Meaning and individual minds: the case of *if*

Magdalena Sztencel
Doctor of Philosophy

School of English Literature, Language and Linguistics/
Centre for Research in Linguistics and Language Sciences
Newcastle University

June, 2013

Abstract

Traditionally (e.g. Sperber & Wilson 1995, Levinson 2000, Jackendoff 2002, Chomsky 2005a), linguistic expressions have meaning in virtue of having linguistic semantic properties. It is often claimed that linguistic semantics is functionally distinct from but related to the semantics of thought. In particular, linguistic semantics is assumed to be deterministically (necessarily and always) decoded in utterance interpretation and fed, as a basic premise, to pragmatic processing. Linguistic semantics is supposed to aid (i.e. constrain) utterance interpretation insofar as it is at least 'widely' shared among speech community members (Carston 2002). However, it has been suggested that linguistic semantics is problematic (e.g. Burton-Roberts 2005, Gibbs 2002, Recanati 2005).

This thesis argues that the notion of linguistic semantics, as well as the process of deterministic decoding of such content, is implausible and explores the consequences of this claim for a theory of meaning and utterance interpretation.

In the first part, I raise questions about the nature of semantics (externalism or internalism) as well as its structure (atomism, molecularism or holism). In line with the Representational Hypothesis (e.g. Burton-Roberts 2012), I maintain that thought is the only locus of semantics and that meaning is not a property of linguistic expressions, but a cognitive relation between an uttered word and semantics (of thought). I argue that whereas semantic content is holistic, meaning (in the sense of Burton-Roberts) is locally – i.e. contextually – constrained to a degree which, all things being equal, allows for successful communication. I argue that utterance interpretation is a wholly pragmatic inferential process, immediately constrained by a personal (i.e. holistic) inference about the communicative intention of a particular speaker in a particular conversational context. I claim that such a process of utterance interpretation can be implemented in terms of Hintzman's (1986) multiple-trace theory of memory.

In the second part, I illustrate my argument by an analysis of the relation between the word *if* and Material Implication (MI). I show that the claim (e.g. Grice 1989, Noh 2000) that *if* semantically encodes MI cannot be maintained. I argue that the application of MI has to be pragmatically determined and, therefore, when MI applies, it does so at the level of (holistic) thought – not at the (anyway problematic) linguistic semantic level. I explain the interpretation of conditionals in terms of Horton & Gerrig's (2005) extension of a multiple-trace theory of memory into the study of common ground. I also discuss the implications of a wholly pragmatic account of utterance interpretation for the distinction between explicit and implicit communication.

To the Memory of my Grandparents

Krystyna and Franciszek

Acknowledgements

First and foremost, I would like to thank my supervisor Noel Burton-Roberts. This thesis would have been simply impossible without his expert guidance. I would like to thank him for his teaching, discussions, detailed and always incredibly helpful comments on the dozens of essays, emails, notes and drafts that I have inundated him with over the past few years. Needless to say, on countless occasions, his encouragement, patience (lots of it) and a great sense of humour were invaluable.

I would also like to express my gratitude for the PhD bursary I received from the School of English Literature, Language and Linguistics and for conference grants offered by the School and by the Centre for Research in Linguistics and Language Sciences. Thanks to this financial support, I was able to spend more time on research and have greatly benefited from attending and presenting at several conferences around Europe. I would also like to thank my examiners, Siobhan Chapman and Richard Waltereit, for many extremely helpful comments and suggestions.

Completing this PhD would have been impossible without the unwavering support from my husband Piotr, who on many evenings patiently put up with my musings instead of listening to his CDs. Completing this PhD would have also been much more difficult without the cheerful presence of my lovely children, who continuously help me put things into perspective. I would also like to thank my mum and my in-laws for their help in times of intensive writing.

Many thanks also go to friends who have devoted their precious time to proof-reading parts of my thesis: Lucy Averley, Louise Carter, Bella Reichard and Mais Sulaiman.

