes in life are attributed to external forces such as chance, luck, or fate (Rotter, 1966).

Enhanced self-agency and internalised LoC have been linked to ‘readiness’ for treatment (Chambers et al., 2008), therapeutic change, empowerment, and progress toward personal goals (Tyler et al., 2020). Indeed, in the context of individualist societies (e.g. the UK), achievement of personal goals is highly valued. Therefore, internalised LoC is considered important (Sullivan et al., 2021).

Diminished self-agency and external LoC, have been documented in relation to psychosis – with one’s thoughts and actions misattributed to external agents believed to be controlling and/or communicating with them (Kozáková et al., 2020). The analysis in relation to the current study, indicated that maintenance formulations brought about self-agency and an internal LoC, in response to SUs understanding that safety behaviours perpetuated ‘vicious cycles’. This could be interpreted as self-blame, but SUs could also feel self-empowered for change. This supports findings from similar research - that the impact of CF in CBTp is experienced in both positive and negative ways (Chadwick et al., 2003; Gibbs et al., 2020; Pain et al., 2008).

One criticism of CBT maintenance formulations is that they are too focused on the individual. Incorporating a small number of *external* perpetuating factors within

maintenance cycles (e.g. wider societal context) may help to acknowledge that personal agency does not (and cannot) always exist in people's lives (DCP, 2011; Bakker, 2008). The intention is to minimise the individualising tendency to locate the difficulties within the person, implicitly conveying a message of blame and deficit (Johnstone, 2018). Helping SUs to discern between the things they can change versus the things they cannot change, may mitigate, and alleviate appraisals of self-blame. Furthermore, strengths-based formulations (Kuyken et al., 2009) and approaches to psychosis (McTiernan et al., 2020), may empower SUs to implement changes that feel attainable (e.g. by enabling SUs to imagine using adaptive coping strategies in areas of difficulty; Kuyken et al., 2009). This would seem important, given that change was perceived as hard work and difficult by participants in the current study, particularly in relation to dropping safety behaviours. Asking SUs to consider how adaptive their safety behaviours may be in the current context of their lives, may further help them to implement change (Gumley & Schwannauer, 2006).

In the context of longitudinal formulations, making links between earlier adverse life events and the development of psychosis was helpful for SUs, as this evoked a shared narrative of non-blame. This is interesting, as previous research has reported that some individuals feel partly or fully 'responsible' for the onset and early development of their psychosis (seemingly, in the absence of such formulations; Jones et al., 2016).

Nevertheless, too much time spent formulating predisposing and precipitating factors (i.e. participants in the current study experienced longitudinal formulation as a lengthy process), may lead to perceptions of external LoC and victimhood, and feelings of disempowerment.

Indeed, models and therapies that emphasise historical precipitants have consistently delivered underwhelming results (Nathan & Gorman, 2002; Roth & Fonagy, 2005). For example, SUs reported dissatisfaction with psychodynamic psychotherapy owing to its preoccupation 'with the past' (Nilsson et al., 2007). Furthermore, the process of working through the timeline in CBT (which helps to inform the longitudinal formulation), has been described by one SU with FEP as akin to hitting "*rock bottom*" (Tong et al., 2017, p. 570), resulting in an exacerbation of psychotic symptoms if traumatic memories were talked about "*for too long or too much*" (Tong et al., 2017, p. 570). SUs also described a temporal element in terms of

not feeling ready to ‘jump into’ sharing traumatic content with a therapist immediately (Tong et al., 2019), and this may help to explain why longitudinal formulations in the earlier stages of CBTp may lead to poorer treatment effects (see Chapter 4, and Spencer et al., 2018).

Whilst H.M.S. is not suggesting that CFs for psychosis should exclude longitudinal/developmental factors, use of a timeline (prior to formulating) appears to act as an exposure to trauma. Owing to this, SUs may reprocess a lot of trauma memories, which can worsen distress and/or exacerbate psychotic/comorbid PTSD symptoms in the short-term (Halpin et al., 2016; Tong et al., 2017). In the absence of a planned preparatory phase to scaffold/prepare SUs for this eventuality, H.M.S. concurs with other authors that advocate for CBTp CFs to focus on presenting symptoms and maintenance processes, to bring about change in the ‘here and now’ (Halpin et al., 2016; Kennerley et al., 2016). Indeed, the current study supports these findings, by suggesting that a focus on perpetuating factors is also a preference for SUs that have experienced CF in CBT for early psychosis.

Formulating a SU’s early childhood to uncover core schema/beliefs within the context of a CBT longitudinal formulation, should be handled sensitively, and with caution, to avoid unintentional harm (James, 2001). As James (2001) explains: working at the schema level “*often appears to shift the blame and responsibility concerning problems away from the patient, and backwards in time towards parents, guardians, teachers, and siblings*” (James, 2001, p. 405). The current study suggests that this may evoke an external LOC, which may be disempowering.

Shaped by early life experiences, core beliefs are believed to be long-standing and deeply entrenched (Morrison, 2007). Literature suggests that following recovery from depression, core schemas may lie dormant (proving difficult to detect; Riso & McBride, 2007). Indeed, core schemas may be re-activated by simple mood induction tasks, which suggests a latent vulnerability (Miranda et al., 1990; Miranda et al., 1998). Nevertheless, it is possible to target these beliefs in CBTp via use of schema change strategies (see Morrison, 2007). This message should be clearly communicated to SUs to mitigate possible appraisals of being ‘damaged’ or ‘broken’, as these appraisals are likely to discourage self-agency, internal LoC, and any hopes of transformative change. Indeed, formulating at the level of schemata may be particularly helpful in

understanding the maintenance of longstanding difficulties, and barriers to change (James, 2001). However, it must be reiterated (as presented in Chapter 3), that an international sample of CBTp experts endorsed core beliefs as ‘important’ components in the CF of voices and delusions, but they were not endorsed as ‘essential’ (Spencer et al., 2020).

For SUs, clear, simple, and parsimonious formulation diagrams facilitated greater understanding of their psychosis, and any associated perpetuating factors. If SUs can understand (and visually ‘see’) how vicious cycles are being maintained, then this may self-empower them for change. Conversely the opposite is true - diagrams that are overly complex, confusing, and too detailed, may be disempowering. Indeed, it is argued elsewhere in Chapter 3, that the CF process in CBT should be parsimonious, and simple as possible (Dudley et al., 2015; Spencer et al., 2020). This would seem crucial for SUs with psychosis in the early stages, that experience visual memory impairments (Smucny et al., 2020).

Finally, collaborative CF diagrams were linked to a sense of ownership, self-empowerment, and readiness for change. This also reflects therapists’ perspectives that active collaboration in CBTp is associated with SU ‘readiness to change’ (Currell et al., 2016). Moreover, research suggests that ‘genuine co-production’ - bringing together the expertise of people with lived experience and professionals on equal terms, can reduce the power imbalance within mental health services (Perkins & Slade, 2012). As Dudley et al. (2011) explained: *“when conceptualization is developed and shared in this collaborative manner, clients are more likely to: provide checks and balances to therapist errors, feel ownership of the emerging conceptualization, and thereby have a better understanding of the process of change”* (p.218).

However, findings from the current study suggest that the CF process may be experienced as less collaborative if therapists take the lead in writing the CF diagram (without checking whether the SU would like to take on this role). Similarly, the CF may feel less collaborative if therapists take away hand-written versions of the CF diagram to type it up neatly, ‘presenting’ it to the SU several weeks later. Furthermore, it may be difficult for SUs to engage collaboratively in the CF process if the therapeutic alliance is impeded by the symptoms of psychosis. A more detailed discussion of the

benefits of collaborative CFs are discussed elsewhere in this thesis (see Chapters 1 and 5).

The findings have several clinical implications. Table 5.2 provides an outline of recommendations for therapists.

Table 5.2 Clinical implications and recommendations for therapists in relation to the overarching theme, themes, and subthemes

Theme	Clinical implications and recommendations for therapists
<i>Overarching theme –</i> Case formulation - A vehicle for change?	• In the early stages of formulation, provide normalising information about psychosis to introduce the idea that change/recovery is possible. • Ensure adequate time in the pace and development of CFs, for SUs to consider the possibility of change.
<i>Theme 1 – Vicious circles</i>	• Closely attend to affect shifts (e.g. shame, sadness) to address appraisals of self-blame in relation to the use of safety behaviours. • Avoid using loaded language (e.g. “bad habits”) which may elicit self-blame. Consider alternative language such as “coping strategies” or “ways of coping” that may have initially helped, but which no longer serve a purpose (and appear to be keeping the difficulties going). • Include a small number of perpetuating factors that consider the wider social context, to mitigate self-blame.
<i>Subtheme – Self-empowerment</i>	• Highlight ways in which SUs can make changes for themselves to promote self-agency, internal LoC, and self-empowerment. • Start with a maintenance formulation to enable SUs to make changes via a quick success experience.

- Explain to SUs from the outset that CBTp will require them to make changes for themselves.
- Offer a small number of CBTp sessions initially to review the acceptability of the CF, and to see whether the SU is ready and willing to make changes.

Theme 2 – Early life experiences

- Starting with a maintenance formulation is regarded as best practice. However, a longitudinal formulation may help the SU to tell their story.
- Support SUs to make links between the past and present - finding parallels and associations between earlier life events, and their current psychotic symptoms.
- Emphasise that bad things in life are not necessarily under one's control/are not necessarily the person's fault; but this does not mean they are destined to be a victim.
- Emphasise that whilst the past may not be changed, our understanding of it (and how it affects us) in the 'here and now' *can* change.

Subtheme – Disempowerment

- Do not spend too long formulation predisposing and precipitating factors. These may be beyond the SU's control, which can reduce self-agency, and increase the sense of external LoC.

- ‘Core beliefs’ may be interpreted as something fundamental about the individual that may be flawed, broken, or unchangeable. Consider whether ‘old rules for living’ (which made sense then), versus ‘new rules for living’ (that make sense now), would be a more empowering message.
- ‘Balance’ the timeline by eliciting positive aspects of one’s life history (alongside adverse life events) to reduce feelings of brokenness, shame, and disempowerment.

Theme 3 – Keep it simple

- Ensure that CF diagrams are simple and parsimonious, so they can be easily understood.
- Use formulation diagram templates (e.g. see Beck et al., 1979, and Morrison, 2017) to aid simplicity, organisation, and structure.
- Revisit and revise maintenance diagrams over time (e.g. pre-, mid, and post-therapy) to highlight changes.
- Offer SUs a copy of their CF diagrams to take away with them, to facilitate memory for change.

Theme 4 – A collaborative process?

- Collaborative CFs may help SUs to feel self-empowered.
 - Encourage SUs to draw their own CF diagram(s), to facilitate a shared understanding and sense of ownership.
 - Foster collaboration by encouraging SUs to complete certain sections of their CF for homework.
-

-
- Check things out with the SU throughout the CF process to ensure that their CF diagram(s) are accurate, and that the content has been agreed.
 - SUs are more likely to remember the original hand-written versions of their CF diagram(s) that were collaboratively written during the therapy session(s).
-

5.5.1 *Strengths, limitations, and recommendations for future research*

Strengths of the study include the real-world representativeness of the sample, and the naturalistic timing of the CFs. The interviews also produced high quality, rich data, that enabled the contributions of all participants to be utilised. That six cognitive therapists were involved in the study, also suggests that these findings extend beyond the practice of an individual (or small number of) therapist(s).

An obvious limitation is that participants were mainly recruited from EIP, with only one participant recruited from an ARMS service. Therefore, this research is limited in its exploration of the impact of CF for those experiencing a pre-clinical state and/or at ultra high-risk of developing a psychotic disorder (Yung et al., 2021). Furthermore, only one participant was recruited that had ‘dropped out’ of therapy. Therefore, this research is limited in its exploration of the impact of CF for those that experience the formulation process, then choose to discontinue therapy.

Most participants engaged with many CBTp sessions (see Table 5.1) delivered by highly qualified therapists. However, this may not reflect the provision of CBTp that is available to other SUs experiencing an ARMS or FEP in other parts of the UK (or indeed, other parts of the world). Consequently, the impact of CF for SUs may vary depending upon the quality of the therapy provided, and the pace at which CFs are co-constructed (within the constraints of the number of therapy sessions offered).

It may also be argued that CF is not an ‘event’ that can be easily disentangled from therapy, given that the process of CF runs throughout therapy. As such, it was, at times, difficult to tease the two apart, both in interview, and whilst undertaking the analysis. The study was designed to anchor certain interview questions to the SUs’ CF diagram(s). Hence, several of the interview questions conceptualised ‘formulation-as-an-event’ (Johnstone, 2018). However, a small number of SUs did not have a copy of their CF diagram with them, that we could refer to in interview. This made this disentangling process even harder, which is undoubtedly a limitation.

A further limitation is that the therapists did not deliver manualised CBTp and so therapist adherence to the cognitive model was left unchecked. Moreover, H.M.S.

cannot report anything about treatment outcomes, and so it is possible that the impact of the CF process for SUs may have been influenced by personal perceptions of recovery. Future research could enquire about (or assess) the severity of psychotic symptoms prior to (and post) receipt of therapy, as this may have direct implications concerning the impact and experiences of the CF process.

Another limitation is that each SU was invited by their cognitive therapist to participate. As such, there may have been selection bias, with therapists approaching those SUs they felt had ‘grasped’, ‘contributed to’, or ‘engaged well’ with the CF process. It is also noted that four participants were still receiving CBT booster sessions (see Table 5.1), and so these individuals may have felt obliged to portray their CF in a more positive light, than they may have done otherwise. Despite this, several participants openly discussed the negative impact of their CF.

It is acknowledged that the analysis captured here is partial, tentative, temporal, situated, and contextual (Braun & Clarke, 2021a; Clarke, 2021b). However, the generation of all new knowledge is contingent (Smith, 2018) and so the design choices in relation to the current study, inevitably shaped (and delimited) the knowledge produced (Braun & Clarke, 2022). For example, SUs with a longer duration of psychosis, that receive CBTp via community treatment teams, or inpatient services, may experience the CF process differently. H.M.S. therefore invites the reader(s) to consider the transferability of these findings, beyond the context in which they were studied (Smith, 2018).

It is also acknowledged that a relatively small number of SUs were recruited overall. However, this is not viewed as inherently problematic as “*bigger isn’t necessarily better*” in the context of qualitative sampling (Braun & Clarke, 2016, p. 742). Smaller sample sizes allow time for intricacies and complexities to be explored in the data (Braun & Clarke, 2016). Moreover, it is noted that the sample size reported here, aligns with recommendations for the design of a ‘medium TA project’, deemed suitable for a Doctorate in the UK (see Braun & Clarke, 2021c).

H.M.S. also acknowledges the limitation of a wholly White British sample of participants, which does not permit broader inferences for the experiences of SUs from minority ethnic groups. Recruiting participants from different ethnic backgrounds may have offered different cultural perspectives on the impact of CFs, as well as helping to

determine the transferability of the experiences interpreted here. Nevertheless, a predominant White British sample is in-keeping with the local population in which this study was carried out. The Office for National Statistics reported (from an ethnicity dataset collected in 2016⁴³) that 92% of the population in the North-East of England identified as White British (Office for National Statistics, 2022).

The reflexive TA method is not without its own limitations. Qualitative methodologies such as GT (see Birks & Mills, 2015), IPA (see Smith et al., 2009) and discourse analysis (DA; see Willig, 2013) offer ‘ready-made’, ‘off-the-shelf’ (Chamberlain, 2012) packages of theory, analytic method, and (elements of) research design (Braun & Clarke, 2020b). Consequently, these methodologies do not allow for theoretical flexibility (as in reflexive TA; Clarke, 2021h) but are likely to be far easier for novice qualitative researchers to follow, without getting enmeshed in theoretical and methodological ‘mash-ups’ that are incompatible with one another (and with reflexive TA). Victoria Clarke recently explained that researchers are unknowingly: “*mashing up codebooks, coding frames, and coding reliability measures with things that we outline [in reflexive TA], because they don’t understand the conceptual foundations of the different approaches*” (Coxon, 2022, p. 41). Indeed, “*limitations and confusions remain in how TA is implemented and understood*” (Braun & Clarke, 2019b, p. 589). This is often seen in the publication of many reflexive TA studies (Braun & Clarke, 2023; Clarke, 2022). To avoid encountering common pitfalls in design incoherence (Braun & Clarke, 2020b, 2023), it would seem essential for researchers to adequately familiarise themselves with the up-to-date literature on reflexive TA, before embarking on the method.

A further limitation of reflexive TA is that the method is concerned with ‘patterns of shared meaning’ (themes) *across* cases, rather than on a focus on individual meaning *within* cases (Braun & Clarke, 2022). This differs from other patterned-based approaches such as IPA, that also seek to incorporate an idiographic orientation, by focusing on the specific details of each case (Braun & Clarke, 2020b). In addition, reflexive TA does not concern itself with language *use* (such pronouns, tone, repetition, degree of fluency etc.), imbedded within IPA. Also, reflexive TA does not concern itself with language *practice* (such as the micro, fine-grained details of language), imbedded

⁴³ Participants from the current research were recruited between 2012-2015.

within some approaches to DA. For example, conversation analysis (e.g. Madill et al., 2001), and discursive psychology (e.g. Wiggins, 2016).

Recommendations for future research include delving deeper into areas not fully explored in the current study. For example, more work might be done to ask probing questions about whether longitudinal formulations brought about change, in the absence of perpetuating factors. Several under-developed themes (not included in the write up of this study) could also be explored in future research. For example, themes that related to the ‘SU as (pseudo) expert’ - with several SUs revealing they had taken on the ‘role of therapist’ by openly formulating their friends/family members difficulties.

Finally, future research could look to assess self-agency and LoC⁴⁴ before and after the co-development of maintenance/longitudinal formulations, to test out some of the theories proposed here.

⁴⁴ For example, using the Sense of Agency Scale (SoAS; Tapal et al., 2017), and the Nowicki-Strickland Locus of Control Scales (Nowicki & Duke, 1974).

Chapter 6. General discussion

“To develop a complete mind - study the science of art; study the art of science. Learn how to ‘see’. Realise that everything connects to everything else”.

– A quote taken from Leonardo da Vinci’s notebooks (n.d.).

6.1 Recap of the thesis aims and summary of key findings

The overarching aim of this thesis (outlined in section 1.10) was to employ the use of mixed methods to further an understanding of CF, particularly in relation to CBT for the treatment of psychosis.

Addressing the specific aims of this thesis (also outlined in section 1.10), the first aim was to undertake an extensive systematic review of the CF literature. Specifically, Chapter 2 investigated the reliability of CFs in relation to various psychological disorders, across a wide range of theoretical modalities. The review updated and extended the findings of an earlier review of the CF literature, by Flinn et al. (2015). Thirty-two studies were yielded in total, with the results demonstrating a wide range of coefficient values (both within, and across studies), that related to the reliability of different *components* of CF. In general, this appeared to reflect overt/descriptive components being more reliable (and easier to formulate) than inferential components. Five different reliability methods were also identified as having been used to determine CF reliability, across the range of studies, and this is the first systematic review of CF to differentiate between these different types of reliability *methods* – an important aspect of the literature which has seemingly been overlooked in previous reviews (see Easden & Kazantzis, 2018; Flinn et al., 2015; Rainforth & Laurensen, 2014).

Owing to the disparate set of literature a meta-analysis was contraindicated. This highlighted a need for consistent approaches when investigating CF reliability, to allow for a) similar studies to be undertaken, b) greater possibility of replication of the findings, and c) a future synthesis of the field, which is lacking. A narrative synthesis was therefore undertaken, and the findings suggested that atheoretical formulations (e.g.

four Ps) produced higher levels of CF reliability, than other theoretical modalities. Methodological quality did not appear to be associated with CF reliability. Another key finding provided a signal that as the number of raters/judges increased (e.g. from one to five), the rate of reliability increased. This suggests that having two (or more) ‘heads’ involved in CF, may help. For example, to counteract problematic heuristics that can affect the clinical decision making of a single clinician (Flinn et al., 2015; Kuyken et al., 2005; Kuyken et al., 2009).

The second aim, addressed in the first empirical study, was to employ the Delphi method to investigate whether a panel of international CBTp experts could reach consensus on the essential ingredients of a CBT CF for auditory hallucinations, and persecutory delusions. Chapter 3 achieved this aim, with successful recruitment of an international panel of 78 CBTp experts from 12 different countries for the main stage (stage 2) of this study. Importantly, a high rate of participant retention was also achieved, with 70 CBTp experts responding again, in the third and final stage. This 3-stage Delphi process (Langlands et al. 2008) involved expert participants producing and rating statements that addressed key areas of CF in terms of the five Ps framework (Dudley & Kuyken, 2014): presenting issues, predisposing, precipitating, perpetuating, and protective factors.

One presenting issue and 7 perpetuating factors were endorsed as essential by >80% of the expert panel. The exact same items were endorsed for both auditory hallucinations, and persecutory delusions. These findings are unique in that a large panel of international experts reached consensus that CFs should be parsimonious and focused on the perpetuating (maintaining) factors to facilitate change (Spencer et al., 2020).

The third aim, addressed in the second empirical study, employed the use of causal inference methods to examine whether therapeutic process variables (e.g. maintenance CF, longitudinal CF) led to changes in symptomatology/features of psychosis. This secondary analysis of the ACTION trial was the first to investigate the treatment effects of CF, in the absence of antipsychotic medication. Indeed, the analyses provided a unique opportunity to investigate mechanisms of change in the absence of an interaction with such medication. A second aim of the study was to estimate the treatment effects of each additional session of therapy, on the PANSS total score. Chapter 4 used a novel statistical approach based on principal stratification (Dunn et al., 2012, 2015; Frangakis

& Rubin, 2002) to examine whether any of the process variables (e.g. maintenance CF, longitudinal CF) modified treatment effects. Secondly, IV methods (Dunn et al., 2015; Emsley et al., 2010) were used to estimate the treatment effects of each additional session of therapy attended on the PANSS total score. The analytic methods allowed for unbiased estimations, even in the presence of ‘hidden confounding’ (Dunn & Bentall, 2007; Landau & Emsley, 2022), and this is an important advance on previous research.

Findings revealed that participants who received a longitudinal formulation in the first four sessions of therapy, had somewhat poorer treatment effects than those who did not. However, this finding was not statistically significant. A ‘dose-response’ was also found, meaning, that each CBT session attended reduced the PANSS total score by approximately 0.6 points. These findings provided a tentative signal that longitudinal CFs in CBTp *may* lead to poorer treatment outcomes, if explored too early on in therapy. Length of therapy also appears to be important for those that receive CBTp in the absence of antipsychotic medication (Spencer et al., 2018).