Table of contents

Abbreviations

A-P	Articulatory-Perceptual
C-I	Conceptual-Intentional
C/IRS	Conceptual/Inferential Role Semantics
CSPR	Conventional System for the Physical Representation of thought
F	False
(F)AC	(Fallacy of) Affirming the Consequent
(F)DA	(Fallacy of) Denying the Antecedent
FL	Faculty of Language
GCI	Generalised Conversational Implicature
Iff	If and only if
LF	Logical form
LOT	Language of Thought
MI	Material implication
MP	Modus Ponens
MT	Modus Tollens
PCI	Particularised Conversational Implicature
PF	Phonetic form
PHON	Phonological properties
PM	Primary memory
RH	The Representational Hypothesis
RT	Relevance Theory
SEM	Semantic properties
SM	Secondary memory
T	True
UG	Universal Grammar

Symbols

⊃	material implication
&	conjunction
v	disjunction
~	negation
≡	equivalence
⊢	entailment
∀	universal quantifier
∃	existential quantifier
⊆	inclusion (‘is a subset of’)

List of figures

Introduction

Twenty men crossing a bridge,
Into a village,
Are twenty men crossing twenty bridges,
Into twenty villages,
Or one man
Crossing a single bridge into a village.

This is old song
That will not declare itself . . .

Twenty men crossing a bridge,
Into a village,
Are
Twenty men crossing a bridge
Into a village.

That will not declare itself
Yet is certain as meaning . . .

The boots of the men clump
On the boards of the bridge.
The first white wall of the village
Rises through fruit-trees.
Of what was it I was thinking?
So the meaning escapes.

The first white wall of the village...
The fruit-trees...

(Wallace Stevens 'Metaphors of a Magnifico')

Traditionally (e.g. Sperber & Wilson 1995, Levinson 2000, Jackendoff 2002, Chomsky 2005a), linguistic expressions have meaning in virtue of having specifically linguistic semantic properties. It is often claimed that linguistic semantics is functionally distinct from but related to the semantics of thought. In particular, linguistic semantics is assumed to be deterministically (necessarily and always) decoded/accessed in utterance interpretation and fed, as a basic premise, to pragmatic processing. Linguistic semantics is supposed to aid (i.e. constrain) utterance interpretation insofar as it is at least 'widely' shared among speech community members (Carston 2002).

In this thesis, I argue that the notion of linguistic semantics, as well as the process of deterministic decoding of such content, is implausible and explore the consequences of this claim for a theory of meaning and utterance interpretation. The arguments put forward in this thesis apply to any theory of meaning which assumes some notion of linguistic semantics – in principle, any theory which places itself within the Saussurean tradition. However, I will focus my attention on Chomsky's theory and its extension (as argued by in Sperber & Wilson 1995, Carston 2002) in Relevance Theory. This is because the theoretical tension I am interested in – i.e. the tension

between the assumption that there is linguistic semantics, on the one hand, and individualism about mental content, on the other – is clearly discernible in both Chomsky (e.g. 2000a) and Relevance Theory (e.g. Sperber & Wilson 1995, Carston 2002).

As for individualism, Chomsky (2000a: 37) argues that there is nothing in the mind-external world such that it could correspond to 'the properties of the intricate modes of reference that a […] name encapsulates'. What we ever talk about is the world as seen through internalist and individualistic perspectives made available by our minds. Indeed, Chomsky (e.g. 2000a: 137) argues that the study of natural language meaning/semantics should be tantamount to the study of (individualistic) beliefs mediating between cognisers and things in the world. In other words, for Chomsky, natural language meaning/semantics is radically individualistic.

However, Chomsky's radical individualism is in tension with his double-interface view of language, itself a legacy of Saussure's notion of the linguistic sign. For Saussure and for Chomsky, the linguistic sign is partly constituted by what it is a sign of – i.e. a signifier, a concept (Burton-Roberts & Poole 2006b). In this tradition, which I refer to as the Saussurean-Chomskyan tradition (after Burton-Roberts 2007), linguistic expressions are double-interface objects which have sound properties and meaning properties as their constitutive parts. In Chomsky's tradition in particular, linguistic expressions are <PHON, SEM> pairs, i.e. double-interface syntactic objects constituted by phonological and semantic properties (Burton-Roberts 2011). Following from this double-interface idea is the view that words, and linguistic expressions in general, have meaning in virtue of constitutively encoding specifically linguistic, context-invariant and stable semantic properties (i.e. linguistic semantics). Whereas Chomsky's radical individualism about natural language meaning/semantics suggests that there is no linguistic semantics, it is precisely his double-interface (<PHON, SEM>) view of linguistic expressions which invokes this notion.

It is usually claimed that linguistic semantics is distinct from but 'related' to the semantics of thought (e.g. Chomsky 2005a, Carston 2002). Linguistic semantics is assumed to be deterministically decoded by the language module and fed, as a basic premise, to pragmatic processes of utterance interpretation. Linguistic semantics is supposed to aid (i.e. constrain) utterance interpretation insofar as it is at least 'widely' shared among members of a given speech community (Carston 2002). However, it has been suggested that linguistic semantics is a problematic notion (e.g. Burton-Roberts 2005, Gibbs 2002, Recanati 2005).