The fourth aim, addressed in the third empirical study, employed the use of reflexive TA (Braun & Clarke, 2020b; 2021c) to explore the personal impact of CBT CF, for SUs experiencing psychosis in the early stages. A second aim of this qualitative study was to explore the personal impact of different levels of CF. Semi-structured interviews with 11 participants generated one overarching theme ‘CF - A vehicle for change?’ which was developed as a pattern of shared meaning across the dataset. Four main themes related to the overarching theme: (1) Vicious cycles: *‘I never really thought about it being me maintaining the problems’* (including one subtheme - Self-empowerment: *‘I can improve things (...) I could do something about it’*), (2) Early life experiences: *‘My experiences have shaped the person that I am, therefore, it's not my fault’* (including one subtheme - Disempowerment: *‘[My] core beliefs have been damaged’*), (3) Keep it simple: *‘Don’t push it too far over the top in case it becomes like spaghetti’*, and (4) A collaborative process?: *‘We both did it together (...) she did all, and I just watched’*.

In relation to themes one and two, there appeared to be key differences in the ways in which the *content* of maintenance, and longitudinal formulations were experienced. The impact of CBTp maintenance formulations (i.e. ‘vicious cycles’) appeared to evoke a sense of self-blame, but also, self-empowerment (and an *internal* LoC; Rotter, 1966), which *may* help to facilitate change. Conversely, the impact of longitudinal formulations appeared to evoke a sense of non-blame, but also, disempowerment (and

an external LoC), which *may* inhibit change. In relation to themes three and four, there appeared to be key similarities in the ways in which the *process* of maintenance, and longitudinal formulations were experienced. The impact of simple, parsimonious, and collaborative CFs may be experienced as self-empowering, which *may* help to facilitate change. Whereas overly complex, too inclusive, and less-collaborative CFs may be experienced as disempowering, which *may* inhibit change (Spencer et al, 2023).

6.2 Overarching themes and reflections

Four overarching themes permeate this thesis and tie the main body of this work together as a whole. These themes are discussed below.

6.2.1 *The CF versus diagnosis debate*

The first theme relating to the CF versus diagnosis debate, spans across the systematic review, and two empirical studies in this thesis. Firstly, the systematic review in Chapter 2, showed mixed evidence for the reliability of CFs which, as highlighted in Chapter 1, is also a challenge to the diagnostic framework (Frances & Nardo, 2013; Insel, 2013; Johnstone, 2014a, 2018; Johnstone et al., 2018; Jones, 2012; Kinderman & Allsopp, 2018; Kraemer et al., 2012; McCarthy, 2013). What is evident is that both psychological formulations and psychiatric diagnoses can suffer from relatively low rates of agreement between clinicians. This indicates that different clinicians may come to different understandings of what a person is experiencing. Secondly, as highlighted in Chapter 1, the presence of a mental health diagnosis does not tell us anything about prognosis (i.e. the predicted course of illness and the probability of future outcomes; Croft et al., 2015). Owing to this, mental health diagnoses are criticised for lacking validity (Frances & Nardo, 2013; Insel, 2013; Johnstone, 2014a, 2018; Johnstone et al., 2018; Jones, 2012; Kinderman & Allsopp, 2018; Kraemer et al., 2012; McCarthy, 2013). Similarly, (whilst acknowledging its limitations), the ACTION trial secondary analysis/causal inference study in Chapter 4, did not provide evidence to suggest that the presence of a CF helped to improve treatment outcomes (Spencer et al., 2018). More broadly, there is not strong evidence that formulation is a better predictor of outcome over diagnosis. Thirdly, whilst the reflexive TA study in Chapter 5 indicated that

formulation has the capacity to be of value to SUs, potential adverse reactions to the process were also described (Spencer et al., 2023). The argument that a mental health diagnosis may convey messages of hopelessness, blame, shame, and loss of self-agency (Brooke, 2004; Colizzi et al., 2020; Bentall, 2009; Boyle, 2002; Johnstone, 2014, 2021; Kirk & Kutchins, 1997; Matsunaga & Kitamura 2016; Mittal et al., 2014; Seery et al., 2021) is also reflected in (some) SU reactions to the CF process (Chadwick et al., 2003; Gibbs et al., 2020; Pain et al., 2008; Spencer et al., 2023). Hence the notion suggested by the PTMF that formulation is ‘good’, and diagnosis is ‘bad’ (Boyle & Johnstone, 2020; Johnstone, 2021; Johnstone et al., 2018) does not marry up with SU experiences of CF, as reported in Chapter 5.

As discussed in Chapter 1, there are challenges to diagnosis on the basis that it has been criticised for lacking reliability, agreement, prognostic value, and acceptability to SUs. It must be noted that the findings from this thesis did not demonstrate that formulation was substantially better than diagnosis in relation to any of these markers. Therefore, we cannot simply substitute diagnosis with formulation, and be confident that it is an improved framework for understanding mental health difficulties.

6.2.2 *Use of the four/five Ps frameworks to improve reliability and achieve consensus*

The second theme relating to use of the four/five Ps framework, spans across the systematic literature review, and one empirical study in this thesis. Firstly, the systematic review in Chapter 2, suggested that formulations constructed using an atheoretical modality (such as the four Ps) produced higher rates of reliability. Secondly, the Delphi study in Chapter 3, demonstrated that CBTp experts can reach CF consensus in relation to key factors organised within a five Ps framework. As discussed in Chapter 2, one hypothesis for higher rates of agreement in relation to the four Ps, may be owing to the framework being generic, and simpler, with wider scope to ‘capture’ many components. An alternative hypothesis is that the four/five Ps framework imposes structure and is prescriptive. Hence this may account for higher levels of reliability. Indeed, variability reduces reliability, and (as outline in Chapter 1), there are many different models of psychosis from which therapists can formulate. As Thrower et al. recently explained: “*there is a lack of consensus regarding the key*

components and processes of formulation that are hypothesized to contribute to the reliability of formulations” (Thrower et al., 2024, p. 1). Consequently, the ‘essential’ Delphi components identified in Chapter 3, now provide therapists with targeted/specific components, which may help them to achieve reliable CFs.

Perhaps though, the four/five Ps framework provides little room for ‘artistic flair’ and ‘creativity’ – as seen in the ‘aha!’ eureka moments described by SU’s in Chapter 5. CF should have both a scientific base and an artistic veneer, hence the Leonardo da Vinci quote, outlined at the start of this chapter.

6.2.3 Parsimony and simplicity

The third theme relating to parsimony and simplicity, spans across two empirical studies in this thesis. Firstly, the Delphi study in Chapter 3, showed that from an original list of 178 items ($n = 89$ items for auditory hallucinations, and $n = 89$ items for persecutory delusions), experts believe only 8 factors are ‘essential’ for the formulation of auditory hallucinations, and persecutory delusions in CBTp. Secondly, the reflexive TA study in Chapter 5 suggests that SUs also prefer to ‘keep it simple’, with complex, elaborate, and ‘spaghetti’ CF diagrams experienced by SUs as being too difficult to understand and/or remember (Spencer et al., 2023). Whilst Chapter 3 relates to parsimony regarding CF *content*, the thoughts of SUs captured in Chapter 5 also appear to reflect this view regarding CF *process*. Nevertheless, Eells (2022) suggests there is a balance to be made, between complex versus simple CFs:

if an overly simple construction is offered, important dimensions of the person’s problems may go unrecognised or misunderstood. If overly complex, the formulation may be unwieldy, too time-consuming, and impractical (...) Thus, a balance between complexity and simplicity is an important aim in case formulation construction. Parsimony is an important guiding principle (p. 23).

Hence, the idea of parsimony and simplicity in CBT CFs is not original (see Butler, 1998; Dudley et al., 2015; Kuyken et al., 2009; Kuyken et al., 2016; McMurran & Bruford, 2016; Morrison, 2007; and Padesky et al., 2011). Nevertheless, the Delphi and reflexive TA studies in this thesis, provide empirical *evidence* in support of this theory, and specifically, in relation to psychosis.

6.2.4 *Different levels of CF: A word of caution surrounding the use of longitudinal formulations versus the use of maintenance formulations*

The fourth (and final) theme relating to different levels of CF, spans across all three empirical studies in this thesis. Firstly, the Delphi study in Chapter 3, showed that of the 8 CF factors endorsed, experts reached consensus that 7 *perpetuating* factors are ‘essential’ for the formulation of auditory hallucinations, and persecutory delusions in CBTp. In contrast, no predisposing or precipitating factors were endorsed as essential (Spencer et al., 2020). This suggests that CBTp experts believe there should be a specific focus on maintenance cycles when formulating key psychotic symptoms. Secondly, the causal inference study in Chapter 4, offers a tentative signal that longitudinal formulations constructed in the first four sessions of therapy, *may* lead to poorer treatment effects. Thirdly, the reflexive TA study in Chapter 5, suggests that SUs experience maintenance formulations as a ‘vehicle for change’. Whereas, longitudinal formulations may be experienced as reducing one’s sense of self-agency, potentially resulting in disempowerment. Furthermore, the ‘quick’ process of the maintenance formulation contrasted with the way SUs talked about the longitudinal formulation, which was experienced as time-consuming owing to the length of time taken to construct (and emotionally process) the content of the timeline. Indeed, as Bighelli et al. (2018) has explained, SUs may find CBTp demanding or challenging, and they may have high expectations about therapy and feel discouraged if they do not see progress/results within a few sessions. That is why a ‘quick’ success experience via a maintenance formulation may so important - to keep SUs engaged in therapy.

The reflexive TA study was also predominantly undertaken with SUs experiencing FEP, and figures suggest that up to 82% of young people with FEP report some type of trauma (Trauelsen et al., 2015), and co-morbid PTSD (Bendall et al., 2012). Therefore (as previously referred to in Chapter 5), young people with FEP that engage in a written timeline or formulation of their major life events/traumatic experiences, may experience temporary distress and/or exacerbation of their psychotic symptoms (Halpin et al., 2016; Tong et al., 2017).

Tong et al. (2017) reported that, in psychosis, the timeline appeared to act as exposure (albeit a smaller dose of exposure to traumatic memories than traditional exposure-based treatments). A recommendation, therefore, is for SUs to be clearly informed of the potential side-effects of longitudinal formulation. As Tong et al.

explains: *“It may be helpful to provide anecdotal examples of possible reactions to ensure that clients have a clear understanding of what the potential risks are (...) and clinicians should be vigilant for the possibility of worsening [symptoms]”* (Tong et al., 2017, p. 572-573). As also mentioned in Chapter 5, a planned ‘preparatory phase’ could be used to develop rapport and scaffold/prepare SUs for this eventuality. Halpin et al. (2016) designed a coping module to help SUs build emotional regulation skills prior to a timeline and longitudinal formulation. Whilst this seems sensible, findings revealed that it did not help to reduce feelings of distress, over-whelm, and in-session dissociation. However, the dose, quality, and/or impact of the module may not have been sufficient. As such, there are currently no evidence-based models with which to scaffold a less emotionally intense timeline and/or longitudinal CF.

What preparatory work might therefore be done to facilitate a less-distressing longitudinal formulation? It is recommended here, that a) SUs use grounding techniques, self-care & comfort to counter-act distress after construction of a timeline and/or longitudinal formulation, b) therapists draw on SU strengths by looking for any adaptive coping strategies that SUs may be currently using, and c) therapists work on the maintenance cycle where possible - listening out for longitudinal factors and keeping those in mind, but aiming to prioritise and address presenting issues in the ‘here and now’.

Finally, it is true that some SUs may wish to start with a timeline – to tell the therapist their story and ‘let it all out’. Perhaps in this instance, it may be appropriate to work on a longitudinal formulation first, because it is on *their* terms. Once the SU and therapist have a good understanding as to what may have led to the problem(s) developing, the therapist can then make the SU aware that treatment is likely to proceed better if they switch focus, to try to work on what might be keeping the problem(s) going.

6.3 Thesis strengths, limitations, and concluding remarks

Strengths, limitations, and recommendations for future research have been outlined in the systematic review and in each of the empirical chapters, therefore, they will not be repeated here. However, the main strength of this thesis is the use of a number of

methods to explore several important questions concerning: 1) CBTp expert's views on the core components of a formulation, 2) use of an archival data set to investigate formulation in relation to outcome, and 3) the impact of formulation from a SUs' perspective. Each of these questions were carefully matched to the most appropriate research design and method. As such, one of the main strengths of this thesis was that H.M.S. produced discrete studies using multiple methodologies, which all converged on somewhat similar findings that emphasised the importance of maintenance formulations. Nevertheless, as a coherent body of research, the switch between methodologies meant that H.M.S. was not replicating, building upon, nor extending an individual study towards the next logical step. An example of this, as mentioned in Chapter 3, may have been to build on the findings from the Delphi study by testing out the 'essential' formulation components, to establish whether they translated to improvements in outcomes for SUs. Consequently, one of the limitations of this thesis is that as individual stand-alone studies, they are not necessarily of sufficient size or scale for us to be confident about any of the single findings. However, this was an explicit choice for H.M.S. to develop skills in several methods, and to match the correct method to the relevant research question.

Another key limitation of the thesis relates to the systematic literature review, Delphi, and the ACTION secondary analysis/causal inference studies. These chapters focused on the *content* of CFs, without consideration of CF as a *process*. Whilst the reflexive TA study did explore SUs' experiences of both content *and* process, the systematic review focused solely on reliability/rate of agreement (e.g. whether therapists can construct CFs that contain similar content, when asked to formulate the same case). Similarly, the Delphi study focused on determining the key ingredients of a CF, and the ACTION secondary analysis looked at the frequency or number of times CFs (longitudinal, maintenance, or both) were present within a course of therapy, in order to determine their treatment effects.

The process of CFs and *how* they are developed primarily focuses on the use of collaboration (Butler, 1998; Kuyken et al., 2009). As outlined in Chapter 1, formulation is about trying to see the world through the SU's eyes, and so CF is considered optimal when it is a process you do *with* someone, rather than *to* them (Kuyken et al., 2009). This contrasts with the use of indirect formulation (e.g. when a multi-disciplinary team formulates a person, without their direct involvement) which does not utilise the checks and balances that are present when SUs are actively involved.

In Chapter 5, SUs talked about the perceived value of completing certain sections of their CF for homework, and/or drawing out the CF for themselves. Hence, for SUs, the ‘how to?’ is central to this process (Spencer et al., 2023). The subtle nuances of understanding how to construct or develop CFs are not captured in the current systematic literature review, Delphi, or ACTION secondary analysis/causal inference studies. Consequently, the neglect of process factors is a limitation to our understanding of formulation in CBTp.

Whilst research on CF content has value, future studies may wish to pay more attention to *how* formulations are developed. For example, the Delphi method could be used to gain consensus with CBTp experts regarding the ‘essential’ processes involved in developing a CF for auditory hallucinations, or persecutory delusions. A causal inference study might also look to investigate CF processes by coding videos/audio recordings of therapy sessions to investigate treatment effects. For example, what was the timing of the CF? How was it constructed (if at all) on paper? Was it developed with the SU in a collaborative way? And do any of these processes lead to improved outcomes? This research would be of great value, but it was not possible to investigate this owing to their being no archive of video/audio recordings of therapy from the ACTION trial. Also, not all SUs consent to video or audio recordings. Therefore, even if this had been possible, there would likely have been large amounts of missing data. Moreover, the practicalities of listening to, transcribing, and coding a mean of 13.3 therapy sessions for each participant, would have been a challenge within the time constraints of this thesis.

A final limitation worth noting, relates to the ontological and epistemological thread running through this thesis. Psychiatric diagnosis, and cognitive models are built on empirical evidence, and support the notion of a single ‘truth’ to be found when it comes to making sense of people's experiences. Owing to this, we can look for things like reliability and validity when it comes to researching CF. However, the reflexive TA method used in Chapter 5, assumes there is no ‘truth’ residing in the data, owing to researcher subjectivity and interpretation (Clarke, 2021e). Hence, themes are reported as having been ‘generated’, or ‘developed’, to acknowledge the active role of the researcher in co-creating new knowledge (Braun et al., 2022; Clarke, 2021a). Consequently, there are ontological and epistemological tensions within this thesis that are acknowledged here.

H.M.S. was keen to embrace Big Q qualitative research in its purest form (Kidder & Fine, 1987). However, it may have made more sense (from an ontological and epistemological perspective) to have employed a ‘small q’ qualitative (Kidder & Fine, 1987) method which (as mentioned in Chapter 5), incorporates aspects of the positivist paradigm whereby accuracy, reliability, and replicability of coding are seen as essential (Braun & Clarke, 2021c; Kidder & Fine, 1987). Moreover, in small q qualitative research, use of multiple independent coders, and consensus between coders (via the use of inter-rater reliability checks) are seen as measures of quality, to control for researcher subjectivity and bias (Braun & Clarke, 2021c, 2021d, 2023; Clarke, 2021b).

In conclusion, psychological CF is an interesting area to research. CF has held an esteemed position in the practice of CBT since the early 1950s (DCP, 2011), despite its empirical basis being limited and somewhat mixed. The work presented in this thesis, broadly supports the notion that formulation is a way of seeing the world through the person’s eyes, and, if done well, can help make sense of a confusing range of psychotic experiences, to help improve people’s lives. The guidance from the Delphi study provides a practical, simple, and easy-to-use formulation model that focuses on maintenance processes, and this has the potential to be widely used. The findings from the ACTION secondary analysis and reflexive TA studies, provide clinicians with a strong steer in how to develop formulations with SU’s, to help direct treatment for psychosis. Owing to work carried out in this thesis, formulation in CBTp is better placed (and with a clearer way forwards), to improve research and clinical practice.

7. References

- Aarseth, E., Bean, A. M., Boonen, H., Colder Carras, M., Coulson, M., Das, D., Deleuze, J., Dunkels, E., Edman, J., Ferguson, C. J., Haagsma, M. C., Helmersson Bergmark, K., Hussain, Z., Jansz, J., Kardefelt-Winther, D., Kutner, L., Markey, P., Nielsen, R. K. L., Prause, N., ... Van Rooij, A. J. (2017). Scholars' open debate paper on the World Health Organization ICD-11 Gaming Disorder proposal. *Journal of Behavioral Addictions*, 6(3), 267-270. <https://doi.org/10.1556/2006.5.2016.088>
- Abel, A., Hayes, A. M., Henley, W., & Kuyken, W. (2016). Sudden gains in cognitive-behavior therapy for treatment-resistant depression: Processes of change. *Journal of Consulting and Clinical Psychology*, 84(8), 726-737. <https://doi.org/10.1037/ccp0000101>
- Aftab, A. (2020, August 14). *Moving beyond psychiatric diagnosis: Lucy Johnstone, PsyD: Conversations in critical psychiatry*. Psychiatric Times. <https://www.psychiatrictimes.com/view/moving-beyond-psychiatric-diagnosis-lucy-johnstone-psyd>
- Akins, R. B., Tolson, H., & Cole, B. R. (2005). Stability of response characteristics of a Delphi panel: Application of bootstrap data expansion. *BMC Medical Research Methodology*, 5, 37. <https://doi.org/10.1186/1471-2288-5-37>
- Albani, C., Pokorny, D., Blaser, G., Grüniger, S., König, S., Marschke, F., Geissler, I., Koerner, A., Geyer, M., & Kächele, H. (2002). Reformulation of the core conflictual relationship theme (CCRT) categories: The CCRT-LU category system. *Psychotherapy Research*, 12(3), 319-338. <https://doi.org/10.1093/ptr/12.3.319>
- Alvarez-Jiménez, M., Hetrick, S. E., González-Blanch, C., Gleeson, J. F., & McGorry, P. D. (2008). Non-pharmacological management of antipsychotic-induced weight gain: Systematic review and meta-analysis of randomised controlled trials. *The British Journal of Psychiatry*, 193(2), 101-107. <https://doi.org/10.1192/bjp.bp.107.042853>
- American Psychiatric Association (1952). *Diagnostic and statistical manual: Mental disorders* (1st ed.). American Psychiatric Association.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Association. <https://doi.org/10.1176/appi.books.9780890425596>
- American Psychological Association (2014, December 18). *Two (or more) heads are better than one for reasoning and perceptual decision-making*. American Psychological Association. <https://www.apa.org/pubs/highlights/peeps/issue->

[36#:~:text=The%20expression%20%22two%20heads%20are,argument%20is%20the%20most%20sound](#)

- American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders* (5th ed., text rev.). American Psychiatric Association. <https://doi.org/10.1176/appi.books.9780890425787>
- Andreasen, N. (1985). *The broken brain: The biological revolution in psychiatry*. HarperCollins INC International Concepts.
- Ashworth, M., Schofield, P., & Das-Munshi, J. (2017). Physical health in severe mental illness. *British Journal of General Practice*, 67(663), 436-437. <https://doi.org/10.3399/bjgp17X692621>
- Aston, R. (2009). A literature review exploring the efficacy of case formulations in clinical practice. What are the themes and pertinent issues? *the Cognitive Behaviour Therapist*, 2(2), 63-74. <https://doi.org/10.1017/S1754470X09000178>
- Aveline, M. (1999). The advantages of formulation over categorical diagnosis in explorative psychotherapy and psychodynamic management. *European Journal of Psychotherapy & Counselling*, 2(2), 199-216. <https://doi.org/10.1080/13642539908402327>
- Bakker, G. M. (2008). Problem-maintaining circles: Case illustrations of formulations that truly guide therapy. *Clinical Psychologist*, 12(1), 30-39. <https://doi.org/10.1080/13284200802069050>
- Baltagi, B. H. (2011). *Econometrics*. Springer.
- Barber, J. P., Luborsky, L., Diguer, L., & Crits-Christoph, P. (1995). A comparison of core conflictual relationship themes before psychotherapy and during early sessions. *Journal of Consulting and Clinical Psychology*, 63(1), 145-148. <https://doi.org/10.1037//0022-006x.63.1.145>
- Bartko, J. J. (1991). Measurement and reliability: Statistical thinking considerations. *Schizophrenia Bulletin*, 17(3), 483-489. <https://doi.org/10.1093/schbul/17.3.483>
- Bazeley, P., & Jackson, K. (2013). *Qualitative data analysis with NVivo* (2nd ed.). SAGE Publications Ltd.
- Beck, A. T. (1963). Thinking and depression: I. Idiosyncratic content and cognitive distortions. *Archives of General Psychiatry*, 9(4), 324-333. <https://doi.org/10.1001/archpsyc.1963.01720160014002>
- Beck, A. T. (1964). Thinking and depression: II. Theory and therapy. *Archives of General Psychiatry*, 10(6), 561-571. <https://doi.org/10.1001/archpsyc.1964.01720240015003>