The thesis is divided into two parts. Part I presents theoretical arguments against the notion of linguistic semantics and for a radically pragmatic (holistic) model of utterance interpretation. It consists of three chapters: chapter 1 ('The externalist-internalist debate'), chapter 2 ('The locus of semantics and the decoding-inferring distinction') and chapter 3 ('Unleashing holism'). Part II looks at whether the arguments made in Part I can and should be applied to linguistic analysis – in particular, to the analysis of the meaning of the word *if*. Part II consists of two chapters: chapter 4 ('Does *if* encode Material Implication?') and chapter 5 ('Holistic and individualistic conditions on interpretation').

In chapter 1, I critically engage with Fodor's (e.g. 2008) externalist position on the nature of semantics and seek to justify the adoption of internalism (e.g. Chomsky 2000a). I argue that Fodor's externalism about semantic content is untenable as it presupposes internalist content in the guise of mind-dependence thesis. Relatedly, I argue that Fodor's lexical-conceptual isomorphism, which is the assumption that 'atomic' concepts equal word meanings, is problematic in the light of cross-linguistic evidence. I then identify four aspects of Chomsky's internalism that I endorse: (a) innateness of concepts, (b) internal compositionality of a word's semantics, (c) its context-variability and (d) its radically individualistic nature. However, I argue that the tension between the consequences of Chomsky's internalist-*individualistic* assumptions (from which it follows that there *is no* linguistic semantics) and his double-interface view of language (from which it follows that there *is* linguistic semantics) needs to be resolved.

Chapter 2 argues for the resolution of the tension between internalism individualism, on the one hand, and the assumption that linguistic semantics is necessary to account for meaning in language, on the other, in favour of internalism-individualism. I seek to show that linguistic semantics is an implausible and unnecessary notion. In particular, I examine the notion of linguistic semantics (and lexical concept) in Relevance Theory, a theory argued to be an extension of Chomsky's paradigm (Sperber & Wilson 1995, Carston 2002). After discussing some problems with the nature and acquisition of linguistic semantics, I argue that the posited process of deterministic decoding of such content is redundant in cases of loose use, cases of so-called concept narrowing and where the communicated concept is the same as the assumed lexical concept. I argue that there is no linguistic semantics and that utterance interpretation is a wholly pragmatic inferential process, immediately constrained by a personal (i.e. holistic) inference about the communicative intention of the speaker in a

given conversational context. I defend my view by dismissing two potential criticisms: (a) If words do not have meaning in virtue of encoding linguistic semantics, how do they mean?; (b) If there is no linguistic semantics constraining word use, how do we ever communicate successfully? I argue that the Representational Hypothesis' (e.g. Burton-Roberts 2012) definition of *meaning-as-relation* invalidates criticism (a) and that Hintzman's (1986) multiple-trace theory of memory and information retrieval invalidates (b).

Chapter 3 looks at the philosophical legacy in thinking about linguistic semantics (in particular, in Relevance Theory's notion of linguistic semantics). I argue that all philosophical notions of shared content which have been posited as linguistic semantics (i.e. causal-externalist wide content, social-externalist wide content and non-truth-theoretic narrow content) are problematic and that holism is the only plausible thesis about mental content. I defend holism against three criticisms made by its opponents (e.g. Fodor & Lepore 1992) – that holism does not allow for mental generalisations (and thus for a theory of *the* human mind), that holism does not offer an accurate account of compositionality, and that it gives rise to the sense-reference problem. I endorse Bilgrami's (1992) holistic thesis about the unity and locality of content and discuss how, in the light of Bilgrami's thesis and the Representational Hypothesis (e.g. Burton-Roberts 2012), we can and should distinguish between a domain of concepts and a domain of associations between semiotic labels and concepts. Finally, I argue that whereas semantic content is holistic, meaning (in the sense of Burton-Roberts) is locally – i.e. contextually – constrained to a degree which, all things being equal, allows for successful communication.

In part II, I illustrate my argument by an analysis of the relation between the word *if* and the logical functor of material implication (MI). The reason for looking at this particular word is its significance to the distinction between deterministically decoded linguistic semantics and pragmatically inferred semantics of thought. While the claim that *if* semantically encodes MI has been controversial in philosophy, pragmatic explanations of the deviations in the interpretation of conditionals from the supposedly encoded MI (e.g. Grice 1989) may be thought of as one of the most successful achievements of pragmatic theory.

However, in chapter 4, I argue that the claim (e.g. Grice 1989, Noh 2000) that *if* semantically encodes MI cannot be maintained even when supported by pragmatic explanation. I show that the problem of pragmatic intrusion into encoded semantics arises for Relevance Theory (Carston 2002, Noh 2000), as it did for Grice (1989). I

argue that the application of MI has to be pragmatically determined and, therefore, when MI applies, it does so at the level of (holistic) thought – not at the (anyway problematic) linguistic semantic level.

In chapter 5, I show how the interpretation of conditionals can be explained in terms of a wholly pragmatic inferential process. This interpretational process is cashed out in terms of Horton & Gerrig's (2005) extension of a multiple-trace theory of memory (e.g. Hintzman 1986) into the study of conversational common ground. I argue that once Horton & Gerrig's argument – that interlocutor-specific information places immediate constraints on utterance interpretation – is acknowledged, it is possible to explain the distinction between the weak (i.e. modelled by MI) and the strong (i.e. modelled by equivalence) interpretations of conditionals. I distinguish between basic uses of conditionals (where *if* signals a relation between two propositional objects *p* and *q*) and extended uses of conditionals (where *if* signals a relation between a proposition and an utterance (or speech act)). I explain why the wholly pragmatic (holistic) analysis I propose is more adequate to handle variation found in the interpretation of basic and extended uses than previous approaches (e.g. Noh (2000), Sweetser (1990), Smith & Smith (1988)). Finally, I discuss the implications of a wholly pragmatic approach to utterance interpretation for the distinction between explicit and implicit communication.

PART I

Chapter 1. The externalist-internalist debate

1.0 Introduction

One of the fundamental questions of semantic theory is how meaning in language arises. Crucial to this question is the distinction between externalism and internalism about mental/semantic content: for internalists (e.g. Chomsky 2000a, 2003), semantic content is to be defined in terms of purely mind-internal states and relations, whereas for externalists (e.g. Fodor 1998, 2008), semantic content is to be defined in terms of the relation between the mind and the mind-external reality. Differences aside, a general assumption shared by both externalist and internalist accounts of semantic content is that meaning in language arises in virtue of (i) the semantics of thought, and (ii) some relation between words and concepts, i.e. constituents of thought. Regarding (i), this chapter critically engages with Fodor's (e.g. 2008) externalist position on the nature of semantics and, in doing so, seeks to justify internalism (e.g. Chomsky 2000a). As for (ii), the chapter provides arguments against Fodor's (*ibid.*) lexical-conceptual isomorphism (i.e. conceptual atomism) and in favour of a compositional account of conceptual correlates of words (i.e. concepts which correspond to words in a language).

In the first section, I look at Fodor's (1998, 2008) notion of semantic content. I argue that Fodor's referentialism about semantic content is untenable for two reasons. First, it presupposes internalist content in the guise of mind-dependent properties that our minds attribute to mind-external entities. Furthermore, it is these properties that compose and not, as Fodor argues, reference. Relatedly, Fodor's 'lexical-conceptual isomorphism' (i.e. the assumption that 'atomic' concepts just are word meanings) raises problems. First, it incorrectly predicts one-to-one mapping between words and concepts. Second, referential equivalents like English *shallow* and French *peu profond* indicate that at least some 'atomic' concepts have compositional content. The isomorphism is also incompatible with Fodor's view that thought is not necessarily language-dependent. In the context of the externalist-internalist debate about the nature of semantic content, the problems with Fodor's referentialism strongly tip the balance in favour of internalism – the subject of section 1.2.

In the second section, I provide an overview of Chomsky's internalism. This includes a brief introduction to the architecture of Chomsky's Faculty of Language and, following from that, his views on the nature of semantic content. Here I begin to

identify some tensions in Chomsky's thinking; whereas for Chomsky concepts are innate and invariant across the species, natural language (e.g. English, Polish, Swahili) semantics, which is 'related to' or perhaps 'identical with' innate conceptual resources, is variable across contexts and individuals and hence not subject to Chomsky's principled inquiry into language. Finally, I discuss a tension between Chomsky's views on natural language semantics and a view of natural language semantics which follows from his double-interface legacy. This is followed by a conclusion.