- Beck, A. T. (1976). *Cognitive therapy and the emotional disorders*. International Universities Press.
- Beck, A. T. (1996). Beyond belief: A theory of modes, personality and psychopathology. In P. M. Salkovskis (Eds.), *Frontiers of cognitive therapy* (pp. 1-25). Guilford Press.
- Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy of depression*. Guilford Press.
- Beck, J. S. (1995). *Cognitive therapy: Basics and beyond*. Guilford Press.
- Beck, J. S. (2020). *Cognitive behavior therapy: Basics and beyond* (3rd ed.). Guilford Press.
- Ben-Aron, M., & McCormick, W. O. (1980). The teaching of formulation: Facts and deficiencies. *Canadian Journal of Psychiatry*, 25(2), 163-166.
<https://doi.org/10.1177/070674378002500211>
- Bendall, S., Alvarez-Jimenez, M., Hulbert, C. A., McGorry, P. D., & Jackson, H. J. (2012). Childhood trauma increases the risk of post-traumatic stress disorder in response to first-episode psychosis. *The Australian and New Zealand Journal of Psychiatry*, 46(1), 35-39. <https://doi.org/10.1177/0004867411430877>
- Benjamin, L. S. (1974). Structural analysis of social behavior. *Psychological Review*, 81(5), 392-425. <https://doi.org/10.1037/h0037024>
- Benjamin, L. S. (1996). An interpersonal theory of personality disorders. In J. Clarkin & M. Lenzenweger (Eds.), *Major theories of personality disorder* (pp. 141-220). Guilford.
- Benkert, F. (1997). *Reliability testing for plananalyses* [Unpublished Master's thesis]. The Eberhard Karl University of Tübingen.
- Bennett-Levy, J., Butler, G., Fennell, M., Hackmann, A., Mueller, M., & Westbrook, D. (2004). *Oxford guide to behavioural experiments in cognitive therapy*. Oxford University Press.
- Bentall, R. P. (2009). *Doctoring the mind: Why psychiatric treatments fail*. Allen Lane.
- Bentall, R. P., & Kaney, S. (1996). Abnormalities of self-representation and persecutory delusions: A test of a cognitive model of paranoia. *Psychological Medicine*, 26(6), 1231-1237. <https://doi.org/10.1017/s0033291700035959>
- Bentall, R. P., Kinderman, P., & Kaney, S. (1994). The self, attributional processes and abnormal beliefs: Towards a model of persecutory delusions. *Behaviour Research and Therapy*, 32(3), 331-341. [https://doi.org/10.1016/0005-7967\(94\)90131-7](https://doi.org/10.1016/0005-7967(94)90131-7)
- Bentall, R. P., Wickham, S., Shevlin, M., & Varese, F. (2012). Do specific early-life adversities lead to specific symptoms of psychosis? A study from the 2007 the Adult Psychiatric

- Morbidity Survey. *Schizophrenia Bulletin*, 38(4), 734-740.
<https://doi.org/10.1093/schbul/sbs049>
- Bhaskar, R. (2020). Critical realism and the ontology of persons. *Journal of Critical Realism*, 19(2), 113-120. <https://doi.org/10.1080/14767430.2020.1734736>
- Bieling, P. J., & Kuyken, W. (2003). Is cognitive case formulation science or science fiction? *Clinical Psychology: Science and Practice*, 10(1), 52-69. <https://doi.org/10.1093/clipsy.10.1.52>
- Bighelli, I., Salanti, G., Huhn, M., Schneider-Thoma, J., Krause, M., Reitmeir, C., Wallis, S., Schwermann, F., Pitschel-Walz, G., Barbui, C., Furukawa, T. A., & Leucht, S. (2018). Psychological interventions to reduce positive symptoms in schizophrenia: Systematic review and network meta-analysis. *World Psychiatry*, 17(3), 316-329.
<https://doi.org/10.1002/wps.20577>
- Birks, M., & Mills, J. (2015). *Grounded theory: A practical guide* (2nd ed.). SAGE Publications Ltd.
- Black, K. J. (2005). Psychiatry and the medical model. In E. H. Rubin & C.F., Zorumski (Eds.), *Adult Psychiatry* (2nd ed., pp. 3-17). Blackwell.
- Blackburn, I.-M., James, I. A., Milne, D. L., Baker, C., Standart, S., Garland, A., & Reichelt, F. K. (2001). The revised cognitive therapy scale (CTS-R): Psychometric properties. *Behavioural and Cognitive Psychotherapy*, 29(4), 431-446. <https://doi.org/10.1017/S1352465801004040>
- Bleuler, E. (1950). *Dementia praecox or the group of schizophrenias*. (J. Zinkin, Trans.). International University Press. (Original work published 1911).
- Bobes, J., Arango, C., Garcia-Garcia, M., Rejas, J., & CLAMORS Study Collaborative Group. (2010). Prevalence of negative symptoms in outpatients with schizophrenia spectrum disorders treated with antipsychotics in routine clinical practice: Findings from the CLAMORS study. *The Journal of Clinical Psychiatry*, 71(3), 280-286.
<https://doi.org/10.4088/JCP.08m04250yel>
- Böge, K., Thomas, N., Jacobsen, P. (2021). Is mindfulness for psychosis harmful? Deconstructing a myth. *British Journal of Psychiatry*, 218(2), 71-72.
<https://doi.org/10.1192/bjp.2020.165>
- Bolton, J. (2014). Case formulation after Engel – The 4P model. *Philosophy, Psychiatry and Psychology*, 21(3), 179-189. <https://dx.doi.org/10.1353/ppp.2014.0027>
- Bottesi, G., Noventa, S., Freeston, M. H., & Ghisi, M. (2019). Seeking certainty about intolerance of uncertainty: Addressing old and new issues through the intolerance of

- uncertainty scale-revised. *PLoS One*, 14(2), e0211929.
<https://doi.org/10.1371/journal.pone.0211929>
- Boyatzis, R. E. (1998). *Transforming qualitative information: Thematic analysis and code development*. SAGE Publications Ltd.
- Boyle, M. (2002). *Schizophrenia: A scientific delusion?* (2nd ed.). Brunner-Routledge.
- Boyle, M., & Johnstone, L. (2014). Alternatives to psychiatric diagnosis. *Lancet Psychiatry*, 1(6), 409-411. [https://doi.org/10.1016/S2215-0366\(14\)70359-1](https://doi.org/10.1016/S2215-0366(14)70359-1)
- Boyle, M., & Johnstone, L. (2020). *A straight talking introduction to the Power Threat Meaning Framework: An alternative to psychiatric diagnosis*. BCCS Books.
- Brabban, A., & Turkington, D. (2002). The search for meaning: Detecting congruence between life events, underlying schema and psychotic symptoms. Formulation-driven and schema-focused cognitive behavioural therapy for a neuroleptic-resistant schizophrenic patient with a delusional memory. In A. P. Morrison (Eds.), *A casebook of cognitive therapy for psychosis* (pp.59-74). Routledge.
- Brady, S. R. (2016). The Delphi method. In L. A. Jason & D. S. Glenwick (Eds.), *Handbook of methodological approaches to community-based research: Qualitative, quantitative, and mixed methods* (pp. 61-68). Oxford University Press.
- Braham, L. G., Trower, P., & Birchwood, M. (2004). Acting on command hallucinations and dangerous behaviour: A critique of the major findings in the last decade. *Clinical Psychology Review*, 24(5), 513-528. <https://doi.org/10.1016/j.cpr.2004.04.002>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101. <https://doi.org/10.1191/1478088706qp063oa>
- Braun, V., & Clarke, V. (2013). *Successful qualitative research: A practical guide for beginners*. SAGE Publications Ltd.
- Braun, V., & Clarke, V. (2016). (Mis) conceptualising themes, thematic analysis, and other problems with Fugard and Potts' (2015) sample-size tool for thematic analysis. *International journal of social research methodology*, 19(6), 739-743.
<https://doi.org/10.1080/13645579.2016.1195588>
- Braun, V., & Clarke, V. (2019a). Novel insights into patients' life-worlds: The value of qualitative research. *Lancet Psychiatry*, 6(9), 720-721. [https://doi.org/10.1016/S2215-0366\(19\)30296-2](https://doi.org/10.1016/S2215-0366(19)30296-2)
- Braun, V., & Clarke, V. (2019b). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, 11(4), 589-597.
<https://doi.org/10.1080/2159676X.2019.1628806>

- Braun, V., & Clarke, V. (2020a). *Answers to frequently asked questions about thematic analysis*. <https://cdn.auckland.ac.nz/assets/psych/about/our-research/documents/Answers%20to%20frequently%20asked%20questions%20about%20thematic%20analysis%20April%202019.pdf>
- Braun, V., & Clarke, V. (2020b). Can I use TA? Should I use TA? Should I *not* use TA? Comparing reflexive thematic analysis and other pattern-based qualitative analytic approaches. *Counselling and Psychotherapy Research*, 21(1), 37-47. <https://doi.org/10.1002/capr.12360>
- Braun, V., & Clarke, V. (2021a). The ebbs and flows of qualitative research: Time, change and the slow wheel of interpretation. In B. C. Clift, J. Gore, S. Gustafsson, S. Bekker, I. C. Batlle & J. Hatchard (Eds.), *Temporality in qualitative inquiry: Theories, methods and practices* (pp. 22-38). Taylor & Francis Group.
- Braun, V., & Clarke, V. (2021b). To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales. *Qualitative Research in Sport, Exercise and Health*, 13(2), 201-216. <https://doi.org/10.1080/2159676X.2019.1704846>
- Braun, V., & Clarke, V. (2021c). *Thematic analysis: A practical guide*. SAGE Publications Ltd.
- Braun, V., & Clarke, V. (2021d). One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology*, 18(3), 328-352. <https://doi.org/10.1080/14780887.2020.1769238>
- Braun, V., & Clarke, V. (2022). Conceptual and design thinking for thematic analysis. *Qualitative Psychology*, 9(1), 3-26. <https://doi.org/10.1037/qup0000196>
- Braun, V., & Clarke, V. (2023). Toward good practice in thematic analysis: Avoiding common problems and be(com)ing a *knowing* researcher. *International Journal of Transgender Health*, 24(1), 1-6. <https://doi.org/10.1080/26895269.2022.2129597>
- Braun, V., Clarke, V., & Hayfield, N. (2022). ‘A starting point for your journey, not a map’: Nikki Hayfield in conversation with Virginia Braun and Victoria Clarke about thematic analysis. *Qualitative Research in Psychology*, 19(2), 424-445. <https://doi.org/10.1080/14780887.2019.1670765>
- British Psychological Society (BPS). (1969). *Regulations for the British Psychological Society diploma in clinical psychology*. British Psychological Society.
- Bromet, E. J., Kotov, R., Fochtmann, L. J., Carlson, G. A., Tanenberg-Karant, M., Ruggero, C., & Chang, S. W. (2011). Diagnostic shifts during the decade following first admission

- for psychosis. *The American Journal of Psychiatry*, 168(11), 1186-1194.
<https://doi.org/10.1176/appi.ajp.2011.11010048>
- Brooke, S. (2004). Formulation in clinical psychology: Past, present and future. *Clinical Psychology*, 41(September), 36-39.
- Brown, P., Waite, F., & Freeman, D. (2019). 'Twisting the lion's tail': Manipulationist tests of causation for psychological mechanisms in the occurrence of delusions and hallucinations. *Clinical Psychology Review*, 68, 25-37.
<https://doi.org/10.1016/j.cpr.2018.12.003>
- Bruch, M. (2015). The development of case formulation approaches. In M. Bruch (Ed.). *Beyond diagnosis: Case formulation in cognitive behavioural therapy* (2nd ed., pp. 1-23). Wiley Blackwell.
- Bucci, S., & Tarrier, N. (2016). A cognitive-behavioural case formulation approach to the treatment of psychosis. In N. Tarrier & J. Johnson (Eds.), *Case formulation in cognitive behaviour therapy: The treatment of challenging and complex cases* (2nd ed., pp. 166-187). Routledge.
- Burgoon, J. K., Guerrero, L. K., & Floyd, K. (2009). *Nonverbal communication*. Routledge.
- Butler, G. (1998). Clinical formulation. In A. S. Bellack & M. Hersen (Eds.), *Comprehensive clinical psychology* (pp. 1-24). Pergamon.
- Butler, G. (2006). The value of formulation: A question for debate. *Clinical Psychology Forum*, 1(160), 9-12. <https://doi.org/10.53841/bpscpf.2006.1.160.9>
- Butler, A. C., Chapman, J. E., Forman, E. M., & Beck, A. T. (2006). The empirical status of cognitive-behavioural therapy: A review of meta-analyses. *Clinical Psychology Review*, 26(1), 17-31. <https://doi.org/10.1016/j.cpr.2005.07.003>
- Byng-Hall, J. (1980). Symptom bearer as marital distance regulator: Clinical implications. *Family Process*, 19(4), 355-365. <https://doi.org/10.1111/j.1545-5300.1980.00355.x>
- Byrne R., & Morrison A. (2014). Service users' priorities and preferences for treatment of psychosis: A user-led Delphi study. *Psychiatric Services*. 65(9), 1167-1169.
<https://doi.org/10.1176/appi.ps.201300289>
- Cairns, D. V. A., Reid, D. G. S., & Murray, D. C. D. (2015). Experience of psychosocial formulation within a biopsychosocial model of care for first- episode psychosis. *International Journal of Psychosocial Rehabilitation*, 19(2), 47-62.

- Carey, T. A., & Pilgrim, D. (2010). Diagnosis and formulation: What should we tell the students? *Clinical Psychology & Psychotherapy*, 17(6), 447-454.
<https://doi.org/10.1002/cpp.695>
- CASP. (2004). *Critical appraisal skills programme*. <http://www.casp-uk.net/>
- Caspar, F. (2019). Plan analysis and the motive-oriented therapeutic relationship. In U. Kramer (Eds.), *Case formulation for personality disorders* (pp. 265-290). Academic Press.
- Caston, J., & Martin, E. (1993). Can analysts agree? The problems of consensus and the psychoanalytic mannequin: II. Empirical tests. *Journal of the American Psychoanalytic Association*, 41(2), 513-548. <https://doi.org/10.1177/000306519304100209>
- Chadwick, P., & Birchwood, M. (1994). The omnipotence of voices. A cognitive approach to auditory hallucinations. *British Journal of Psychiatry*, 164(2), 190-201.
<https://doi.org/10.1192/bjp.164.2.190>
- Chadwick, P. D. J., Lowe, C. F., Horne, P. J., & Higson, P. J. (1994). Modifying delusions: The role of empirical testing. *Behavior Therapy*, 25(1), 35-49.
[https://doi.org/10.1016/S0005-7894\(05\)80144-3](https://doi.org/10.1016/S0005-7894(05)80144-3)
- Chadwick, P., Lees, S., & Birchwood, M. (2000). The revised beliefs about voices questionnaire (BAVQ-R). *British Journal of Psychiatry*, 177(3), 229-232.
<https://doi.org/10.1192/bjp.177.3.229>
- Chadwick, P., Williams, C., & Mackenzie, J. (2003). Impact of case formulation in cognitive behaviour therapy for psychosis. *Behaviour Research and Therapy*, 41(6), 671-680.
[https://doi.org/10.1016/s0005-7967\(02\)00033-5](https://doi.org/10.1016/s0005-7967(02)00033-5)
- Chamberlain, K. (2012). Do you really need a methodology? *Qualitative Methods in Psychology Bulletin*, 13, 59-63.
- Chambers, J. C., Eccleston, L., Day, A., Ward, T., & Howells, K. (2008). Treatment readiness in violent offenders: The influence of cognitive factors on engagement in violence programs. *Aggression and Violent Behavior*, 13(4), 276-284.
<https://doi.org/10.1016/j.avb.2008.04.003>
- Chmielewski, M., Clark, L. A., Bagby, R. M., & Watson, D. (2015). Method matters: Understanding diagnostic reliability in DSM-IV and DSM-5. *Journal of Abnormal Psychology*, 124(3), 764-769. <https://doi.org/10.1037/abn0000069>
- Christiansen, C. (2007). Adolf Meyer revisited: Connections between lifestyles, resilience and illness. *Journal of Occupational Science*, 14(2), 63-76.
<https://doi.org/10.1080/14427591.2007.9686586>

- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 284-290. <https://doi:10.1037/1040-3590.6.4.284>
- Cicchetti, D. V., & Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiency*, 86(2), 127-137.
- Cieri, F., & Esposito, R. (2019). Psychoanalysis and neuroscience: The bridge between mind and brain. *Frontiers in Psychology*, 10, 1790. <https://doi.org/10.3389/fpsyg.2019.01983>
- Clarke, V. (2021a, November 30). *Foundations of qualitative research 2. Part 1: Key characteristics of qualitative research* [Webinar]. YouTube. <https://www.youtube.com/watch?v=-XWrpqPrHPk&t=144s>
- Clarke, V. (2021b, November 30). *Foundations of qualitative research 1. Part 1: What is qualitative research?* [Webinar]. YouTube. <https://www.youtube.com/watch?v=qH0XvEe6DzI&t=1s>
- Clarke, V. (2021c, November 30). *Foundations of qualitative research 2. Part 3: Key (theoretical and philosophical) concepts in qualitative research* [Webinar]. YouTube. <https://www.youtube.com/watch?v=bgZuPmO9130&t=5s>
- Clarke, V. (2021d, November 30). *Thematic analysis. Part 3: Six phases of reflexive thematic analysis*. [Webinar]. YouTube. <https://www.youtube.com/watch?v=BhL113ye9Ss>
- Clarke, V. (2021e, November 30). *Foundations of qualitative research 1. Part 3: Orientations to meaning in Big Q qualitative* [Webinar]. YouTube. https://www.youtube.com/watch?v=fG6-I-c_Oc&t=206s
- Clarke, V. (2021f, November 30). *Foundations of qualitative research 2. Part 2: Experiential and critical qualitative research* [Webinar]. YouTube. <https://www.youtube.com/watch?v=OfwtzhE4HtQ&t=5s>
- Clarke, V. (2021g, November 30). *Thematic analysis. Part 4: Avoiding common problems* [Webinar]. YouTube. <https://www.youtube.com/watch?v=tHeLh1XrWS0>
- Clarke, V. (2021h, November 30). *Thematic analysis. Part 2: Thematic analysis is uniquely flexible* [Webinar]. YouTube. <https://www.youtube.com/watch?v=-VqcGKvcWm8&t=1s>
- Clarke, V. (2022, May 19). *Common problems in published thematic analysis research in health psychology* [Webinar]. YouTube. <https://www.youtube.com/watch?v=J3q4hmqo7BE>

- Clements, K., & Turpin, G. (1992). Vulnerability models and schizophrenia: The assessment and prediction of relapse. In M. Birchwood & N. Tarrier (Eds.), *Innovations in the psychological management of schizophrenia* (pp. 21-47). Wiley.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46. <https://doi.org/10.1177/001316446002000104>
- Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70(4), 213-220. <https://doi.org/10.1037/h0026256>
- Coid, J. W., Ullrich, S., Kallis, C., Keers, R., Barker, D., Cowden, F., & Stamps, R. (2013). The relationship between delusions and violence: Findings from the East London first episode psychosis study. *JAMA Psychiatry*, 70(5), 465-471. <https://doi.org/10.1001/jamapsychiatry.2013.12>
- Collins, W. D., & Messer, S. B. (1991). Extending the plan formulation method to an object relations perspective: Reliability, stability, and adaptability. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, 3(1), 75-81. <https://doi.org/10.1037/1040-3590.3.1.75>
- Colizzi, M., Ruggeri, M., & Lasalvia, A. (2020). Should we be concerned about stigma and discrimination in people at risk for psychosis? A systematic review. *Psychological Medicine*, 50(5), 705-726. <https://doi.org/10.1017/S0033291720000148>
- Conway, C. C., Krueger, R. F., & HiTOP Consortium Executive Board (2021). Rethinking the diagnosis of mental disorders: Data-driven psychological dimensions, not categories, as a framework for mental-health research, treatment, and training. *Current Directions in Psychological Science*, 30(2), 151-158. <https://doi.org/10.1177%2F0963721421990353>
- Cooper, J. E., Kendell, R. E., Gurland, B. J., Sharpe, L., Copeland, J. R. M., & Simon, R. (1972). *Psychiatric diagnosis in New York and London*. Oxford University Press.
- Correll, C. U., & De Hert, M. (2013). Antipsychotics for acute schizophrenia: Making choices. *Lancet*, 382(9896), 919-920. [https://doi.org/10.1016/S0140-6736\(13\)61032-6](https://doi.org/10.1016/S0140-6736(13)61032-6)
- Cortese, S., Solmi, M., Michelini, G., Bellato, A., Blanner, C., Canozzi, A., Eudave, L., Farhat, L. C., Højlund, M., Köhler-Forsberg, O., Leffa, D. T., Rohde, C., de Pablo, G. S., Vita, G., Wesselhoeft, R., Martin, J., Baumeister, S., Bozhilova, N. S., Carlisi, C. O., ... Correll, C. U. (2023). Candidate diagnostic biomarkers for neurodevelopmental disorders in children and adolescents: A systematic review. *World Psychiatry: Official Journal of the World Psychiatric Association (WPA)*, 22(1), 129-149. <https://doi.org/10.1002/wps.21037>

- Coxon, A. (2022). 'Thematic analysis has travelled to places that we've never heard of'. *The Psychologist*, 37(2), 38-43.
- Craddock, N., & Mynors-Wallis, L. (2014). Psychiatric diagnosis: impersonal, imperfect and important. *The British Journal of Psychiatry*, 204(2), 93-95.
<https://doi.org/10.1192/bjp.bp.113.133090>
- Critchfield, K. L., Benjamin, L. S., & Levenick, K. (2015). Reliability, sensitivity, and specificity of case formulations for comorbid profiles in interpersonal reconstructive therapy: Addressing mechanisms of psychopathology. *Journal of Personality Disorders*, 29(4), 547-573. <https://doi.org/10.1521/pedi.2015.29.4.547>
- Critical Psychiatry Network. (2013, May 23). *Blogs: Statement by the Critical Psychiatry Network on the new DSM-5*. <https://www.criticalpsychiatry.co.uk/news/statement-by-the-critical-psychiatry-network-on-the-new-dsm-5/>
- Crits-Christoph, P. (1992). The efficacy of brief dynamic psychotherapy: A meta-analysis. *The American Journal of Psychiatry*, 149(2), 151-158. <https://doi.org/10.1176/ajp.149.2.151>
- Crits-Christoph, P., Luborsky, L., Dahl, L., Popp, C., Mellon, J., & Mark, D. (1988). Clinicians can agree in assessing relationship patterns in psychotherapy. The core conflictual relationship theme method. *Archives of General Psychiatry*, 45(11), 1001-1004.
<https://doi.org/10.1001/archpsyc.1988.01800350035005>
- Croft, P., Altman, D. G., Deeks, J. J., Dunn, K. M., Hay, A. D., Hemingway, H., LeResche, L., Peat, G., Perel, P., Petersen, S. E., Riley, R. D., Roberts, I., Sharpe, M., Stevens, R. J., Van Der Windt, D. A., Von Korff, M., & Timmis, A. (2015). The science of clinical practice: disease diagnosis or patient prognosis? Evidence about "what is likely to happen" should shape clinical practice. *BMC Medicine*, 13(20), 1-8.
<https://doi.org/10.1186/s12916-014-0265-4>
- Cromarty, P., & Gallagher, D. (2023). What guides guided self-help? Recognising the role of formulation in low intensity CBT. *Mental Health Science*, 1(1), 48-54.
- Currell, S., Christodoulides, T., Siitarinen, J., & Dudley, R. (2016). Patient factors that impact upon cognitive behavioural therapy for psychosis: Therapists' perspectives. *Behavioural and Cognitive Psychotherapy*, 44(4), 493-498.
<https://doi.org/10.1017/S1352465815000260>
- Curtis, J. T., & Silberschatz, G. (1989). *The plan formulation method: A reliable procedure for case formulation*. [Unpublished manuscript].