1.1 Fodor's externalism and the Language of Thought hypothesis

On the Language of Thought (LOT) hypothesis, thinking is computation, i.e. it is a causal chain of operations on mental symbols or concepts. Fodor's LOT hypothesis specifically is a version of a Representational Theory of Mind, a theory that aims to explain how minds represent things. The Representational Theory of Mind, in other words, is a theory concerned with how it is possible that we think about what we think about – that is, often things in the mind-external world.

Before I discuss Fodor's account, let me briefly point out that there are several important assumptions underlying his referentialism. The first of these is that conceptual correlates of words are atomic; that is, they are not compositionally constituted.[1] Fodor's atomicity follows straightforwardly from his rejection of the analytic-synthetic distinction (Fodor 1998: 70-72, 2007: 3). But consider (1) and (2) below (from Rey 2010).

(1) People who run move their bodies.
(2) People who run damage their bodies.

By definition, analytic sentences are those whose truth value is known solely in virtue of knowing the meaning/semantics of the expressions they are composed by. For example, because the verb *run* is definable in terms of the verb *move* (i.e. *run* entails *move*)[2], (1) is known to be true by anyone who knows the meaning/semantics of the expression *run*. In contrast, the truth value of a synthetic sentence is not known simply by knowing the meaning/semantics of expressions it is composed by, but requires reference to the mind-external world; in order to know the truth value of a synthetic

[1] For Fodor, an atomic concept is just any concept that corresponds to some word, even if it might intuitively seem that the meaning of the relevant word is compositional.
[2] Later in this chapter, I argue that entailment does not actually hold between words but between concepts.

sentence, we need to go beyond logic and consult our so-called encyclopaedic knowledge (we need to know what the world is like). For example, the truth value of (2) cannot be determined by looking at the meanings/semantics of *run* and *damage* because there is no relation of entailment between these two expressions. Put differently, running is not definable in terms of damaging – running may but does not have to damage one's body. The idea behind the analytic-synthetic distinction is that the meanings/semantics of expressions can be defined compositionally in terms of the expression's logical, but not synthetic, relations with other expressions. Such an approach to meaning/semantics has been termed MEANING MOLECULARISM (e.g. Lepore 1999, Pagin 2006).

An example often rehearsed in the context of molecularist approaches to word meaning/semantics is the word *bachelor*. The molecularist assumption is that the word *bachelor* can be defined in terms of *adult*, *male*, *single* and *human*. This is to say that the meaning/semantics of the word *bachelor* is defined in terms of (i.e. is compositionally constituted by/contains) the concepts ADULT, MALE, SINGLE and HUMAN or that the word *bachelor* semantically licences inferences to these concepts and no others.[3]

Despite being quite intuitive, the analytic-synthetic distinction has come under severe criticism. For example, Fodor (1998: 107) concedes that analytic relations, if there were any, could in principle provide a satisfactory account of word meaning/conceptual content. He argues, however, that no version of molecularism can be right. Fodor agrees with Quine (1951) that since no one has been able to draw a serious distinction between conceptual connections that are analytic (i.e. those which enter into content-constitutive relation with a concept) and those that are not, there is no such distinction.[4] In definitional terms, this means that word meanings/semantics cannot be construed as definitions which we all share; in inferential terms, this means that there are no inferences which a word necessarily and sufficiently gives rise to. Furthermore, Fodor's rejection of the analytic-synthetic distinction means that any compositional account of the semantics of conceptual correlates of words will inevitably be holistic, i.e. psychological and individualistic. This would be in conflict with Fodor's second assumption.

[3] Laurence & Margolis (1999) make a distinction between containment and inferential molecular accounts. I assume that the fact that a word licences particular inferences arises *because* it is used to represent particular conceptual *structures*, i.e. compositionally complex concepts. Inferential relations and conceptual containment are thus two sides of the same coin.

[4] Lahav (1989) provides arguments against molecularism by discussing adjective-noun expressions.

The second assumption is actually a constraint that Fodor sets for his theory. That is, Fodor holds that a theory of concepts should be able to account for content identity amongst the species. In Fodor's (e.g. 1998: 28, 34) terms, his theory should meet the publicity constraint on concepts, i.e. a constraint by which a theory of concepts should predict that we all share types of the same primitive concepts. This is necessary because, for Fodor, only content identity can account for intentional generalisations across individuals and across slices of time in the same individual. Such intentional generalisations are, in turn, assumed to be necessary to explain rational behaviour.[5]

The third of Fodor's assumptions is that we need referential semantics to account for the fact that we often think and therefore talk about the mind-external world: '*What minds do is think about things*' (Fodor 2008: 8, italics original). This assumption is grounded in Fodor's (Fodor 2008: 16 fn 29) rejection of solipsism[6]. Solipsism is wrong, Fodor argues, because (among other things) it seems to deny that at least some of the time we do think about the world (e.g. 2007: 7). Fodor's referential semantics, which is a theory of co-variances between symbols in the mind and things in the world, is supposed to connect concepts and thus words directly to the world. It will shortly become clear that whereas I share with Fodor the intuition that very often we think and talk about the mind-external word, I do not agree that this intuition can be explained in terms of referential semantics.