- Curtis, J. T., Silberschatz, G., Weiss, J., Sampson, H., & Rosenberg, S. E. (1988). Developing reliable psychodynamic case formulations: An illustration of the plan diagnosis method. *Psychotherapy*, 25(2), 256-265. <https://doi.org/10.1037/h0085340>
- Daalman, K., Dierenen, K. M. J., Derks, E. M., van Lutterveld, R., Kahn, R. S., & Sommer, I. E. C. (2012). Childhood trauma and auditory verbal hallucinations. *Psychological Medicine*, 42(12), 2475-2484. <https://doi.org/10.1017/S0033291712000761>
- Dalkey, N. C. (1972). The Delphi method: An experimental study of group opinion. In N. C. Dalkey, D. L. Rourke, R. Lewis, & D. Snyder (Eds.), *Studies in the quality of life: Delphi and decision-making* (pp. 13-54). Lexington Books.
- Dalkey, N. C., & Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Management Science*, 9(3), 458-467. <https://doi.org/10.1287/mnsc.9.3.458>
- Dallos, R. (2006). Integrative formulation: CAT and NAT. In L. Johnstone & R. Dallos (Eds.), *Formulation in psychology and psychotherapy: Making sense of people's problems* (1st ed., pp. 182-207). Routledge.
- Dallos, R., & Stedmon, J. (2014). Systemic formulation: Mapping the family dance. In L. Johnstone & R. Dallos (Eds.), *Formulation in psychology and psychotherapy: Making sense of people's problems* (2nd ed., pp. 67-95). Routledge.
- Dancey C. P., & Reidy, J. (2020). *Statistics without maths for psychology* (8th ed.). Pearson.
- Danermark, B., Ekström, M., & Karlsson, J. C. (2019). Explaining society: Critical realism in the social sciences (2nd ed.). Routledge.
- de Paula-Ravagnani, G. S., Guanaes-Lorenzi, C., & Rasera, E. F. (2017). Use of theoretical models in family therapy: Focus on social constructionism. *Paidéia*, 27(67), 84-92. <https://doi.org/10.1590/1982-43272767201710>
- Derogatis, L. R. (1983). *The SCL-90-R: Administration, scoring, and procedures manual II*. Clinical Psychometric Research.
- DeVellis, R. F. (2005). Inter-Rater Reliability. In K. Kempf-Leonard (Ed.), *Encyclopedia of social measurement* (2nd ed., pp. 317-322). Elsevier.
- DeWitt, K. N., Kaltreider, N. B., Weiss, D. S., & Horowitz, M. J. (1983). Judging change in psychotherapy. Reliability of clinical formulations. *Archives of General Psychiatry*, 40(10), 1121-1128. <https://doi.org/10.1001/archpsyc.1983.01790090083013>
- Diamond, I. R., Grant, R. C., Feldman, B. M., Pencharz, P. B., Ling, S. C., Moore, A. M., & Wales, P. W. (2014). Defining consensus: A systematic review recommends

- methodologic criteria for reporting of Delphi studies. *Journal of Clinical Epidemiology*, 67(4), 401-409. <https://doi.org/10.1016/j.jclinepi.2013.12.002>
- Division of Clinical Psychology. (2010). *The core purpose and philosophy of the profession*. The British Psychological Society. https://www.google.co.uk/url?sa=i&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=0CAMQw7AJahcKEwjghZnm4Z_8AhUAAAAAHQAAAAAQAw&url=https%3A%2F%2Fhull-repository.worktribe.com%2FOutputFile%2F369342&psig=AOvVaw1nIdup5KbLDFjeZaLWQJ8G&ust=1672435447672546
- Division of Clinical Psychology. (2011). *Good practice guidelines on the use of psychological formulation*. The British Psychological Society. <https://shop.bps.org.uk/good-practice-guidelines-on-the-use-of-psychological-formulation>
- Division of Clinical Psychology. (2013). *Classification of behaviour and experience in relation to functional psychiatric diagnoses: Time for a paradigm shift. DCP position statement*. The British Psychological Society. <https://www.madinamerica.com/wp-content/uploads/2015/06/DCP-classification-of-behavior-and-experience.pdf>
- Dudley, R., Bryant, C., Hammond, K., Siddle, R., Kingdon, D., & Turkington, D. (2007). Techniques in cognitive behavioural therapy: Using normalising in schizophrenia. *Journal of the Norwegian Psychological Association*, 44(5), 562-572.
- Dudley, R., Ingham, B., Sowerby, K., & Freeston, M. (2015). The utility of case formulation in treatment decision making: The effect of experience and expertise. *Journal of Behavior Therapy and Experimental Psychiatry*, 48, 66-74. <https://doi.org/10.1016/j.jbtep.2015.01.009>
- Dudley, R., & Kuyken, W. (2014). Case formulation in cognitive behavioural therapy: A principle-driven approach. In L. Johnstone & R. Dallos (Eds.), *Formulation in psychology and psychotherapy: Making sense of people's problems* (2nd ed., pp. 18-44). Routledge.
- Dudley, R., Kuyken, W., & Padesky, C. A. (2011). Disorder specific and trans-diagnostic case conceptualisation. *Clinical Psychology Review*, 31(2), 213-224. <https://doi.org/10.1016/j.cpr.2010.07.005>
- Dudley, R., Park, I., James, I., & Dodgson, G. (2010). Rate of agreement between clinicians on the content of a cognitive formulation of delusional beliefs: The effect of qualifications and experience. *Behavioural and Cognitive Psychotherapy*, 38(2), 185-200. <https://doi.org/10.1017/S1352465809990658>

- Dudley, R., Shaftoe, D., Cavanagh, K., **Spencer, H.**, Ormrod, J., Turkington, D., & Freeston, M. (2011). 'Jumping to conclusions' in first-episode psychosis. *Early Intervention in Psychiatry*, 5(1), 50-56. <https://doi.org/10.1111/j.1751-7893.2010.00258.x>
- Dudley, R., Siitarinen, J., James, I., & Dodgson, G. (2009). What do people with psychosis think caused their psychosis? A Q methodology study. *Behavioural and Cognitive Psychotherapy*, 37(1), 11-24. <https://doi.org/10.1017/S1352465808004955>
- Dudley, R., Taylor, P., Wickham, S., & Hutton, P. (2016). Psychosis, delusions and the "Jumping to Conclusions" reasoning bias: A systematic review and meta-analysis. *Schizophrenia Bulletin*, 42(3), 652-665. <https://doi.org/10.1093/schbul/sbv150>
- Duncan, E. A. S., Nicol, M. M., & Ager, A. (2004). Factors that constitute a good cognitive behavioural treatment manual: A Delphi study. *Behavioural and Cognitive Psychotherapy*, 32(2), 199-213. <https://doi.org/10.1017/S135246580400116X>
- Dunn, G., & Bentall, R. (2007). Modelling treatment-effect heterogeneity in randomized controlled trials of complex interventions (psychological treatments). *Statistics in Medicine*, 26(26), 4719-4745. <https://doi.org/10.1002/sim.2891>
- Dunn, G., Emsley, R., Liu, H., Landau, S., Green, J., White, I., & Pickles, A. (2015). Evaluation and validation of social and psychological markers in randomised trials of complex interventions in mental health: A methodological research programme. *Health Technology Assessment*, 19(93), 1-115. <https://doi.org/10.3310/hta19930>
- Dunn, G., Fowler, D., Rollinson, R., Freeman, D., Kuipers, E., Smith, B., Steel, C., Onwumere, J., Jolley, S., Garety, P., & Bebbington, P. (2012). Effective elements of cognitive behaviour therapy for psychosis: Results of a novel type of subgroup analysis based on principal stratification. *Psychological Medicine*, 42(5), 1057-1068. <https://doi.org/10.1017/S0033291711001954>
- Dunn, H., Morrison, A. P., & Bentall, R. P. (2006). The relationship between patient suitability, therapeutic alliance, homework compliance and outcome in cognitive therapy for psychosis. *Clinical Psychology and Psychotherapy*, 13(3), 145-152. <https://doi.org/10.1002/cpp.481>
- Easden, M. H., & Fletcher, R. B. (2020). Therapist competence in case conceptualization and outcome in CBT for depression. *Psychotherapy Research*, 30(2), 151-169. <https://doi.org/10.1080/10503307.2018.1540895>
- Easden, M. H., & Kazantzis, N. (2018). Case conceptualization research in cognitive behavior therapy: A state of the science review. *Journal of Clinical Psychology*, 74(3), 356-384. <https://doi.org/10.1002/jclp.22516>

- Editorial (1988). Where next with psychiatric illness? *Nature*, 336(6195), 95-96.
<https://doi.org/10.1038/336095a0>
- Eells, T. D. (2022). The history and current status of psychotherapy case formulation. In T. D. Eells (Ed.), *Handbook of psychotherapy case formulation* (3rd ed., pp. 1-35). Guilford Press.
- Eells, T. D. (2009). Contemporary themes in case formulation. In P. Sturmey (Ed.), *Clinical case formulation: Varieties of approaches* (pp. 291-315). John Wiley & Sons.
<https://doi.org/10.1002/9780470747513>
- Eells, T. D., Horowitz, M. J., Singer, J., Salovey, P., Daigle, D., & Turvey, C. (1995). The relationship models method: A comparison of independently derived case formulations. *Psychotherapy Research*, 5(2), 154-168. <https://doi.org/10.1080/10503309512331331276>
- Eells, T. D., Kendjelic, E. M., & Lucas, C. P. (1998). What's in a case formulation? Development and use of a content coding manual. *The Journal of Psychotherapy Practice and Research*, 7(2), 144-153.
- Eells, T. D., & Lombart, K. G. (2003). Case formulation and treatment concepts among novice, experienced, and expert cognitive-behavioral and psychodynamic therapists. *Psychotherapy Research*, 13(2), 187-204. <https://doi.org/10.1093/ptr/kpg018>
- Eells, T., Lombart, K. G., Kendjelic, E. M., Turner, L., & Lucas, C. P. (2005). The quality of psychotherapy case formulations: A comparison of expert, experienced, and novice cognitive-behavioral and psychodynamic therapists. *Journal of Consulting and Clinical Psychology*, 73(4), 579-589. <https://doi.org/10.1037/0022-006X.73.4.579>
- Eells, T. D., Lombart, K. G., Salsman, N., Kendjelic, E. M., Schneiderman, C. T., & Lucas, C. P. (2011). Expert reasoning in psychotherapy case formulation. *Psychotherapy Research: Journal of the Society for Psychotherapy Research*, 21(4), 385-399.
<https://doi.org/10.1080/10503307.2010.539284>
- Elliott, R., Fischer, C. T., & Rennie, D. L. (1999). Evolving guidelines for publication of qualitative research studies in psychology and related fields. *British Journal of Clinical Psychology*, 38, 215-229. <https://doi.org/10.1348/014466599162782>
- Ellis, A. (1957). Rational psychotherapy and individual psychology. *Journal of Individual Psychology*, 13(1), 38-44.
- Ellis, A. (1962). *Reason and emotion in psychotherapy*. Lyle Stuart.
- Elsevier (2023, January 18). Elsevier Support Center. How do I search in Scopus? Tips for proximity operators. https://service.elsevier.com/app/answers/detail/a_id/34325/

- Emsley, R., Dunn, G., & White, I. R. (2010). Mediation and moderation of treatment effects in randomised controlled trials of complex interventions. *Statistical Methods in Medical Research*, 19(3), 237-270. <https://doi.org/10.1177/0962280209105014>
- Engel, G. L. (1977). The need for a new medical model: A challenge for biomedicine. *Science*, 196(4286), 129-136. <https://doi.org/10.1126/science.847460>
- Etelämäki, M., Voutilainen, L., & Weiste, E. (2021). Distributing agency and experience in therapeutic interaction: Person references in therapists' responses to complaints. *Frontiers in Psychology*, 12, 585321. <https://doi.org/10.3389/fpsyg.2021.585321>
- Ferrer, E., & Nesselroade, J. R. (2003). Modeling affective processes in dyadic relations via dynamic factor analysis. *Emotion*, 3(4), 344-360. <https://doi.org/10.1037/1528-3542.3.4.344>
- Flach, C., French, P., Dunn, G., Fowler, D., Gumley, A. I., Birchwood, M., Stewart, S. L. K., & Morrison, A. P. (2015). Components of therapy as mechanisms of change in cognitive therapy for people at risk of psychosis: Analysis of the EDIE-2 trial. *The British Journal of Psychiatry*, 207(2), 123-129. <https://doi.org/10.1192/bjp.bp.114.153320>
- Fleiss, J. L. (1981). *Statistical methods for rates and proportions* (2nd ed.). John Wiley.
- Flinn, L., Braham, L., & das Nair, R. (2015). How reliable are case formulations? A systematic literature review. *The British Journal of Clinical Psychology*, 54(3), 266-290. <https://doi.org/10.1111/bjc.12073>
- Flitcroft, A., James, I. A., Freeston, M., & Wood-Mitchell, A. (2007). Determining what is important in a good formulation. *Behavioural and Cognitive Psychotherapy*, 35(3), 325-333. <http://doi.org/10.1017/S135246580600350X>
- Fothergill, C., & Kuyken, W. (2002). The quality of cognitive therapy case formulation rating scale. [Unpublished manuscript].
- Fowler, D., Freeman, D., Smith, B., Kuipers, E., Bebbington, P., Bashforth, H., Coker, S., Hodgekins, J., Gracie, A., Dunn, G., & Garety, P. (2006). The Brief Core Schema Scales (BCSS): Psychometric properties and associations with paranoia and grandiosity in non-clinical and psychosis samples. *Psychological Medicine*, 36(6), 749-759. <https://doi.org/10.1017/S0033291706007355>
- Frances, A. (2013a). *Saving normal: An insider's revolt against out-of-control psychiatric diagnosis, DSM-5, big pharma, and the medicalization of ordinary life*. HarperCollins.

- Frances, A. (2013b, May 5). *One manual shouldn't dictate US mental health research*. NewScientist. <https://www.newscientist.com/article/dn23490-one-manual-shouldnt-dictate-us-mental-health-research/>
- Frances, A., & Nardo, J. (2013). ICD-11 should not repeat the mistakes made by DSM-5. *The British Journal of Psychiatry*, 203(1), 1-2. <https://doi.org/10.1192/bjp.bp.113.127647>
- Frances, A. J., & Widiger, T. (2012). Psychiatric diagnosis: Lessons from the DSM-IV past and cautions for the DSM-5 future. *Annual Review of Clinical Psychology*, 8, 109-130. <https://doi.org/10.1146/annurev-clinpsy-032511-143102>
- Frangakis, C. E., & Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics*, 58(1), 21-29. <https://doi.org/10.1111/j.0006-341X.2002.00021.x>
- Frankl, V. E. (1960). *Man's search for meaning*. Beacon Press.
- Freeman, D. (2016). Persecutory delusions: A cognitive perspective on understanding and treatment. *Lancet Psychiatry*, 3(7), 685-692. [https://doi.org/10.1016/S2215-0366\(16\)00066-3](https://doi.org/10.1016/S2215-0366(16)00066-3)
- Freeman, D., Bradley, J., Antley, A., Bourke, E., DeWeever, N., Evans, N., Černis, E., Sheaves, B., Waite, F., Dunn, G., Slater, M., & Clark, D. M. (2016). Virtual reality in the treatment of persecutory delusions: Randomised controlled experimental study testing how to reduce delusional conviction. *The British Journal of Psychiatry*, 209(1), 62-67. <https://doi.org/10.1192/bjp.bp.115.176438>
- Freeman, D., Bradley, J., Waite, F., Sheaves, B., DeWeever, N., Bourke, E., McInerney, J., Evans, N., Černis, E., Lister, R., Garety, P., & Dunn, G. (2016). Targeting recovery in persistent persecutory delusions: A proof of principle study of a new translational psychological treatment (the Feeling Safe Programme). *Behavioural and Cognitive Psychotherapy*, 44(5), 539-552. <https://doi.org/10.1017/S1352465816000060>
- Freeman, D., Dunn, G., Murray, R. M., Evans, N., Lister, R., Antley, A., Slater, M., Godlewska, B., Cornish, R., Williams, J., Di Simplicio, M., Igoumenou, A., Brenneisen, R., Tunbridge, E. M., Harrison, P. J., Harmer, C. J., Cowen, P., & Morrison, P. D. (2015). How cannabis causes paranoia: Using the intravenous administration of Δ^9 -tetrahydrocannabinol (THC) to identify key cognitive mechanisms leading to paranoia. *Schizophrenia Bulletin*, 41(2), 391-399. <https://doi.org/10.1093/schbul/sbu098>
- Freeman, D., Dunn, G., Startup, H., Pugh, K., Cordwell, J., Mander, H., Černis, E., Wingham, G., Shirvell, K., & Kingdon, D. (2015). Effects of cognitive behaviour therapy for worry on persecutory delusions in patients with psychosis (WIT): A parallel, single-

- blind, randomised controlled trial with a mediation analysis. *Lancet Psychiatry*, 2(4), 305-313. [https://doi.org/10.1016/S2215-0366\(15\)00039-5](https://doi.org/10.1016/S2215-0366(15)00039-5)
- Freeman, D., Emsley, R., Diamond, R., Collett, N., Bold, E., Chadwick, E., Isham, L., Bird, J. C., Edwards, D., Kingdon, D., Fitzpatrick, R., Kabir, T., Waite, F., & Oxford Cognitive Approaches to Psychosis Trial Study Group. (2021). Comparison of a theoretically driven cognitive therapy (the Feeling Safe Programme) with befriending for the treatment of persistent persecutory delusions: A parallel, single-blind, randomised controlled trial. *Lancet. Psychiatry*, 8(8), 696-707. [https://doi.org/10.1016/S2215-0366\(21\)00158-9](https://doi.org/10.1016/S2215-0366(21)00158-9)
- Freeman, D., Garety, P. A., Kuipers, P., Fowler, D., & Bebbington, P. E. (2002). A cognitive model of persecutory delusions. *The British Journal of Clinical Psychology*, 41(Pt 4), 331-347. <https://doi.org/10.1348/014466502760387461>
- Freeman, D., Garety, P. A., Kuipers, E., Fowler, D., Bebbington, P. E., & Dunn, G. (2007). Acting on persecutory delusions: The importance of safety seeking. *Behaviour Research and Therapy*, 45(1), 89-99. <https://doi.org/10.1016/j.brat.2006.01.014>
- Freeman, D., McManus, S., Brugha, T., Meltzer, H., Jenkins, R., & Bebbington, P. (2011). Concomitants of paranoia in the general population. *Psychological Medicine*, 41(5), 923-936. <https://doi.org/10.1017/S0033291710001546>
- Freeman, D., Pugh, K., Dunn, G., Evans, N., Sheaves, B., Waite, F., Cernis, E., Lister, R., & Fowler, D. (2014). An early Phase II randomised controlled trial testing the effect on persecutory delusions of using CBT to reduce negative cognitions about the self: The potential benefits of enhancing self-confidence. *Schizophrenia Research*, 160(1-3), 186-192. <https://doi.org/10.1016/j.schres.2014.10.038>
- Freeman, D., Taylor, K. M., Molodynski, A., & Waite, F. (2019). Treatable clinical intervention targets for patients with schizophrenia. *Schizophrenia Research*, 211, 44-50. <https://doi.org/10.1016/j.schres.2019.07.016>
- Freeman, D., Waite, F., Emsley, R., Kingdon, D., Davies, L., Fitzpatrick, R., & Dunn, G. (2016). The efficacy of a new translational treatment for persecutory delusions: Study protocol for a randomised controlled trial (The Feeling Safe Study). *Trials*, 17(1), 134. <https://doi.org/10.1186/s13063-016-1245-0>
- Freeman, D., Waite, F., Startup, H., Myers, E., Lister, R., McNerney, J., Harvey, A. G., Geddes, J., Zaiwalla, Z., Luengo-Fernandez, R., Foster, R., Clifton, L., & Yu, L. M. (2015). Efficacy of cognitive behavioural therapy for sleep improvement in patients with persistent delusions and hallucinations (BEST): A prospective, assessor-blind,

- randomised controlled pilot trial. *Lancet Psychiatry*, 2(11), 975-983.
[https://doi.org/10.1016/S2215-0366\(15\)00314-4](https://doi.org/10.1016/S2215-0366(15)00314-4)
- Freeston, M. (2023). What if we have too many models of worry and GAD? *Behavioural and Cognitive Psychotherapy*, 1-20. Advance online publication.
<https://doi.org/10.1017/S1352465822000649>
- Freud, S. (1957). *Neurosis and psychosis*. (A. Strachev & J. Strachev, Trans.). Collected papers, Vol. II. Hogarth Press.
- Friedman, R. S., & Lister, P. (1987). The current status of psychodynamic formulation. *Psychiatry*, 50(2), 126-141.
- Fritscher, L. (2020, December 2). *Object relations theory and the mom factor*. Verywellmind.
<https://www.verywellmind.com/what-is-object-relations-theory-2671995>
- Gaebel, W., Stricker, J., & Kerst, A. (2020). Changes from ICD-10 to ICD-11 and future directions in psychiatric classification^[P]_{SEP}. *Dialogues in Clinical Neuroscience*, 22(1), 7–15. <https://doi.org/10.31887/DCNS.2020.22.1/wgaebel>
- Garb, H. N. (1998). *Studying the clinician: Judgment research and psychological assessment*. American Psychological Association.
- Gates, V., Hsiao, M., Zieve, G. G., Courry, R., & Persons, J. B. (2021). Relationship to CBT outcome and dropout of decision support tools of the written case formulation, list of treatment goals and plot of symptom scores. *Behaviour Research and Therapy*, 142, 103874. <https://doi.org/10.1016/j.brat.2021.103874>
- Gibbs, M., Griffiths, M., & Dilks, S. (2020). A grounded theory of how service users experience and make use of formulation in therapy for psychosis. *Psychosis: Psychological, Social and Integrative Approaches*, 12(3), 245-256.
<https://doi.org/10.1080/17522439.2020.1742198>
- Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond “heuristics and biases”. *European Review of Social Psychology*, 2(1), 83-115.
<https://doi.org/10.1080/14792779143000033>
- Gilbert, P. & Leahy, R. L. (Eds.). (2007). *The therapeutic relationship in the cognitive behavioral psychotherapies*. Routledge.
- Gilhooly, K. (1983). *Thinking*. Academic Press.
- Goldsmith, L., Lewis, S., Dunn, G., & Bentall, R. (2015). Psychological treatments for early psychosis can be beneficial or harmful, depending on the therapeutic alliance: An instrumental variable analysis. *Psychological Medicine*, 45(11), 2365-2373.
<https://doi.org/10.1017/S003329171500032X>

- Graham, M. J., Milanowski, A. T., & Miller, J. B. (2012). *Measuring and promoting inter-rater agreement of teacher and principal performance ratings*. Center for Educator Compensation Reform. <https://files.eric.ed.gov/fulltext/ED532068.pdf>
- Grandjean, L., Hummel, J., Wyer, D., Beuchat, H., Caspar, F., Sachse, R., Berger, T., & Kramer, U. (2021). Psychotherapeutic case formulation: Plan analysis for narcissistic personality disorder. *Personality and Mental Health*, 15(4), 309-316. <https://doi.org/10.1002/pmh.1521>
- Grant, A., Townsend, M., Mills, J., & Cockx, A. (2010). *Assessment and care formulation in cognitive behavioural therapy*. SAGE.
- Greenberger, D., & Padesky, C. A. (1995). *Mind over mood: Change how you feel by changing the way you think*. Guilford Press.
- Grenyer, B. F. S., & Luborsky, L. (1998). Positive versus negative CCRT patterns. In L. Luborsky & P. Crits-Christoph (Eds.), *Understanding transference: The core conflictual relationship theme method* (2nd ed., pp. 55-63). American Psychological Association. <https://doi.org/10.1037/10250-004>
- Grove, W. M., Andreasen, N. C., McDonald-Scott, P., Keller, M. B., & Shapiro, R. W. (1981). Reliability studies of psychiatric diagnosis. Theory and practice. *Archives of General Psychiatry*, 38(4), 408-413. <https://doi.org/10.1001/archpsyc.1981.01780290042004>
- Guest, G., MacQueen, K. M., & Namey, E. E. (2012). *Applied thematic analysis*. SAGE Publications Ltd.
- Gumley, A., Braehler, C., Laithwaite, H., MacBeth, A., & Gilbert, P. (2010). A compassion focused model of recovery after psychosis. *International Journal of Cognitive Therapy*, 3(2), 186-201. <https://doi.org/10.1521/ijct.2010.3.2.186>
- Gumley, A., & Schwannauer, M. (2006). *Staying well after psychosis: A cognitive interpersonal approach to recovery and relapse prevention*. John Wiley & Sons.
- Gusenbauer, M., & Haddaway, N. R. (2020). Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. *Research Synthesis Methods*, 11(2), 181-217. <https://doi.org/10.1002/jrsm.1378>
- Guze, S. B. (1992). *Why psychiatry is a branch of medicine*. Oxford University Press.
- Haarhoff, B. (2008). *The map, the navigator, and the explorer: evaluating the content and quality of CBT case conceptualisations and the role of self-practice/self-reflection as a training intervention* [Doctoral thesis, Massey University]. Massey University, University of New Zealand. <https://mro.massey.ac.nz/handle/10179/1077>

- Haarhoff, B. A., Flett, R. A., & Gibson, K. L. (2011). Evaluating the content and quality of cognitive behavioural therapy case conceptualisations. *New Zealand Journal of Psychology*, 40(3), 104-114.
- Haddaway, N. R., Collins, A. M., Coughlin, D., & Kirk, S. (2015). The role of Google Scholar in evidence reviews and its applicability to grey literature searching. *PLoS ONE*, 10(9), e0138237. <https://doi.org/10.1371/journal.pone.0138237>
- Haddock, G., Eisner, E., Boone, C., Davies, G., Coogan, C., & Barrowclough, C. (2014). An investigation of the implementation of NICE-recommended CBT interventions for people with schizophrenia. *Journal of Mental Health*, 23(4), 162-165. <https://doi.org/10.3109/09638237.2013.869571>
- Halpin, E., Kugathasan, V., Hulbert, C., Alvarez-Jimenez, M., & Bendall, S. (2016). Case formulation in young people with post-traumatic stress disorder and first-episode psychosis. *Journal of Clinical Medicine*, 5(11), 106. <https://doi.org/10.3390/jcm5110106>
- Hariton, E., & Locascio, J. J. (2018). Randomised controlled trials - the gold standard for effectiveness research: Study design: Randomised controlled trials. *BJOG: An International Journal of Obstetrics and Gynaecology*, 125(13), 1716. <https://doi.org/10.1111/1471-0528.15199>
- Harper, D. & Spellman, D. (2014). Formulation and narrative therapy: Telling a different story. In L. Johnstone & R. Dallos (Eds.), *Formulation in psychology and psychotherapy: Making sense of people's problems* (2nd ed., pp. 96-120). Routledge.
- Harris, K., Collinson, C., & das Nair, R. (2012). Service-users' experiences of an early intervention in psychosis service: An interpretative phenomenological analysis. *Psychology and Psychotherapy*, 85(4), 456-469. <https://doi.org/10.1111/j.2044-8341.2011.02043.x>
- Hartley, S., Jovanoska, J., Roberts, S., Burden, N., & Berry, K. (2016). Case formulation in clinical practice: Associations with psychological mindedness, attachment and burnout in staff working with people experiencing psychosis. *Psychology and Psychotherapy*, 89(2), 133-147. <https://doi.org/10.1111/papt.12074>
- Hartmann, D. P. (1977). Considerations in the choice of interobserver reliability estimates. *Journal of Applied Behavior Analysis*, 10(1), 103-116. <https://doi.org/10.1901/jaba.1977.10-103>

- Hasson, F., & Keeney, S. (2011). Enhancing rigor in the Delphi technique research. *Technology Forecasting and Social Change*, 78(9), 1695-1704.
<https://doi.org/10.1016/j.techfore.2011.04.005>
- Hasson, F., Keeney, S., & McKenna, H. (2000). Research guidelines for the Delphi survey technique. *Journal of Advanced Nursing*, 32(4), 1008-1015.
<https://doi.org/10.1046/j.1365-2648.2000.t01-1-01567.x>
- Hazell, C. M., Hayward, M., Cavanagh, K., & Strauss, C. (2016). A systematic review and metaanalysis of low intensity CBT for psychosis. *Clinical Psychology Review*, 45, 183-192. <https://doi.org/10.1016/j.cpr.2016.03.004>
- Heatherington, L., & Johnson, B. (2019). Social constructionism in couple and family therapy: Narrative, solution-focused, and related approaches. In B. H. Fiese, M. Celano, K. Deater-Deckard, E. N. Jouriles, & M. A. Whisman (Eds.), *APA handbook of contemporary family psychology: Foundations, methods, and contemporary issues across the lifespan* (pp. 127-142). American Psychological Association. <https://doi.org/10.1037/0000099-008>
- Hegarty, B. D., Marceau, E. M., Gusset, M., & Grenyer, B. F. S. (2020). Early treatment response in psychotherapy for depression and personality disorder: Links with core conflictual relationship themes. *Psychotherapy Research*, 30(1), 112-123.
<https://doi.org/10.1080/10503307.2019.1609114>
- Ho, B. C., Andreasen, N. C., Ziebell, S., Pierson, R., & Magnotta, V. (2011). Long-term antipsychotic treatment and brain volumes: A longitudinal study of first-episode schizophrenia. *Archives of General Psychiatry*, 68(2), 128-137.
<https://doi.org/10.1001/archgenpsychiatry.2010.199>
- Hodgetts, A., & Wright, J. (2007). Researching clients' experiences: A review of qualitative studies. *Clinical Psychology & Psychotherapy*, 14(3), 157-163.
<https://doi.org/10.1002/cpp.527>
- Holloway, F. (2005). *The forgotten need for rehabilitation in contemporary mental health services: A position statement from the executive committee of the faculty of rehabilitation and social psychiatry*. Royal College of Psychiatrists.
- Holmes, E. A., Ghaderi, A., Harmer, C. J., Ramchandani, P. G., Cuijpers, P., Morrison, A. P., Roiser, J. P., Bockting, C., O'Connor, R. C., Shafran, R., Moulds, M. L., & Craske, M. G. (2018). The Lancet Psychiatry Commission on psychological treatments research in tomorrow's science. *Lancet Psychiatry*, 5(3), 237-286. [https://doi.org/10.1016/S2215-0366\(17\)30513-8](https://doi.org/10.1016/S2215-0366(17)30513-8)

- Horowitz, M. J. (1989). Relationship schema formulation: Role-relationship models and intrapsychic conflict. *Psychiatry: Interpersonal and Biological Processes*, 52(3), 260-274.
- Horowitz, M. J. (1991a). Short-term dynamic therapy of stress response syndromes. In P. Crits-Christoph & J. P. Barber (Eds.), *Handbook of short-term dynamic psychotherapy* (pp. 166-198). Basic Books.
- Horowitz, M. J. (Ed.). (1991b). *Person schemas and maladaptive interpersonal patterns*. The University of Chicago Press.
- Howard, A., Forsyth, A., **Spencer, H.**, Wisniewska Young, E., & Turkington, D. (2012). Do voice hearers naturally use focusing and metacognitive coping techniques? *Psychosis*, 5(2), 119-6. <https://doi.org/10.1080/17522439.2012.668926>
- Hsu, C.-C., & Sandford, B. A. (2007). The Delphi technique: Making sense of consensus. *Practical Assessment, Research & Evaluation*, 12(10), 1-8. <https://doi.org/10.7275/pdz9-th90>
- Hugdahl, K. (2015). [Auditory hallucinations: A review of the ERC "VOICE" project](https://doi.org/10.5498/wjp.v5.i2.193). *World Journal of Psychiatry*, 5(2), 193-209. <https://doi.org/10.5498/wjp.v5.i2.193>
- Humphrey-Murto, S., & de Wit, M. (2019). The Delphi method – more research please. *Journal of Clinical Epidemiology*, 106, 136-139. <https://doi.org/10.1016/j.jclinepi.2018.10.011>
- Insel T. (2013, April 29). *Director's blog: Transforming diagnosis*. National Institute of Mental Health. <http://www.nimh.nih.gov/about/director/2013/transforming-diagnosis.shtml>
- Iqbal, S., & Pipon-Young, L. (2009). 'The Delphi method: A step-by-step guide'. *The Psychologist*, 22(7), 598-601.
- Iyegbe, C. O., & O'Reilly, P. F. (2022). Genetic origins of schizophrenia find common ground. *Nature*, 604(7906), 433-435. <https://doi.org/10.1038/d41586-022-00773-5>
- Jablensky A. (2016). Psychiatric classifications: Validity and utility. *World Psychiatry: Official Journal of the World Psychiatric Association (WPA)*, 15(1), 26-31. <https://doi.org/10.1002/wps.20284>
- Jacoby, I. (1988). Evidence and consensus. *The Journal of the American Medical Association*, 259(20), 3039. <https://doi.org/10.1001/jama.1988.03720200061036>
- James, I. A. (2001). Schema Therapy: The next generation, but should it carry a health warning? *Behavioural and Cognitive Psychotherapy*, 29(4), 401-407. <http://doi.org/10.1017/S1352465801004015>

- Jaspers, K. (1997). *General psychopathology*. (J. Hoenig & M. W. Hamilton, Trans.). The Johns Hopkins University Press. (Original work published 1963).
- Jefferis, S., Fantarrow, Z., & Johnston, L. H. (2021). The torchlight model of mapping in cognitive analytic therapy (CAT) reformulation: A qualitative investigation. *Psychology and Psychotherapy*, 94(S1), 137-150. <https://doi.org/10.1111/papt.12311>
- Jenaway, A. (2019, January 21). CAT concepts, reciprocal roles: Relationship roles. Association for cognitive analytic therapy ltd. <https://www.engage.acat.org.uk/relationship-roles/>
- Jenkins, P. (2020, November). *Single session formulation: An alternative to the waiting list*. BACP: University and College Counselling. <https://www.bacp.co.uk/bacp-journals/university-and-college-counselling/november-2020/single-session-formulation/>
- Jenkins, D., & Smith, T. E. (1994). Applying Delphi methodology in family therapy research. *Contemporary Family Therapy*, 16(5), 411-430. <https://doi.org/10.1007/BF02197902>
- Johns, L. C., Kompus, K., Connell, M., Humpston, C., Lincoln, T. M., Longden, E., Preti, A., Alderson-Day, B., Badcock, J. C., Cella, M., Fernyhough, C., McCarthy-Jones, S., Peters, E., Raballo, A., Scott, J., Siddi, S., Sommer, I. E., & Larøi, F. (2014). Auditory verbal hallucinations in persons with and without a need for care. *Schizophrenia Bulletin*, 40(Suppl 4), S255–S264. <https://doi.org/10.1093/schbul/sbu005>
- Johnstone, L. (2014a). *A straight-talking introduction to psychiatric diagnosis*. PCCS Books.
- Johnstone, L. (2014b). Using formulation in teams. In N. Tarrier & J. Johnson (Eds.), *Case formulation in cognitive behaviour therapy: The treatment of challenging and complex cases* (2nd ed., pp. 216-242). Routledge.
- Johnstone, L. (2018). Psychological formulation as an alternative to psychiatric diagnosis. *Journal of Humanistic Psychology*, 58(1), 30-46. <https://doi.org/10.1177/0022167817722230>
- Johnstone, L. (2021, May 5). *Beyond the mental health paradigm*. IAI News. <https://iai.tv/articles/beyond-the-mental-health-paradigm-the-power-threat-meaning-framework-auid-1803>
- Johnstone, L. (2022). *Users and abusers of psychiatry: A critical look at psychiatric practice* (Classic Edition). Routledge.
- Johnstone, L., Boyle, M., Cromby, J., Dillon, J., Harper, D., Kinderman, P., Longden, E., Pilgrim, D., & Read, J. (2018). *The power threat meaning framework: Overview*. British Psychological Association.

<https://www.bps.org.uk/sites/www.bps.org.uk/files/Policy/Policy%20-%20Files/PTM%20Overview.pdf>Kaiser

- Johnstone, L., & Dallos, R. (2014). Introduction to formulation. In L. Johnstone & R. Dallos (Eds.), *Formulation in psychology and psychotherapy: Making sense of people's problems* (2nd ed., pp. 1-17). Routledge.
- Jones K. D. (2012). A critique of the DSM-5 field trials. *The Journal of Nervous and Mental Disease*, 200(6), 517-519. <https://doi.org/10.1097/NMD.0b013e318257c699>
- Jones, J., & Hunter, D. (1995). Consensus methods for medical and health services research. *The BMJ*, 311(7001), 376-380. <https://doi.org/10.1136/bmj.311.7001.376>
- Jones, N., Shattell, M., Kelly, T., Brown, R., Robinson, L., Renfro, R., Harris, B., & Luhrmann, T. M. (2016). "Did I push myself over the edge?": Complications of agency in psychosis onset and development. *Psychosis: Psychological, Social and Integrative Approaches*, 8(4), 324-335. <https://doi.org/10.1080/17522439.2016.1150501>
- Jünger, S., Payne, S. A., Brine, J., Radbruch, L., & Brearley, S. G. (2017). Guidance on Conducting and REporting DELphi Studies (CREDES) in palliative care: Recommendations based on a methodological systematic review. *Palliative Medicine*, 31(8), 684-706. <https://doi.org/10.1177/0269216317690685>
- Kahlon, S., Neal, A., & Patterson, T. (2014). Experiences of cognitive behavioural therapy formulation in clients with depression. *The Cognitive Behaviour Therapist*, 7, e8. <http://doi.org/10.1017/S1754470X14000075>
- Kahneman D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *The American psychologist*, 58(9), 697-720. <https://doi.org/10.1037/0003-066X.58.9.697>
- Kay, S. R., Fiszbein, A., & Opler, L. A. (1987). The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophrenia Bulletin*, 13(2), 261-276. <https://doi.org/10.1093/schbul/13.2.261>
- Kannis-Dymand, L., Myers, A., Human, B., Love, S., & Lovell, G. P. (2021). Evaluation of collaborative cognitive-behavioural case formulations in a clinical psychology program: A client perspective. *Clinical Psychologist*, 25(3), 339-349. <https://doi.org/10.1080/13284207.2021.1981745>
- Kazdin, A. E. (2023). Drawing causal inferences from randomized controlled trials in psychotherapy research. *Psychotherapy Research*, 33(8), 991-1003. <https://doi.org/10.1080/10503307.2022.2130112>

- Keeney, S., Hasson, F., McKenna (2011). *The Delphi technique in nursing and health research*. John Wiley & Sons.
- Kendell, R. E., Everett, B., Cooper, J. E., Sartorius, N., & David, M. E. (1968). The reliability of the "Present State Examination." *Social Psychiatry*, 3(3), 123-129. <https://doi.org/10.1007/BF00577836>
- Kendell, R., & Jablensky, A. (2003). Distinguishing between the validity and utility of psychiatric diagnoses. *The American Journal of Psychiatry*, 160(1), 4-12. <https://doi.org/10.1176/appi.ajp.160.1.4>
- Kendjelic, E. M., & Eells, T. D. (2007). Generic psychotherapy case formulation training improves formulation quality. *Psychotherapy: Theory, Research, Practice, Training*, 44(1), 66-77. <https://doi.org/10.1037/0033-3204.44.1.66>
- Kennerley, H., Kirk, J., & Westbrook, D. (2016). *An introduction to cognitive behaviour therapy: Skills and applications*. SAGE Publications Ltd.
- Kerr, I. B., Birkett, P. B. L., & Chanen, A. (2003). Clinical and service implications of a cognitive analytic therapy model of psychosis. *The Australian and New Zealand Journal of Psychiatry*, 37(5), 515-523. <https://doi.org/10.1046/j.1440-1614.2003.01200.x>
- Kerr, I. B., Crowley, V., & Beard, H. (2006). A cognitive analytic therapy-based approach to psychotic disorder. In J. O. Johannessen, B. V. Martindale, & J. Cullberg (Eds.), *Evolving psychosis: Different stages, different treatments* (pp. 172-184). Routledge.
- Kidder, L. H., & Fine, M. (1987). Qualitative and quantitative methods. When stories converge. In M. M. Mark & L. Shotland (Eds.), *New directions for program evaluation* (pp. 57-75). Jossey-Bass.
- Kinderman, P. (2011). Cognitive models of auditory hallucinations. In R. Hagen, D. Turkington, T. Berge, & R. W. Gråwe (Eds.), *CBT for psychosis: A symptom-based approach* (pp. 12-21). Routledge.
- Kinderman, P., & Allsopp, K. (2018). Non-diagnostic recording of mental health difficulties in ICD-11. *Lancet Psychiatry*, 5(12), 966. [https://doi.org/10.1016/S2215-0366\(18\)30394-8](https://doi.org/10.1016/S2215-0366(18)30394-8)
- Kinderman, P., & Lobban, F. (2000). Evolving formulations: Sharing complex information with clients. *Behavioural and Cognitive Psychotherapy*, 28(3), 307-310. <http://doi.org/10.1017/S1352465800003118>
- Kinderman, P., Read, J., Moncrieff, J., & Bentall, R. P. (2013). Drop the language of disorder. *Evidence-Based Mental Health*, 16(1), 2-3. <https://doi.org/10.1136/eb-2012-100987>

- King, N., & Brooks, J. M. (2017). *Template analysis for business and management students*. SAGE Publications Ltd.
- Kingdon, D., Gibson, A., Kinoshita, Y., Turkington, D., Rathod, S., & Morrison, A. (2008). Acceptable terminology and subgroups in schizophrenia: An exploratory study. *Social Psychiatry and Psychiatric Epidemiology*, 43(3), 239-243.
<https://doi.org/10.1007/s00127-007-0284-y>
- Kingdon, D. G., & Turkington, D. (1999). Cognitive behavioural therapy of schizophrenia: Styles and methods. In T. Wykes, N. Tarrier, & S. Lewis (Eds.), *Outcome and innovation in psychological treatment of schizophrenia* (pp. 59-80). John Wiley & Sons.
- Kingdon, D. G., & Turkington, D. (1994). *Cognitive-behavioral therapy of schizophrenia*. Guilford Press.
- Kingdon, D. G., & Turkington, D. (2008). *Cognitive therapy of schizophrenia*. Guilford Press.
- Kirk, S. A., & Kutchins, H. (1997). *Making us crazy: DSM—The psychiatric bible and the creation of mental disorder*. Free Press.
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155-163.
<https://doi.org/10.1016/j.jcm.2016.02.012>
- Korver-Nieberg, N., Quee, P. J., Boos, H. B., Simons, C. J., & GROUP. (2011). The validity of the DSM-IV diagnostic classification system of non-affective psychoses. *The Australian and New Zealand Journal of Psychiatry*, 45(12), 1061-1068.
<https://doi.org/10.3109/00048674.2011.620562>
- Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., Brown, T. A., Carpenter, W. T., Caspi, A., Clark, L. A., Eaton, N. R., Forbes, M. K., Forbush, K. T., Goldberg, D., Hasin, D., Hyman, S. E., Ivanova, M. Y., Lynam, D. R., Markon, K., . . . Zimmerman, M. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, 126(4), 454-477. <https://doi.org/10.1037/abn0000258>
- Kozáková, E., Bakštein, E., Havlíček, O., Bečev, O., Knytl, P., Zaytseva, Y., & Španiel, F. (2020). Disrupted sense of agency as a state marker of first-episode schizophrenia: A large-scale follow-up study. *Frontiers in Psychiatry*, 11, 570570.
<https://doi.org/10.3389/fpsy.2020.570570>

- Kraemer, H. C., Kupfer, D. J., Clarke, D. E., Narrow, W. E., & Regier, D. A. (2012). DSM-5: how reliable is reliable enough?. *The American Journal of Psychiatry*, 169(1), 13-15. <https://doi.org/10.1176/appi.ajp.2011.11010050>
- Kraepelin, E. (1919). *Dementia praecox and paraphrenia*. Thoenmes Continuum.
- Kuyken, W. (2006). Evidence-based case formulation: Is the emperor clothed? In N. Tarrier (Ed.), *Case formulation in cognitive behaviour therapy: The treatment of challenging and complex cases* (pp. 12-35). Routledge/Taylor & Francis Group.
- Kuyken, W., Beshai, S., Dudley, R., Abel, A., Görg, N., Gower, P., McManus, F., & Padesky, C. A. (2016). Assessing competence in collaborative case conceptualization: Development and preliminary psychometric properties of the Collaborative Case Conceptualization Rating Scale (CCC-RS). *Behavioural and Cognitive Psychotherapy*, 44(2), 179-192. <https://doi.org/10.1017/S1352465814000691>
- Kuyken, W., Fothergill, C. D., Musa, M., & Chadwick, P. (2005). The reliability and quality of cognitive case formulation. *Behaviour Research and Therapy*, 43(9), 1187-1201. <https://doi.org/10.1016/j.brat.2004.08.007>
- Kuyken, W., Padesky, C. A., & Dudley, R. (2009). *Collaborative case conceptualization: Working effectively with clients in cognitive-behavioral therapy*. Guilford Press.
- Laikin, M., Winston, A., & McCullough, L. (1991). Intensive short-term dynamic psychotherapy. In P. Crits-Christoph & J. P. Barber (Eds.), *Handbook of short-term dynamic psychotherapy* (pp. 80-109). Basic Books.
- Landau, S., & Emsley, R. (2022). Causal inference: Efficacy and mechanism evaluation. In S. Piantadosi & C. L. Meinert (Eds.), *Principles and practice of clinical trials* (pp. 1981-2002). Springer. https://doi.org/10.1007/978-3-319-52636-2_137
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-174.
- Langlands, R. L., Jorm, A. F., Kelly, C. M., & Kitchener, B. A. (2008). First aid recommendations for psychosis: Using the Delphi method to gain consensus between mental health consumers, carers, and clinicians. *Schizophrenia Bulletin*, 34(3), 435-443. <https://doi.org/10.1093/schbul/sbm099>
- Law, H., & Morrison, A. (2014). Recovery in psychosis: A Delphi study with experts by experience. *Schizophrenia Bulletin*, 40(6), 1347-1355. <https://doi.org/10.1093/schbul/sbu047>
- Lawlis, G. F., & Lu, E. (1972). Judgment of counseling process: Reliability, agreement, and error. *Psychological Bulletin*, 78(1), 17-20. <https://doi.org/10.1037/h0032935>

- Lawlor, C., Hall, K., & Ellett, L. (2015). Paranoia in the therapeutic relationship in cognitive behavioural therapy for psychosis. *Behavioural and Cognitive Psychotherapy*, 43(4), 490-501. <https://doi.org/10.1017/S1352465814000071>
- Lebert, L., Turkington, D., Freeston, M., & Dudley, R. (2021) Rumination, intolerance of uncertainty and paranoia in treatment resistant psychosis. *Psychosis*, 13(1), 65-70. <https://doi.org/10.1080/17522439.2020.1798489>
- Leiper, R. (2014). Psychodynamic formulation: Looking beneath the surface. In L. Johnstone & R. Dallos (Eds.), *Formulation in psychology and psychotherapy: Making sense of people's problems* (2nd ed., pp. 45-66). Routledge.
- Leiper, R., & Maltby, M. (2004). *The psychodynamic approach to therapeutic change*. Sage.
- Levenson, H., & Strupp, H. H. (2010). Cyclical maladaptive patterns: Case formulation in time-limited dynamic psychotherapy. In T. D. Eells (Eds.), *Handbook of psychotherapy case formulation* (2nd ed., pp. 164-197). Guilford.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 22(140), 5-55.
- Lincoln, T. M., Hartmann, M., Köther, U., & Moritz, S. (2015). Do People with psychosis have specific difficulties regulating emotions? *Clinical Psychology & Psychotherapy*, 22(6), 637-646. <https://doi.org/10.1002/cpp.1923>
- Linscott, R. J., & van Os, J. (2010). Systematic reviews of categorical versus continuum models in psychosis: Evidence for discontinuous subpopulations underlying a psychometric continuum. Implications for DSM-V, DSM-VI, and DSM-VII. *Annual Review of Clinical Psychology*, 6, 391-419. <https://doi.org/10.1146/annurev.clinpsy.032408.153506>
- Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). John Wiley & Sons.
- López-López, J. A., Davies, S. R., Caldwell, D. M., Churchill, R., Peters, T. J., Tallon, D., Dawson, S., Wu, Q., Li, J., Taylor, A., Lewis, G., Kessler, D. S., Wiles, N., & Welton, N. J. (2019). The process and delivery of CBT for depression in adults: A systematic review and network meta-analysis. *Psychological Medicine*, 49(12), 1937-1947. <https://doi.org/10.1017/S003329171900120X>
- Low, J. (2019). A pragmatic definition of the concept of theoretical saturation. *Sociological Focus*, 52(2), 131-139. <https://doi.org/10.1080/00380237.2018.1544514>

- Luborsky, L. (1977). Measuring a pervasive psychic structure in psychotherapy: The core conflictual relationship theme. In N. Freedman & S. Grand (Eds.), *Communicative structures and psychic structures* (pp. 367-395). Plenum Press.
- Luborsky, L., Barber, J. P., Binder, J., Curtis, J., Dahl, H., Horowitz, L. M., & Teller, V. (1993). Transference-related measures: A new class based on psychotherapy sessions. In N. E. Miller, L. Luborsky, J. P. Barber, & J. P. Docherty (Eds.), *Psychodynamic treatment research: A handbook for clinical practice* (pp. 326-341). Basic Books.
- Luborsky, L., & Crits-Christoph, P. (1990). *Understanding transference: The core conflictual relationship theme method*. Basic Books.
- Luborsky, L., & Crits-Christoph, P. (1998). The early life of the idea for the core conflictual relationship theme method. In L. Luborsky & P. Crits-Christoph (Eds.), *Understanding transference: The core conflictual relationship theme method* (2nd ed., pp. 3-13). American Psychological Association. <https://doi.org/10.1037/10250-001>
- Luborsky, L., & Diguer, L. (1998). The reliability of the core conflictual relationship theme method measure: Results from eight samples. In L. Luborsky & P. Crits-Christoph (Eds.), *Understanding transference: The core conflictual relationship theme method* (2nd ed., pp. 97-107). American Psychological Association. <https://doi.org/10.1037/10250-001>
- Luborsky, L., Mark, D., Hole, A. V., Popp, C. A., Goldsmith, B., & Cacciola, J. (1995). Supportive-expressive dynamic psychotherapy for depression: Axis I. In J. P. Barber & P. Crits-Christoph (Eds.), *Dynamic therapies for psychiatric disorders: Axis I* (pp. 13-42). Basic Books.
- Luce, A., & Barclay, N. (2022). 'I just feel like there's just lots of people in my head!'. In A. Woods, B. Alderson-Day & C. Fernyhough (Eds.), *Voices in psychosis: Interdisciplinary perspectives* (pp. 184-193). Oxford University Press.
- Macneil, C. A., Hasty, M. K., Conus, P., & Berk, M. (2012). Is diagnosis enough to guide interventions in mental health? Using case formulation in clinical practice. *BMC Medicine*, 10(111), 1-3. <https://doi.org/10.1186/1741-7015-10-111>
- Madill, A., Widdicombe, S., & Barkham, M. (2001). The potential of conversation analysis for psychotherapy research. *The Counseling Psychologist*, 29(3), 413-434. <https://doi.org/10.1177/0011000001293006>
- Makkena, A. (2024). The stalwart – Adolf Meyer. *Telangana Journal of Psychiatry*, 10(1), 96-99. https://doi.org/10.4103/tjp.tjp_34_24

- Malan, D. (1959). On assessing the results of psychotherapy. *The British Journal of Medical Psychology*, 32(2), 86-105. <https://doi.org/10.1111/j.2044-8341.1959.tb00472.x>
- Malan, D. H. (1973). The outcome problem in psychotherapy research: A historical review. *Archives of General Psychiatry*, 29(6), 719-729. <https://doi.org/10.1001/archpsyc.1973.04200060005001>
- Malan, D. H. (1976). *Toward the validation of dynamic psychotherapy: A replication*. Plenum Press.
- Mann, J. (1973). *Time limited psychotherapy*. Harvard University Press.
- Mann, J., & Goldman, K. (1982). *A case book in time limited psychotherapy*. McGraw-Hill.
- Markowitz, L. J., & Purdon, C. (2008). Predictors and consequences of suppressing obsessional thoughts. *Behavioural and Cognitive Psychotherapy*, 36(2), 179-192. <https://doi.org/10.1017/S1352465807003992>
- Markon, K. E., Chmielewski, M., & Miller, C. J. (2011). The reliability and validity of discrete and continuous measures of psychopathology: A quantitative review. *Psychological Bulletin*, 137(5), 856-879. <https://doi.org/10.1037/a0023678>
- Marshall, M., Lockwood, A., Lewis, S., & Fiander, M. (2004). Essential elements of an early intervention service for psychosis: The opinions of expert clinicians. *BMC Psychiatry*, 4(17), 1-7. <https://doi.org/10.1186/1471-244X-4-17>
- Marwaha, S., Broome, M. R., Bebbington, P. E., Kuipers, E., & Freeman, D. (2014). Mood instability and psychosis: Analyses of British national survey data. *Schizophrenia Bulletin*, 40(2), 269-277. <https://doi.org/10.1093/schbul/sbt149>
- Marwaha, S., Johnson, S., Bebbington, P., Stafford, M., Angermeyer, M. C., Brugha, T., Azorin, J-M., Kilian, R., Hansen, K., & Toumi M. (2007). Rates and correlates of employment in people with schizophrenia in the UK, France and Germany. *British Journal of Psychiatry*, 191(1), 30-37. <https://doi.org/10.1192/bjp.bp.105.020982>
- Matsunaga, A., & Kitamura, T. (2016). The effects of symptoms, diagnostic labels, and education in psychiatry on the stigmatization towards schizophrenia: A questionnaire survey among a lay population in Japan. *Mental Illness*, 8(1), 6344. <https://doi.org/10.4081/mi.2016.6344>
- McCabe, R., & Priebe, S. (2004). Explanatory models of illness in schizophrenia: Comparison of four ethnic groups. *The British Journal of Psychiatry*, 185, 25-30. <https://doi.org/10.1192/bjp.185.1.25>

- McCarthy, M. (2013). Director of top research organization for mental health criticizes DSM for lack of validity. *BMJ (Clinical research ed.)*, 346, f2954.
<https://doi.org/10.1136/bmj.f2954>
- Mc Glanaghy, E., Turner, D., Davis, G. A., Sharpe, H., Dougall, N., Morris, P., Prentice, W., & Hutton, P. (2021). A network meta-analysis of psychological interventions for schizophrenia and psychosis: Impact on symptoms. *Schizophrenia Research*, 228, 447-459. <https://doi.org/10.1016/j.schres.2020.12.036>
- McGlashan, T. H., Docherty, J. P., & Siris, S. (1976). Integrative and sealing-over recoveries from schizophrenia: Distinguishing case studies. *Psychiatry*, 39(4), 325-338.
<https://doi.org/10.1080/00332747.1976.11023903>
- McHugh M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, 22(3), 276-282.
- McMurran, M., & Bruford, S. (2016). Case formulation quality checklist: A revision based upon clinicians' views. *Journal of Forensic Practice*, 18(1), 31-38.
<https://doi.org/10.1108/JFP-05-2015-0027>
- McTiernan, K., Gullon-Scott, F., & Dudley, R. (2020). An exploration of strength use and its relationship with life satisfaction, positive self-beliefs and paranoia. *International Journal of Wellbeing*, 10(2), 53-70. <https://doi.org/10.5502/ijw.v10i2.883>
- Mellsop, G. W., & Banzato, C. E. (2006). A concise conceptualization of formulation. *Academic Psychiatry*, 30(5), 424-425.
<https://doi.org/10.1176/appi.ap.30.5.424>
- Meltzer, H., Singleton, N., Lee, A., Bebbington, P., Brugha, T., & Jenkins R. (2002). *The social and economic circumstances of adults with mental disorders*. Stationery Office.
<https://doi.org/10.1037/e623732007-001>
- Messer, S. B., & Warren, C. S. (1995). *Models of brief psychodynamic therapy: A comparative approach*. Guilford Press.
- methodsMcr. (2013, February 23). *What is causal inference by Dr Richard Emsley* [Presentation]. methodsMcr. YouTube.
https://www.youtube.com/watch?v=Q3_WlIzSxHk
- Meyer, A. (1931). Psychobiology. The Thomas Salmon lectures. In E. E. Winters & S. Bowers (Eds.), *Psychobiology: A science of man*. Charles C. Thomas.
- Meyer, A. (1951). *The collected papers of Adolf Meyer - Volume III: Medical Teaching*. Johns Hopkins Press.

- Meyer, V. (1966). Modification of expectations in cases with obsessional rituals. *Behaviour Research and Therapy*, 4(4), 273-280. [https://doi.org/10.1016/0005-7967\(66\)90023-4](https://doi.org/10.1016/0005-7967(66)90023-4)
- Meyer, V. (1975). The impact of research on the clinical application of behaviour therapy. In R. I. Thompson & W. S. Dockens (Eds.), *Applications of behaviour modification*. Academic Press.
- Meyer, V., & Turkat, I. D. (1979). Behavioral analysis of clinical cases. *Journal of Behavioral Assessment*, 1(4), 259-270. <https://doi.org/10.1007/BF01321368>
- Mind. (2008). *In the red: Debt and mental health*. Mind.
- Miranda, J., Gross, J. J., Persons, J. B., & Hahn, J. (1998). Mood matters: Negative mood induction activates dysfunctional attitudes in women vulnerable to depression. *Cognitive Therapy and Research*, 22(4), 363-376. <https://doi.org/10.1023/A:1018709212986>
- Miranda, J., Persons, J. B., & Byers, C. N. (1990). Endorsement of dysfunctional beliefs depends on current mood state. *Journal of Abnormal Psychology*, 99(3), 237-241. <https://doi.org/10.1037//0021-843x.99.3.237>
- Mittal, D., Corrigan, P., Sherman, M. D., Chekuri, L., Han, X., Reaves, C., Mukherjee, S., Morris, S., & Sullivan, G. (2014). Healthcare providers' attitudes toward persons with schizophrenia. *Psychiatry Rehabilitation Journal*, 37(4), 297-303. <https://doi.org/10.1037/prj0000095>
- Morrison, A. P. (1998). A cognitive analysis of the maintenance of auditory hallucinations: Are voices to schizophrenia what bodily sensations are to panic? *Behavioural and Cognitive Psychotherapy*. 26(4), 289-302. <http://doi.org/10.1017/S1352465898264010>
- Morrison, A. (2001). The interpretation of intrusions in psychosis: An integrative cognitive approach to hallucinations and delusions. *Behavioural and Cognitive Psychotherapy*, 29(3), 257-276. <https://doi:10.1017/S1352465801003010>
- Morrison, A. P. (2007). Case formulation and cognitive schemas in cognitive therapy for psychosis. In L. P. Riso, P. L. du Toit, D. J. Stein, & J. E. Young (Eds.), *Cognitive schemas and core beliefs in psychological problems: A scientist-practitioner guide* (pp. 177-197). American Psychological Association.
- Morrison, A. P. (2017). A manualised treatment protocol to guide delivery of evidence-based cognitive therapy for people with distressing psychosis: Learning from clinical trials. *Psychosis: Psychological, Social and Integrative Approaches*, 9(3), 271-281. <https://doi.org/10.1080/17522439.2017.1295098>

- Morrison, A. P., & Barratt, S. (2010). What are the components of CBT for psychosis? A Delphi study. *Schizophrenia Bulletin*, 36(1), 136-142.
<https://doi.org/10.1093/schbul/sbp118>
- Morrison, A. P., Gonçalves, C. C., Peel, H., Larkin, A., & Bowe, S. E. (2023). Identifying types of problems and relative priorities in the problem lists of participants in CBT for psychosis trials. *Behavioural and Cognitive Psychotherapy*, 1-12. Advance online publication. <https://doi.org/10.1017/S1352465822000583>
- Morrison, A. P., Gumley, A. I., Schwannauer, M., Campbell, M., Gleeson, A., Griffin, E., & Gillan, K. (2005). The beliefs about paranoia scale: Preliminary validation of a metacognitive approach to conceptualizing paranoia. *Behavioural and Cognitive Psychotherapy*, 33(2), 153-164. <http://doi.org/10.1017/S1352465804001900>
- Morrison, A. P., Hutton, P., Shiers, D., & Turkington, D. (2012). Antipsychotics: Is it time to introduce patient choice? *The British Journal of Psychiatry*, 201, 83-84.
<https://doi.org/10.1192/bjp.bp.112.112110>
- Morrison, A. P., Hutton, P., Wardle, M., **Spencer, H.**, Barratt, S., Brabban, A., Callcott, P., Christodoulides, T., Dudley, R., French, P., Lumley, V., Tai, S. J., & Turkington, D. (2012). Cognitive therapy for people with a schizophrenia spectrum diagnosis not taking antipsychotic medication: An exploratory trial. *Psychological Medicine*, 42(5), 1049-1056. <https://doi.org/10.1017/S0033291711001899>
- Morrison, A. P., Law, H., Carter, L., Sellers, R., Emsley, R., Pyle, M., French, P., Shiers, D., Yung, A. R., Murphy, E. K., Holden, N., Steele, A., Bowe, S. E., Palmier-Claus, J., Brooks, V., Byrne, R., Davies, L., & Haddad, P. M. (2018). Antipsychotic drugs versus cognitive behavioural therapy versus a combination of both in people with psychosis: A randomised controlled pilot and feasibility study. *Lancet Psychiatry*, 5(5), 411-423.
[https://doi.org/10.1016/S2215-0366\(18\)30096-8](https://doi.org/10.1016/S2215-0366(18)30096-8)
- Morrison, A. P., Pyle, M., Maughan, D., Johns, L., Freeman, D., Broome, M. R., Husain, N., Fowler, D., Hudson, J., MacLennan, G., Norrie, J., Shiers, D., Hollis, C., James, A., & MAPS group. (2020). Antipsychotic medication versus psychological intervention versus a combination of both in adolescents with first-episode psychosis (MAPS): A multicentre, three-arm, randomised controlled pilot and feasibility study. *Lancet. Psychiatry*, 7(9), 788-800. [https://doi.org/10.1016/S2215-0366\(20\)30248-0](https://doi.org/10.1016/S2215-0366(20)30248-0)
- Morrison, A. P., Renton, J. C., Dunn, H., Williams, S., & Bentall, R. P. (2004). *Cognitive therapy for psychosis: A formulation-based approach*. Routledge.

- Morrison, A. P., Renton, J. C., French, P., & Bentall, R. (2008). *Think you're crazy? Think again: A resource book for cognitive therapy for psychosis*. Routledge.
- Morrison, A.P., Turkington, D., Pyle, M., **Spencer, H.**, Brabban, A., Dunn, G., Christodoulides, T., Dudley, R., Chapman, N., Callcott, P., Grace, T., Lumley, V., Drage, L., Tully, S., Irving, K., Cummings, A., Byrne, R., Davies, L. M., & Hutton, P. (2014). Cognitive therapy for people with schizophrenia spectrum disorders not taking antipsychotic drugs: A single-blind randomised controlled trial. *Lancet*, 383(9926), 1395-1403. [http://dx.doi.org/10.1016/S0140-6736\(13\)62246-1](http://dx.doi.org/10.1016/S0140-6736(13)62246-1)
- Morrison, A. P., Turkington, D., Wardle, M., **Spencer, H.**, Barratt, S., Dudley, R., Brabban, A., & Hutton, P. (2012). A preliminary exploration of predictors of outcome and cognitive mechanisms of change in cognitive behaviour therapy for psychosis in people not taking antipsychotic medication. *Behaviour Research and Therapy*, 50(2), 163-167. <https://doi.org/10.1016/j.brat.2011.12.001>
- Morrison, A. P., Wardle, M., Hutton, P., Davies, L., Dunn, G., Brabban, A., Byrne, R., Drage, L., **Spencer, H.**, & Turkington, D. (2013). Assessing cognitive therapy instead of neuroleptics: Rationale, study design and sample characteristics of the ACTION trial. *Psychosis: Psychological, Social and Integrative Approaches*, 5(1), 82-91. <http://dx.doi.org/10.1080/17522439.2012.756539>
- Mosolov, S. N., & Yaltonskaya, P. A. (2022). Primary and secondary negative symptoms in schizophrenia. *Frontiers in Psychiatry*, 12, 766692. <https://doi.org/10.3389/fpsy.2021.766692>
- Mosri, D. F. (2021). Clinical applications of neuropsychanalysis: Hypotheses toward an integrative model. *Frontiers in Psychology*, 12, 718372. <https://doi.org/10.3389/fpsyg.2021.718372>
- Mumma, G. H. (1998). Improving cognitive case formulations and treatment planning in clinical practice and research. *Journal of Cognitive Psychotherapy*, 12(3), 251-274.
- Mumma, G. H. (2001). Manual for the cognitive-behavioral-interpersonal semi-structured assessment interview. (CBISSAI). Unpublished manual.
- Mumma, G. H. (2011). Current issues in case formulation. In P. Sturmey & M. McMurrin (Eds.), *Forensic case formulation* (pp. 33-60). Wiley-Blackwell.
- Mumma, G. H., & Mooney, S. R. (2007). Comparing the validity of alternative cognitive case formulations: A latent variable, multivariate time series approach. *Cognitive Therapy and Research*, 31(4), 451-481. <https://psycnet.apa.org/doi/10.1007/s10608-006-9062-5>

- Mumma, G. H., & Smith, J. L. (2001). Cognitive-behavioral interpersonal scenarios: Inter-formulator reliability and convergent validity. *Journal of Psychopathology and Behavioral Assessment*, 23(4), 203-221.
<https://psycnet.apa.org/doi/10.1023/A:1012738802126>
- Muran, J. C., Samstag, L.W., Segal, Z. V., & Winston, A. (1998). Interpersonal scenarios: A measure of self-schemas. *Psychotherapy Research*, 8(3), 321-333.
<https://psycnet.apa.org/doi/10.1093/ptr/8.3.321>
- Muran, J. C., & Segal, V. (1992). The development of an idiographic measure of self-schemas: An illustration of the construction and use of self-scenarios. *Psychotherapy: Theory, Research, Practice, Training*, 29(4), 524-535. <https://doi.org/10.1037/0033-3204.29.4.524>
- Muran, J. C., Segal, Z. V., & Samstag, L. W. (1994). Self-scenarios as a repeated measures outcome measurement of self-schemas in short-term cognitive therapy. *Behavior Therapy*, 25(2), 255-274. [https://doi.org/10.1016/S0005-7894\(05\)80287-4](https://doi.org/10.1016/S0005-7894(05)80287-4)
- Murphy, M. K., Black, N. A., Lamping, D. L., McKee, C. M., Sanderson, C. F., Askham, J., & Marteau, T. (1998). Consensus development methods, and their use in clinical guideline development. *Health Technology Assessment*, 2(3), 1-88.
<https://doi.org/10.3310/hta2030>
- Muse, K., & McManus, F. (2016). Expert insight into the assessment of competence in cognitive-behavioural therapy: A qualitative exploration of experts' experiences, opinions and recommendations. *Clinical Psychology & Psychotherapy*, 23(3), 246-259.
<https://doi.org/10.1002/cpp.1952>
- Naeem, F., Farooq, S., & Kingdon, D. (2015). Cognitive behavioural therapy (brief versus standard duration) for schizophrenia. *Cochrane Database of Systematic Reviews*, 10, CD010646. <https://doi.org/10.1002/14651858.CD010646.pub3>
- Naeem, F., Houry, B., Munshi, T., Ayub, M., Lecomte, T., Kingdon, D., & Farooq, S. (2016). Brief cognitive behavioral therapy for psychosis (CBTp) for schizophrenia: Literature review and meta-analysis. *International Journal of Cognitive Therapy*, 9(1), 73-86.
https://doi.org/10.1521/ijct_2016_09_04
- Nathan, P. E., & Gorman, J. M. (2002). *A guide to treatments that work*. Oxford University Press.
- National Institute for Health and Clinical Excellence (NICE). (2014). *Psychosis and schizophrenia in adults: Prevention and management*. (Clinical Guideline 178).

<https://www.nice.org.uk/guidance/cg178/resources/psychosis-and-schizophrenia-in-adults-prevention-and-management-pdf-35109758952133>

- National Institute for Health and Care Excellence (NICE). (2020). *Implementing the early intervention in psychosis access and waiting time standard: Updated commissioning guidance*. NHS England Publications. http://iris-initiative.org.uk/wordpress/wp-content/uploads/2020/12/EIP-Guidance-2020_Final_2_KE.pdf
- Nattrass, A., Kellett, S., Hardy, G. E., & Ricketts, T. (2015). The content, quality and impact of cognitive behavioural case formulation during treatment of obsessive compulsive disorder. *Behavioural and Cognitive Psychotherapy*, 43(5), 590-601.
<https://doi.org/10.1017/S135246581400006X>
- Newman-Taylor, K., Wood, R., Ellis, A., & Isham, L. (2022). Formulating psychosis: A thematic analysis of CBTp trainees' experiences. *the Cognitive Behaviour Therapist*, 15, Article e54. <https://doi.org/10.1017/S1754470X22000459>
- Nezu, A. M., & Nezu, C. M. (1993). Identifying and selecting target problems for clinical interventions: A problem-solving model. *Psychological Assessment*, 5(3), 254-263. <https://doi.org/10.1037/1040-3590.5.3.254>
- NHS England. (2019). *The NHS long term plan*. NHS England Publications.
<https://www.longtermplan.nhs.uk/wp-content/uploads/2019/01/nhs-long-term-plan-june-2019.pdf>
- NHS England. (2023). *Implementing the early intervention in psychosis access and waiting time standard*. NHS England Publications. <https://www.england.nhs.uk/wp-content/uploads/2023/03/B1954-implementing-the-early-intervention-in-psychosis-access-and-waiting-time-standard.pdf>
- Niederberger, M., & Spranger, J. (2020). Delphi technique in health sciences: A map. *Frontiers in Public Health*, 8(457), 1-10. <https://doi.org/10.3389/fpubh.2020.00457>
- Nilsson, T., Svensson, M., Sandell, R., & Clinton, D. (2007). Patients' experiences of change in cognitive-behavioral therapy and psychodynamic therapy: A qualitative comparative study. *Psychotherapy Research*, 17(5), 553-566.
<https://doi.org/10.1080/10503300601139988>
- Nowicki, S., & Duke, M. P. (1974). A locus of control scale for noncollege as well as college adults. *Journal of Personality Assessment*, 38(2), 136-137.
<https://doi.org/10.1080/00223891.1974.10119950>

- O’Cathain, A., Thomas, K. J., Drabble, S. J., Rudolph, A., & Hewison, J. (2013). What can qualitative research do for randomised controlled trials? A systematic mapping review. *BMJ Open*, 3(6), e002889. <https://doi.org/10.1136/bmjopen-2013-002889>
- Office for National Statistics. (2022). *Dataset: Developing admin-based ethnicity statistics for England: 2016*. <https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/ethnicity/datasets/developingadminbasedethnicitystatisticsforengland>
- Owen, G. (2023). What is formulation in psychiatry? *Psychological Medicine*, 53(5), 1700-1707. <https://doi.org/10.1017/S0033291723000016>
- Padesky, C. A., Kuyken, W., & Dudley, R. (Eds.). (2011). *The collaborative case conceptualization rating scale (CCC-RS) and coding manual* (Version 5). <http://padesky.com/clinical-corner/clinical-tools>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... McKenzie, J. E. (2021). PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *BMJ*, 372, n160. <https://doi.org/10.1136/bmj.n160>
- Pain, C. M., Chadwick, P., & Abba, N. (2008). Clients’ experience of case formulation in cognitive behaviour therapy for psychosis. *British Journal of Clinical Psychology*, 47(2), 127-138. <https://doi.org/10.1348/014466507X235962>
- Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health*, 42(5), 533-544. <https://doi.org/10.1007/s10488-013-0528-y>
- Perkins, R., & Slade, M. (2012). Recovery in England: Transforming statutory services? *International Review of Psychiatry*, 24(1), 29-39. <https://doi.org/10.3109/09540261.2011.645025>

- Perry, A. (2012). CAT with people who hear distressing voices. *Reformulation, Summer*(38), 16-22.
- Perry, J. C., Augusto, F., & Cooper, S. H. (1989). Assessing psychodynamic conflicts: I. Reliability of the idiographic conflict formulation method. *Psychiatry*, 52(3), 289-301. <https://doi.org/10.1080/00332747.1989.11024451>
- Perry, S., Cooper, A. M., & Michels, R. (1987). The psychodynamic formulation: Its purpose, structure and clinical application. *American Journal of Psychiatry*, 144, 543- 550. <https://doi.org/10.1176/foc.4.2.297>
- Persons, J. B. (1989). *Cognitive therapy in practice: A case formulation approach*. Norton.
- Persons, J. B. (1989). *Cognitive therapy in practice: A case formulation approach*. Norton.
- Persons, J. B. (2005). Empiricism, mechanism, and the practice of cognitive-behavior therapy. *Behavior Therapy*, 36(2), 107-118. [https://doi.org/10.1016/S0005-7894\(05\)80059-0](https://doi.org/10.1016/S0005-7894(05)80059-0)
- Persons, J. B., & Bertagnolli, A. (1999). Inter-rater reliability of cognitive-behavioral case formulations of depression: A replication. *Cognitive Therapy and Research*, 23(3), 271-283. <https://doi.org/10.1023/A:1018791531158>
- Persons, J. B., Mooney, K. A., & Padesky, C. A. (1995). Interrater reliability of cognitive-behavioral case formulations. *Cognitive Therapy and Research*, 19(1), 21-34. <https://doi.org/10.1007/BF02229674>
- Persons, J. B., Roberts, N. A., Zalecki, C. A., & Brechwald, W. A. G. (2006). Naturalistic outcome of case formulation-driven cognitive-behavior therapy for anxious depressed outpatients. *Behaviour Research and Therapy*, 44, 1041-1051. <https://doi.org/10.1016/j.brat.2005.08.005>
- Peters, E., Ward, T., Jackson, M., Woodruff, P., Morgan, C., McGuire, P., & Garety, P. A. (2017). Clinical relevance of appraisals of persistent psychotic experiences in people with and without a need for care: An experimental study. *Lancet Psychiatry*, 4(12), 927-936. [https://doi.org/10.1016/S2215-0366\(17\)30409-1](https://doi.org/10.1016/S2215-0366(17)30409-1)
- Pilgrim, D. (2000). Psychiatric diagnosis: More questions than answers. *The Psychologist*, 13(6), 302-305.
- Pilgrim, D. (2020). *Critical realism for psychologists*. Routledge.
- Pollack, J., Flegenheimer, W., Kaufman, J., & Sadow, J. (1992). *Brief adaptive psychotherapy: A treatment manual*. Social and Behavioral Documents.

- Pope, C., & Mays, N. (1995). Reaching the parts other methods cannot reach: An introduction to qualitative methods in health and health services research. *BMJ*, 311(6996), 42-45. <https://doi.org/10.1136/bmj.311.6996.42>
- Popp, C. A., Diguer, L., Luborsky, L., Faude, J., Johnson, S., Morris, M., Schaffer, N., Schaffler, P., & Schmidt, K. (1996). Repetitive relationship themes in waking narratives and dreams. *Journal of Consulting and Clinical Psychology*, 64(5), 1073-1078. <https://doi.org/10.1037//0022-006x.64.5.1073>
- Potter, S. (2010). Words with arrows: The benefits of mapping whilst talking. *Reformulation, Summer*(34), 37-45.
- Powell, C. (2003). The Delphi technique: Myths and realities. *Journal of Advanced Nursing*, 41(4), 376-382. <https://doi.org/10.1046/j.1365-2648.2003.02537.x>
- Proglar, Y. (2009). Mental illness and social stigma: notes on "How Mad Are You?". *Journal of Research in Medical Sciences*, 14(5), 331-334.
- Psychotherapy Expert Talks. (2016a). *Tracy D. Eells on psychotherapy case formulation: An integrative, evidence-based model*. [Webinar]. YouTube. <https://www.youtube.com/watch?v=qTPkI6lpifA&list=WL&index=20&t=110s>
- Psychotherapy Expert Talks. (2016b). *Jacqueline Persons on CBT case formulation, Aaron T. Beck, and therapy research*. [Webinar]. YouTube. <https://www.youtube.com/watch?v=F505PAIFAGk>
- Pyle, M., Broome, M. R., Joyce, E., MacLennan, G., Norrie, J., Freeman, D., Fowler, D., Haddad, P. M., Shiers, D., Hollis, C., Smith, J., Liew, A., Byrne, R. E., French, P., Peters, S., Hudson, J., Davies, L., Emsley, R., Yung, A., ... Morrison, A. P. (2019). Study protocol for a randomised controlled trial of CBT vs antipsychotics vs both in 14-18-year-olds: Managing Adolescent first episode Psychosis: A feasibility study (MAPS). *Trials*, 20(395). <https://doi.org/10.1186/s13063-019-3506-1>
- Pyle, M., & Morrison, A. P. (2014). "It's just a very taboo and secretive kind of thing": Making sense of living with stigma and discrimination from accounts of people with psychosis. *Psychosis*, 6(3), 195-205. <https://doi.org/10.1080/17522439.2013.834458>
- QSR International Pty Ltd. (2018) NVivo (Version 12), <https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home>
- Quinlan, E. & Deane, F. P. (2021). A longitudinal study of trainee psychologists' tolerance of uncertainty, state anxiety and confidence in case formulation. *Australian Psychologist*, 56(6), 499-510. <https://doi.org/10.1080/00050067.2021.1965855>

- Quinlan, E., Deane, F. P., Schilder, S., & Read, E. (2022). Confidence in case formulation and pluralism as predictors of psychologists' tolerance of uncertainty. *Counselling Psychology Quarterly*, 35(4), 943-958. <https://doi.org/10.1080/09515070.2021.1997918>
- Rainforth, M., & Laurenson, M. (2014). A literature review of case formulation to inform mental health practice. *Journal of Psychiatric and Mental Health Nursing*, 21(3), 206-213. <https://doi.org/10.1111/jpm.12069>
- Ramus, F. (2013). What's the point of neuropsychanalysis? *The British Journal of psychiatry: The Journal of Mental Science*, 203(3), 170-171. <https://doi.org/10.1192/bjp.bp.113.127217>
- Ray, W. A., Chung, C. P., Murray, K. T., Hall, K., & Stein, C. M. (2009). Atypical antipsychotic drugs and the risk of sudden cardiac death. *The New England Journal of Medicine*, 360(3), 225-235. <https://doi.org/10.1056/NEJMoa0806994>
- Read, J., van Os, J., Morrison, A. P., & Ross, C. A. (2005). Childhood trauma, psychosis and schizophrenia: A literature review with theoretical and clinical implications. *Acta Psychiatrica Scandinavica*, 112(5), 330-350. <https://doi.org/10.1111/j.1600-0447.2005.00634.x>
- Rector, N. A., & Cassin, S. E. (2010). Clinical expertise in cognitive behavioural therapy: Definition and pathways to acquisition. *Journal of Contemporary Psychotherapy: On the Cutting Edge of Modern Developments in Psychotherapy*, 40(3), 153-161. <https://doi.org/10.1007/s10879-010-9136-2>
- Redhead, S., Johnstone, L., & Nightingale, J. (2015). Clients' experiences of formulation in cognitive behaviour therapy. *Psychology and Psychotherapy*, 88(4), 453-467. <https://doi.org/10.1111/papt.12054>
- Redmond, C., Larkin, M., Harrop, C. (2010). The personal meaning of romantic relationships for young people with psychosis. *Clinical Child Psychology and Psychiatry*, 15(2), 151-170. <https://doi.org/10.1177/1359104509341447>
- Rethlefsen, M. L., Kirtley, S., Waffenschmidt, S., Ayala, A. P., Moher, D., Page, M. J., Koffel, J. B., & PRISMA-S Group (2021). PRISMA-S: An extension to the PRISMA statement for reporting literature searches in systematic reviews. *Systematic Reviews*, 10(1), 39. <https://doi.org/10.1186/s13643-020-01542-z>
- Riso, L. P., & McBride, C. (2007). Introduction: A return to a focus on cognitive schemas. In L. P. Riso, P. L. du Toit, D. J. Stein & J. E. Young (Eds.), *Cognitive schemas and core beliefs in psychological problems: A scientist-practitioner guide* (pp. 3-9). American Psychological Association.

- Ritchie, J., & Spencer, L. (1994). Qualitative data analysis for applied policy research. In A. Bryman & R. G. Burgess (Eds.), *Analysing qualitative data* (pp. 173-194). Taylor & Francis.
- Roberts, G. (1991). Delusional belief systems and meaning in life: A preferred reality? *The British Journal of Psychiatry*, 159(Suppl 14), 19-28.
- Robins, L. N., Helzer, J. E., Croughan, J., & Ratcliff, K. S. (1981). National Institute of Mental Health Diagnostic Interview Schedule. Its history, characteristics, and validity. *Archives of General Psychiatry*, 38(4), 381-389.
<https://doi.org/10.1001/archpsyc.1981.01780290015001>
- Romme, M. A. J., & Escher, A. D. M. A. C. (1989). Hearing voices. *Schizophrenia Bulletin*, 15(2), 209-216. <https://doi.org/10.1093/schbul/15.2.209>
- Rosenberg, S. E, Silberschatz, G., Curtis, J. T., Sampson, H., & Weiss, J. (1986). A method for establishing reliability of statements from psychodynamic case formulations. *The American Journal of Psychiatry*, 143(11), 1454-1456. <https://doi.org/10.1176/ajp.143.11.1454>
- Rosenthal, J. (1987). *Judgment studies: Design analysis and meta-analysis*. Cambridge University Press.
- Roth, A., & Fonagy, P. (2005). *What works for whom?* Guilford Press.
- Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs: General and Applied*, 80(1), 1-28.
<https://doi.org/10.1037/h0092976>
- Routledge Mental Health's Podcast. (Producer). (2010, November 10). 'Collaborative Case Conceptualization', an interview with the authors [Audio podcast]. Podomatic.
https://www.podomatic.com/podcasts/routledgementalhealth-podcast/episodes/2010-11-10T02_42_06-08_00
- Royal College of Psychiatrists. (2014). *Report of the second round of the national audit of schizophrenia (NAS) 2014*. (No. CCQI183). Healthcare Quality Improvement Partnership. https://www.rcpsych.ac.uk/docs/default-source/improving-care/ccqi/national-clinical-audits/ncap-library/national-audit-of-schizophrenia-document-library/nas_round-2-report.pdf?sfvrsn=6356a4b0_4
- Royal College of Psychiatrists (2022). *Core psychiatry: Royal college of psychiatrists core training curriculum (CT1-CT3)* (Version 1.0). https://www.rcpsych.ac.uk/docs/default-source/training/curricula-and-guidance/2022-curricula/core-psychiatry-curriculum-final-17-august-20222f69f413dad4e07a91fc38aab87c931.pdf?sfvrsn=36b5ba25_8

- Ruffalo, M. L. (2020, July 19). *Psychoanalysis. Major psychodynamic paradigms: The basics. A brief overview of ego psychology, object relations, and self psychology.*
- Psychology Today. <https://www.psychologytoday.com/gb/blog/freud-fluoxetine/202007/major-psychodynamic-paradigms-the-basics>
- Ryle, A. (1985). Cognitive theory, object relations and the self. *British Journal of Medical Psychology*, 58(1), 1-7. <https://doi.org/10.1111/j.2044-8341.1985.tb02608.x>
- Ryle, A. (1994). Persuasion or education? The role of reformulation in cognitive analytic therapy. *International Journal of Short-Term Psychotherapy*, 9(2-3), 111-117.
- Ryle, A. (2003). History and use of the SDR. *Reformulation, Autumn*(20), 18-21.
- Ryle, A. (2005). Cognitive analytic therapy. In J. Norcross & M. Goldfried (Eds.), *Handbook of psychotherapy integration* (2nd ed., pp. 196-218). Oxford University Press. <https://doi.org/10.1093/med:psych/9780195165791.003.0009>
- Ryle, A., & Kerr, I. B. (2020). *Introducing cognitive analytic therapy: Principles and practice of a relational approach to mental health.* Wiley.
- Safran, J. D., & Muran, J. C. (1995). Resolving therapeutic alliance ruptures: Diversity and integration. *In Session: Psychotherapy in Practice*, 1(1), 81-92.
- Safran, J. D., & Segal, Z. V. (1990). *Interpersonal process in cognitive therapy.* Basic.
- Sage. (2021, November 17). *Thematic analysis: In conversation with Virginia Braun and Victoria Clarke* [Webinar]. YouTube. https://www.youtube.com/watch?v=hns-tlUx1_Q
- Sage. (2022, August 11). *Build good practice in reporting thematic analysis with Virginia Braun and Victoria Clarke* [Webinar]. YouTube. <https://www.youtube.com/watch?v=tpWLsckpM78>
- Sage. (2023, June 15). *Common challenges in thematic analysis and how to avoid them with Virginia Braun and Victoria Clarke* [Webinar]. YouTube. https://www.youtube.com/watch?v=En1HA_xwM4M&t=4s
- Schizophrenia Commission. (2012). *The abandoned illness: A report by the Schizophrenia Commission.* Rethink Mental Illness. https://www.rethink.org/media/2303/tsc_main_report_14_nov.pdf
- Schulze, K., Freeman, D., Green, C., & Kuipers, E. (2013). Intrusive mental imagery in patients with persecutory delusions. *Behaviour Research and Therapy*, 51(1), 7-14. <https://doi.org/10.1016/j.brat.2012.10.002>
- Scott, M. M., & Hatfield, J. C. (1985). Problems of analyst and observer agreement in naturalistic narrative data. *Journal of Educational Measurement*, 22(3), 207-218. <https://doi.org/10.1111/j.1745-3984.1985.tb01059.x>

- Scott, M., & Sembi, S. (2006). Cognitive Behaviour Therapy Treatment Failures in Practice: The Neglected Role of Diagnostic Inaccuracy. *Behavioural and Cognitive Psychotherapy*, 34(4), 491-495. <https://doi:10.1017/S1352465806003055>
- Seery, C., Bramham, J., & O'Connor, C. (2021). Effects of a psychiatric diagnosis vs a clinical formulation on lay attitudes to people with psychosis. *Psychosis*, 13(4), 361-372. <https://doi.org/10.1080/17522439.2021.1901302>
- Segal, Z. V. (1988). Appraisal of the self-schema construct in cognitive models of depression. *Psychological Bulletin*, 103(2), 147-162. <https://doi.org/10.1037/0033-2909.103.2.147>
- Segen's Medical Dictionary. (2012). *Bad habit*. Retrieved May 13, 2022, from <https://medical-dictionary.thefreedictionary.com/bad+habit>
- Seitz, P. F. (1966). The consensus problem in psychoanalytic research. In L. Gottschalk & L. Auerbach (Eds.), *Methods of research and psychotherapy* (pp. 209-225). Appleton-Century-Crofts.
- Selvini, M. P., Boscolo, L., Cecchin, G., & Prata, G. (1980). Hypothesizing-circularity-neutrality: Three guidelines for the conductor of the session. *Family Process*, 19(1), 3-12. <https://doi.org/10.1111/j.1545-5300.1980.00003.x>
- Sensky, T., Turkington, D., Kingdon, D., Scott, J. L., Scott, J., Siddle, R., O'Carroll, M., & Barnes, T. R. (2000). A randomized controlled trial of cognitive-behavioral therapy for persistent symptoms in schizophrenia resistant to medication. *Archives of General Psychiatry*, 57(2), 165-172. <https://doi.org/10.1001/archpsyc.57.2.165>
- Shah, V. (2018, December 1). *YourQuote*. <https://www.yourquote.in/vivek-shah-jjld/quotes/though-being-firm-believer-law-causality-i-still-have-events-kb1c0>
- Shakoor, S., McGuire, P., Cardno, A. G., Freeman, D., Plomin, R., & Ronald, A. (2015). A shared genetic propensity underlies experiences of bullying victimization in late childhood and self-rated paranoid thinking in adolescence. *Schizophrenia Bulletin*, 41(3), 754-763. <https://doi.org/10.1093/schbul/sbu142>
- Shalom, J. G., & Aderka, I. M. (2020). A meta-analysis of sudden gains in psychotherapy: Outcome and moderators. *Clinical Psychology Review*, 76, 101827. <https://doi.org/10.1016/j.cpr.2020.101827>
- Shamseer, L., Moher, D., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L. A., & PRISMA-P Group (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: Elaboration and explanation. *BMJ*, 350, g7647. <https://doi.org/10.1136/bmj.g7647>

- Sharkey, S. B., & Sharples, A. Y. (2001). An approach to consensus building using the Delphi technique: Developing a learning resource in mental health. *Nurse Education Today*, 21(5), 398-408. <https://doi.org/10.1054/nedt.2001.0573>
- Shefler, G., & Tishby, O. (1998). Interjudge reliability and agreement about the patient's central issue in time-limited psychotherapy (TLP) and its relation to TLP outcome. *Psychotherapy Research*, 8(4), 426-438. <https://doi.org/10.1080/10503309812331332507>
- Shine, L., & Westacott, M. (2010). Reformulation in cognitive analytic therapy: Effects on the working alliance and the client's perspective on change. *Psychology and Psychotherapy: Theory, Research and Practice*, 83(Pt 2), 161-177. <https://doi.org/10.1348/147608309X471334>
- Short, V., Covey, J. A., Webster, L. A., Wadman, R., Reilly, J., Hay-Gibson, N., & Stain, H. J. (2019). Considering the team in team formulation: A systematic review. *Mental Health Review Journal*, 24(1), 11-29. <https://doi.org/10.1108/MHRJ-12-2017-0055>
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420-428. <https://doi.org/10.1037//0033-2909.86.2.420>
- Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). McGraw-Hill.
- Sieradzka, D., Power, R. A., Freeman, D., Cardno, A. G., McGuire, P., Plomin, R., Meaburn, E. L., Dudbridge, F., & Ronald, A. (2014). Are genetic risk factors for psychosis also associated with dimension-specific psychotic experiences in adolescence?. *PLoS One*, 9(4), e94398. <https://doi.org/10.1371/journal.pone.0094398>
- Singh, T., Poterba, T., Curtis, D., Akil, H., Al Eissa, M., Barchas, J. D., Bass, N., Bigdeli, T. B., Breen, G., Bromet, E. J., Buckley, P. F., Bunney, W. E., Bybjerg-Grauholm, J., Byerley, W. F., Chapman, S. B., Chen, W. J., Churchhouse, C., Craddock, N., Cusick, C. M., ... Daly, M. J. (2022). Rare coding variants in ten genes confer substantial risk for schizophrenia. *Nature*, 604(7906), 509-516. <https://doi.org/10.1038/s41586-022-04556-w>
- Smith, B. (2018). Generalizability in qualitative research: Misunderstandings, opportunities and recommendations for the sport and exercise sciences. *Qualitative Research in Sport, Exercise and Health*, 10(1), 137-149. <https://doi.org/10.1080/2159676X.2017.1393221>
- Smith, J., Flowers, P., & Larkin, M. (2009). *Interpretative phenomenological analysis: Theory, method and research*. SAGE Publications Ltd.

- Smucny, J., Zarubin, V. C., Ragland, J. D., & Carter, C. S. (2020). Are visual memory deficits in recent-onset psychosis degenerative? *The American Journal of Psychiatry*, 177(4), 355-356. <https://doi.org/10.1176/appi.ajp.2019.19080845>
- Sørbye, Ø., Dahl, H. J., Eells, T. D., Amlo, S., Hersoug, A. G., Haukvik, U. K., Hartberg, C. B., Høglend, P. A., & Ulberg, R. (2019). Psychodynamic case formulations without technical language: A reliability study. *BMC Psychology*, 7(1), 67. <https://doi.org/10.1186/s40359-019-0337-5>
- Spencer, H. M.** (2019). Making sense of psychosis. In D. Turkington & **H. M. Spencer** (Eds.), *Back to life, back to normality: CBT informed recovery for families with relatives with schizophrenia and other psychoses* (volume 2, pp. 27-32). Cambridge University Press. <https://doi.org/10.1017/9781316410523.007>
- Spencer, H. M.**, Dudley, R., Freeston, M. H., & Turkington, D. (2020). What are the essential ingredients of a CBT case conceptualization for voices and delusions in schizophrenia spectrum disorders? A study of expert consensus. *Schizophrenia Research*, 224, 74-81. <https://doi.org/10.1016/j.schres.2020.09.026>
- Spencer, H. M.**, Dudley, R., Johnston, L., Freeston, M. H., Turkington, D., & Tully, S. (2023). Case formulation – A vehicle for change? Exploring the impact of Cognitive Behavioural Therapy formulation in first episode psychosis: A reflexive thematic analysis. *Psychology and Psychotherapy: Theory, Research and Practice*, 96(2), 328-346. <https://doi.org/10.1111/papt.12442>
- Spencer, H. M.**, McMenamin, M., Emsley, R., Turkington, D., Dunn, G., Morrison, A. P., Brabban, A., Hutton, P., & Dudley, R. (2018). Cognitive behavioural therapy for antipsychotic free schizophrenia spectrum disorders: Does therapy dose influence outcome? *Schizophrenia Research*, 202, 385-386. <https://doi.org/10.1016/j.schres.2018.07.016>
- Sperry, L., Gudeman, J. E., Blackwell, B., & Faulkner, L. R. (1992). *Psychiatric case formulations*. American Psychiatric Press.
- Startup, H., Pugh, K., Dunn, G., Cordwell, J., Mander, H., Černis, E., Wingham, G., Shirvell, K., Kingdon, D. & Freeman, D. (2016). Worry processes in patients with persecutory delusions. *British Journal of Clinical Psychology*, 55(4), 387-400. <https://doi.org/10.1111/bjc.12109>
- StataCorp. (2015). *Stata Statistical Software: Release 14*. College Station, TX: StataCorp LP.

- Steel, C., Tarrier, N., Stahl, D., & Wykes, T. (2012). Cognitive behaviour therapy for psychosis: The impact of therapist training and supervision. *Psychotherapy and Psychosomatics*, 81(3), 194-195. <https://doi.org/10.1159/000334250>
- Stemler, S. E., (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research, and Evaluation* 9(4), 1-11. <https://doi.org/10.7275/96jp-xz07>
- Stephenson, J. (2018). I, Me, You & Us: can a psychodynamic model provide a relational perspective on psychosis? *Psychoanalytic Psychotherapy*, 32(3), 267-281. <https://doi.org/10.1080/02668734.2018.1487464>
- Stern, A. (2017, May 20). Beyond Pluto: Nasa's next new horizon target is its biggest mystery yet. <https://www.wired.co.uk/article/alan-stern-new-horizons-pluto-mu69-nasa-interview>
- Stevens, L. H., **Spencer, H. M.**, & Turkington, D. (2017). Identifying four subgroups of trauma in psychosis: Vulnerability, psychopathology, and treatment. *Frontiers in Psychiatry*, 8(21). <https://doi.org/10.3389/fpsy.2017.00021>
- Stevens, L. H., Turkington, D., Drage, L., Morrison, T., Muncer, S., **Spencer, H. M.**, & Dudley, R. (2019). Investigation of a traumatic psychosis subgroup: A cluster analysis of an antipsychotic free cohort. *Psychosis*, 11(4), 298-307. <https://doi.org/10.1080/17522439.2019.1628290>
- Strunk, D. R., Brotman, M. A., DeRubeis, R. J., & Hollon, S. D. (2010). Therapist competence in cognitive therapy for depression: Predicting subsequent symptom change. *Journal of Consulting and Clinical Psychology*, 78(3), 429-437. <http://dx.doi.org/10.1037/a0019631>
- Strupp, H. H. (1955). Psychotherapeutic technique, professional affiliation, and experience level. *Journal of Consulting Psychology*, 19(2), 97-102. <https://doi.org/10.1037/h0046025>
- Strupp, H. H. (1958). The psychotherapist's contribution to the treatment process. *Behavioral Science*, 3(1), 34-67. <https://doi.org/10.1002/bs.3830030105>
- Sturgiss, E. A., & Clark, A. M. (2020). Using critical realism in primary care research: An overview of methods. *Family Practice*, 37(1), 143-145. <https://doi.org/10.1093/fampra/cmz084>
- Sullivan, S., Yamasaki, S., Ando, S., Endo, K., Kasai, K., Culpin, I., Dardani, C., Zammit, S., & Nishida, A. (2021). The association between locus of control and psychopathology: A cross-cohort comparison between a UK (Avon longitudinal study of parents and

- children) and a Japanese (Tokyo teen cohort) Cohort. *Frontiers in Psychology*, 12, 600941. <https://doi.org/10.3389/fpsyg.2021.600941>
- Summers, A., & Martindale, B. (2013). Using psychodynamic principles in formulation in everyday practice. *Advances in Psychiatric Treatment*, 19(3), 203-211. <https://doi.org/10.1192/apt.bp.112.010181>
- Sumsion, T. (1998). The Delphi technique: An adaptive research tool. *British Journal of Occupational Therapy*, 61(4), 153-156. <https://doi.org/10.1177/030802269806100403>
- Surís, A., Holliday, R., & North, C. S. (2016). The evolution of the classification of psychiatric disorders. *Behavioral Sciences*, 6(1), 5. <https://doi.org/10.3390/bs6010005>
- Sutton, J. (2022, November 26). *What is pathologizing & overpathologizing in psychology?* PositivePsychology.com. <https://positivepsychology.com/pathologizing/>
- Sweeney, A., Filson, B., Kennedy, A., Collinson, L., & Gillard, S. (2018). A paradigm shift: Relationships in trauma-informed mental health services. *BJPsych Advances*, 24(5), 319-333. <https://doi.org/10.1192/bja.2018.29>
- Tallberg, P., Ulberg, R., Johnsen Dahl, H. S., & Høglend, P. A. (2020). Core conflictual relationship theme: The reliability of a simplified scoring procedure. *BMC Psychiatry*, 20(1), 150. <https://doi.org/10.1186/s12888-020-02558-4>
- Tandon, R., Gaebel, W., Barch, D. M., Bustillo, J., Gur, R. E., Heckers, S., Malaspina, D., Owen, M. J., Schultz, S., Tsuang, M., Van Os, J., & Carpenter, W. (2013). Definition and description of schizophrenia in the DSM-5. *Schizophrenia Research*, 150(1), 3-10. <https://doi.org/10.1016/j.schres.2013.05.028>
- Tandon, R., Nasrallah, H. A., & Keshavan, M. S. (2009). Schizophrenia, "just the facts" 4. Clinical features and conceptualization. *Schizophrenia Research*, 110(1-3), 1-23. <https://doi.org/10.1016/j.schres.2009.03.005>
- Tapal, A., Oren, E., Dar, R., & Eitam, B. (2017). The sense of agency scale: A measure of consciously perceived control over one's mind, body, and the immediate environment. *Frontiers in Psychology*, 8(1552), 1-11. <https://doi.org/10.3389/fpsyg.2017.01552>
- Taplin, K., Greenhill, B., Seddon, C., & McGuire, J. (2018). Service user experiences of CAT diagrams: An interpretative phenomenological analysis. *International Journal of Cognitive Analytic Therapy and Relational Mental Health*, 2, 75-96.
- Tarrier, N. (2006a). An introduction to case formulation and its challenges. In N. Tarrier (Ed.), *Case formulation in cognitive behaviour therapy: The treatment of challenging and complex cases* (pp. 1-11). Routledge.

- Tarrier, N. (2006b). A cognitive-behavioural case formulation approach to the treatment of schizophrenia. In N. Tarrier (Ed.), *Case formulation in cognitive behaviour therapy: The treatment of challenging and complex cases* (pp. 167-187). Routledge.
- Tarrier, N., & Calam, R. (2002). New developments in cognitive-behavioural case formulation. Epidemiological, systemic and social context: An integrative approach. *Behavioural and Cognitive Psychotherapy*, 30(3), 311-328.
<https://doi.org/10.1017/S1352465802003065>
- Tarrier, N., & Johnson, J. (2016). Introduction. In N. Tarrier & J. Johnson (Eds.), *Case formulation in cognitive behaviour therapy: The treatment of challenging and complex cases* (2nd ed., pp. 1-13). Routledge.
- Taylor, P. J., Perry, A., Hutton, P., Tan, R., Fisher, N., Focone, C., Griffiths, D., & Seddon, C. (2019). Cognitive Analytic Therapy for psychosis: A case series. *Psychology and psychotherapy*, 92(3), 359-378. <https://doi.org/10.1111/papt.12183>
- Thangaratinam, S., & Redman, C. W. E. (2005). The Delphi technique. *The Obstetrician & Gynaecologist*, 7(2), 120-125. <https://doi.org/10.1576/toag.7.2.120.27071>
- Thew, G., & Krohnert, N. (2015). Formulation as intervention: Case report and client experience of formulating in therapy. *The Cognitive Behaviour Therapist*, 8, e25.
<http://doi.org/10.1017/S1754470X15000641>
- Thompson, M. (2009). Considering the implication of variations within Delphi research. *Family Practice*, 26(5), 420-424. <https://doi.org/10.1093/fampra/cmp05>
- Thrower, N. E., Bucci, S., Morris, L., & Berry, K. (2024). The key components of a clinical psychology formulation: A consensus study. *The British Journal of Clinical Psychology*, 1-14. Advance online publication. <https://doi.org/10.1111/bjc.12455>
- Thwaites, R., & Freeston, M. H. (2005). Safety-seeking behaviour: Fact or function? How can we clinically differentiate between safety behaviours and adaptive coping strategies across anxiety disorders? *Behavioural and Cognitive Psychotherapy*, 33(2), 177-188.
<http://doi.org/10.1017/S1352465804001985>
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society*, 58(1), 267-288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- Tien, A. Y. (1991). Distribution of hallucinations in the population. *Social Psychiatry and Psychiatric Epidemiology*, 26, 287-292. <https://doi.org/10.1007/BF00789221>
- Tong, J., Simpson, K., Alvarez-Jimenez, M., & Bendall, S. (2017). Distress, psychotic symptom exacerbation, and relief in reaction to talking about trauma in the context of beneficial trauma therapy: Perspectives from young people with post-traumatic stress

- disorder and first episode psychosis. *Behavioural and cognitive psychotherapy*, 45(6), 561-576. <https://doi.org/10.1017/S1352465817000236>
- Tong, J., Simpson, K., Alvarez-Jimenez, M., & Bendall, S. (2019). Talking about trauma in therapy: Perspectives from young people with post-traumatic stress symptoms and first episode psychosis. *Early intervention in psychiatry*, 13(5), 1236-1244. <https://doi.org/10.1111/eip.12761>
- Trauelsens, A. M., Bendall, S., Jansen, J. E., Nielsen, H. G., Pedersen, M. B., Trier, C. H., Haahr, U. H., & Simonsen, E. (2015). Childhood adversity specificity and dose-response effect in non-affective first-episode psychosis. *Schizophrenia Research*, 165(1), 52-59. <https://doi.org/10.1016/j.schres.2015.03.014>
- Trower, P., & Chadwick, P. (1995). Pathways to defense of the self: A theory of two types of paranoia. *Clinical Psychology Science and Practice*, 2(3), 263-278. <https://doi.org/10.1111/j.1468-2850.1995.tb00044.x>
- Trubetskoy, V., Pardiñas, A. F., Qi, T., Panagiotaropoulou, G., Awasthi, S., Bigdeli, T. B., Bryois, J., Chen, C. Y., Dennison, C. A., Hall, L. S., Lam, M., Watanabe, K., Frei, O., Ge, T., Harwood, J. C., Koopmans, F., Magnusson, S., Richards, A. L., Sidorenko, J., ... Schizophrenia Working Group of the Psychiatric Genomics Consortium (2022). Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature*, 604(7906), 502-508. <https://doi.org/10.1038/s41586-022-04434-5>
- Turkat, I. D. (2012). Case formulation — the heart of the matter. *The Psychologist*, 25, 728-729.
- Turkat, I. D. (2014). An historical perspective on the impact of case formulation. *the Behavior Therapist*, 37(7), 180-188.
- Turkat, I. D. (2015). Forward. In M. Bruch (Ed.). *Beyond diagnosis: Case formulation in cognitive behavioural therapy* (2nd ed., pp. ix-xi). [Foreword]. Wiley Blackwell.
- Turkat, I. D., & Maisto, S. A. (1983). Functions of and differences between psychiatric diagnosis and case formulation. *the Behavior Therapist*, 6, 184-185.
- Turkington, D., Bryant, C., & Lumley, V. (2011). Cognitive models for delusions. In R. Hagen, D. Turkington, T. Berge, & R. W. Gråwe (Eds.), *CBT for psychosis: A symptom-based approach* (pp. 22–39). Routledge.
- Turkington, D., Lebert, L., & **Spencer, H.** (2016). Auditory hallucinations in schizophrenia: Helping patients to develop effective coping strategies. *BJPsych Advances*, 22, 391-396. <http://doi.org/10.1192/apt.bp.115.015214>

- Turkington, D., & **Spencer, H. M.** (Eds.). (2019). *Back to life, back to normality: CBT informed recovery for families with relatives with schizophrenia and other psychoses*. (Volume 2). Cambridge University Press.
- Turkington, D., **Spencer, H.**, Jassal, I., & Cummings, A. (2015). Cognitive behavioural therapy for the treatment of delusional systems. *Psychosis*, 7(1), 48-59.
<https://doi.org/10.1080/17522439.2014.899385>
- Turkington, D., **Spencer, H.**, Lebert L., & Dudley, R. (2017). Befriending: Active placebo or effective psychotherapy? *The British Journal of Psychiatry*, 211(1), 5-6.
<https://doi.org/10.1192/bjp.bp.116.197467>
- Turner, D. T., Burger, S., Smit, F., Valmaggia, L. R., & van der Gaag, M. (2020). What constitutes sufficient evidence for case formulation-driven CBT for psychosis? Cumulative meta-analysis of the effect on hallucinations and delusions. *Schizophrenia Bulletin*, 46(5), 1072-1085. <https://doi.org/10.1093/schbul/sbaa045>
- Turner, D.T., van der Gaag, M., Karyotaki, E., & Cuijpers, P. (2014). Psychological interventions for psychosis: A meta-analysis of comparative outcome studies. *The American Journal of Psychiatry*, 171(5), 523-538.
<https://doi.org/10.1176/appi.ajp.2013.13081159>
- Tyler, N., Heffernan, R., & Fortune, C. A. (2020). Reorienting locus of control in individuals who have offended through strengths-based interventions: Personal agency and the good lives model. *Frontiers in Psychology*, 11, 553240.
<https://doi.org/10.3389/fpsyg.2020.553240>
- Tyrer, R., & Masterson, C. (2019). Clients' experience of change: An exploration of the influence of reformulation tools in cognitive analytic therapy. *Clinical Psychology & Psychotherapy*, 26(2), 167-174. <https://doi.org/10.1002/cpp.2339>
- Ullrich, S., Keers, R., Shaw, J., Doyle, M., & Coid, J. W. (2018). Acting on delusions: The role of negative affect in the pathway towards serious violence. *The Journal of Forensic Psychiatry and Psychology*, 29(5), 691-704.
<https://doi.org/10.1080/14789949.2018.1434227>
- Ushioda, E. (2009). A person-in-context relational view of emergent motivation, self and identity. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 215-228). Multilingual Matters.
- van der Gaag, M., Nieman, D., & van den Berg, D. (2013). *CBT for those at risk of a first episode psychosis: Evidence-based psychotherapy for people with an 'At Risk Mental State'*. Routledge Taylor & Francis Group.

- van Os, J., Hanssen, M., Bijl, R. V., & Ravelli, A. (2000). Strauss (1969) revisited: A psychosis continuum in the general population? *Schizophrenia Research*, 45(1-2), 11-20.
[https://doi.org/10.1016/s0920-9964\(99\)00224-8](https://doi.org/10.1016/s0920-9964(99)00224-8)
- Varese, F., Smeets, F., Drukker, M., Lieverse, R., Lataster, T., Viechtbauer, W., Read, J., van Os, J., & Bentall, R. P. (2012). Childhood adversities increase the risk of psychosis: A meta-analysis of patient-control, prospective- and cross-sectional cohort studies. *Schizophrenia Bulletin*, 38(4), 661-671. <https://doi.org/10.1093/schbul/sbs050>
- Völlm, B. (2014). Case formulation in personality disordered offenders - a Delphi survey of professionals. *Criminal Behaviour and Mental Health*, 24(1), 60-80.
<https://doi.org/10.1002/cbm.1884>
- Volpato, E., Cavallera, C., Castelnuovo, G., Molinari, E., & Pagnini, F. (2022). The "common" experience of voice-hearing and its relationship with shame and guilt: A systematic review. *BMC Psychiatry*, 22(1), 281. <https://doi.org/10.1186/s12888-022-03902-6>
- Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Waite, F., Myers, E., Harvey, A. G., Espie, C. A., Startup, H., Sheaves, B., & Freeman, D. (2016). Treating sleep problems in patients with schizophrenia. *Behavioural and Cognitive Psychotherapy*, 44(3), 273-287. <https://doi.org/10.1017/S1352465815000430>
- Waller, H., Emsley, R., Freeman, D., Bebbington, P., Dunn, G., Fowler, D., Hardy, A., Kuipers, E., & Garety, P. (2015). Thinking Well: A randomised controlled feasibility study of a new CBT therapy targeting reasoning biases in people with distressing persecutory delusional beliefs. *Journal of Behavior Therapy and Experimental Psychiatry*, 48, 82-89. <https://doi.org/10.1016/j.jbtep.2015.02.007>
- Wallerstein, R. S. (2002). The trajectory of psychoanalysis: A prognostication. *The International Journal of Psychoanalysis*, 83(6), 1247-1267.
<https://doi.org/10.1516/002075702320888620>
- Walsh, E., Buchanan, A., & Fahy, T. (2002). Violence and schizophrenia: Examining the evidence. *British Journal of Psychiatry*, 180(6), 490-495.
<https://doi.org/10.1192/bjp.180.6.490>
- Watson, J. (Ed.). (2019). *Drop the disorder! Challenging the culture of psychiatric diagnosis*. PCCS Books.
- Watson, D., Clark, L. A., Weber, K., Assenheimer, J. S., Strauss, M. E., & McCormick, R. A. (1995). Testing a tripartite model: II. Exploring the symptom structure of anxiety and

- depression in student, adult, and patient samples. *Journal of Abnormal Psychology*, 104(1), 15-25. <https://doi.org/10.1037//0021-843x.104.1.15>
- Wells, G. A., Shea, B., O'Connell, D., Peterson, J., Welch, V., Losos, M., & Tugwell, P. (2010). *The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomized studies in metaanalyses*. http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp
- Westermann, S., & Lincoln, T. M. (2011). Emotion regulation difficulties are relevant to persecutory ideation. *Psychology and Psychotherapy*, 84(3), 273-287. <https://doi.org/10.1348/147608310X523019>
- Wheable, V., & Davis, J. (2020). Examining the evidence base for forensic case formulation: An integrative review of recent research. *International Journal of Forensic Mental Health*, 19(3), 315-328. <https://doi.org/10.1080/14999013.2019.1707331>
- Wiggins, S. (2016). *Discursive psychology: Theory, method and applications*. SAGE Publications Ltd.
- Willig, C. (2013). *Introducing qualitative research in psychology* (3rd ed.). Open University Press.
- Wilson, F. E., & Evans, I. A. (1983). The reliability of target-behavior selection in behavioral assessment. *Behavioral Assessment*, 5, 15-32.
- Wilson, R. S., Yung, A. R., & Morrison, A. P. (2020). Comorbidity rates of depression and anxiety in first episode psychosis: A systematic review and meta-analysis. *Schizophrenia Research*, 216, 322-329. <https://doi.org/10.1016/j.schres.2019.11.035>
- World Health Organization. (1992). *The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines* (10th revision). World Health Organization. https://cdn.who.int/media/docs/default-source/classification/other-classifications/9241544228_eng.pdf?sfvrsn=933a13d3_1&download=true
- World Health Organization. (2019). *ICD-11: International classification of diseases* (11th revision). World Health Organization. <https://icd.who.int/>
- Wright, N. P., Turkington, D., Kelly, O. P., Davies, D., Jacobs, A. M., & Hopton, J. (2014). *Treating psychosis: A clinician's guide to integrating acceptance and commitment therapy, compassion-focused therapy and mindfulness approaches within the cognitive-behavioral therapy tradition*. New Harbinger Publications.
- Wykes, T., Steel, C., Everitt, B., & Tarrier, N. (2008). Cognitive behavioural therapy for schizophrenia: Effects sizes, clinical models and methodological rigour. *Schizophrenia Bulletin*, 34(3), 523-537. <https://doi.org/10.1093/schbul/sbm114>

- Yung, A. R., Wood, S. J., Malla, A., Nelson, B., McGorry, P., & Shah, J. (2021). The reality of at risk mental state services: A response to recent criticisms. *Psychological Medicine*, 51(2), 212-218. <https://doi.org/10.1017/S003329171900299X>
- Yung, A. R., Yuen, H. P., McGorry, P. D., Phillips, L. J., Kelly, D., Dell'Olio, M., Francey, S. M., Cosgrave, E. M., Killackey, E., Stanford, C., Godfrey, K., & Buckby, J. (2005). Mapping the onset of psychosis: The comprehensive assessment of at-risk mental states. *The Australian and New Zealand Journal of Psychiatry*, 39(11-12), 964-971. <https://doi.org/10.1080/j.1440-1614.2005.01714.x>
- Zavos, H. M., Freeman, D., Haworth, C. M., McGuire, P., Plomin, R., Cardno, A. G., & Ronald, A. (2014). Consistent etiology of severe, frequent psychotic experiences and milder, less frequent manifestations: A twin study of specific psychotic experiences in adolescence. *JAMA Psychiatry*, 71(9), 1049-1057. <https://doi.org/10.1001/jamapsychiatry.2014.994>
- Zivov, M., Salkovskis, P. M., Oldfield, V. B., & Kushnir, J. (2013). Formulation in cognitive behavior therapy for obsessive–compulsive disorder: Aligning therapists, perceptions and practice. *Clinical Psychology: Science and Practice*, 20(2), 143-151. <https://doi.org/10.1111/cpsp.12030>