As mentioned, Fodor is concerned with the question of how it is possible to think about the mind-external world. The two big questions that Fodor (1998, 2008) expects his theory to answer are: (a) in what sense do concepts hook up/refer to things out there in the world?, and (b) in virtue of what do atomic concepts compose to form complex concepts? In simple terms, Fodor asks what it is about the content of a concept like CAT that it hooks up to cats in the mind-external world, and what it is about the concept CAT that it can combine with another atomic concept like BLACK to form the complex concept BLACK CAT.

Before I engage with Fodor's recent answers to these questions, let me briefly consider one issue surrounding the development of Fodorian thinking on the ontology of concepts.

---

[5] I return to this point in chapter 3.

[6] There are two kinds of solipsism: methodological and ontological. Methodological solipsism holds that mental states and processes are insensitive to the mind-external environment and thus to properties such as truth and reference (e.g. Fodor 1980). Thus, for a methodological solipsist it does not matter whether there exists anything in the mind-external world. Ontological solipsism is a much stronger thesis which holds that nothing exists except one's own mind. Fodor (2008) rejects both kinds of solipsism.

> **Quote 1**: […] it used to seem to me that atomism about concepts means that **DOORKNOB is innate**. But now I think that you can trade a certain amount of innateness for a certain amount of mind-dependence. *Being a doorknob* is just: striking our kinds of minds the way that doorknobs do. So, what you need to **acquire the concept DOORKNOB** "from experience" is just: the kind of mind that experience causes to be struck that way by doorknobs. The price of making this trade of innateness for mind-dependence is, however, a touch of Wotan's problem. It turns out that much of what we find in the world is indeed "only ourselves". It turns out, in lots of cases, that we *make things be of a kind* by being disposed to *take them to be of a kind.* (Fodor 1998: 162, my emphasis in bold)

One interesting thing about this quote is the evident importance for Fodor of conceptual atomism and, relatedly, of lexical-conceptual isomorphism, which is the claim that atomic concepts are correlates of words. Fodor's atomicity claim, which, as mentioned, follows from his rejection of the analytic-synthetic distinction, has withstood Fodor's radical transition from concept innateness to concept acquisition. It is important now to look more closely at what possibly motivated Fodor's transition to concept acquisition and whether this move was advantageous and genuine. I turn to these questions now.

We have seen that it is Fodor's rejection of compositional semantics for conceptual correlates of words that leads him to atomism. For Fodor (2008: 66-67), primitive concepts are, roughly, correlates of words: hence concepts like CAT, BLACK, DOORKNOB or CARBURETTOR are not decomposable, he argues.

Now, there are two ways in which one may, in principle, develop the atomicity story. It may be claimed that atomicity is tantamount to innateness, i.e. that all atomic concepts, including CARBURETTOR and BUREAUCRAT, are innate. In fact, as Quote 1 above shows, Fodor used to be committed to such a view. But this way of thinking is highly controversial. CARBURETTOR and BUREAUCRAT are not even natural kind concepts – it is difficult to think that one could have an atomic concept CARBURETTOR prior to some experience with carburettors.

Note that whereas it is controversial to hold that *atomic* concepts like CARBURETTOR are innate, it seems plausible to assume that concepts like CARBURETTOR are innate as long as they are *compositionally constituted*. In other words, there is no problem with BUREAUCRAT and CARBURETTOR being innate in virtue of being composed by innately specified primitive content. But, as we have seen, Fodor rejects compositional content for conceptual correlates of words.

More recently, however, Fodor (2008) has claimed that atomic concepts, which he takes to correlate with words, are not innate, but acquired in experience. More

specifically, concepts get their 'content' referentially, by locking onto properties which things in the mind-external world possess.

1.1.1 *Referential content and Fodor's mind-dependence thesis*

For Fodor (1998, 2008), mental states are representational because they carry information about things in the world. This representational relation is established via referential semantics; semantic 'content' of the so-called atomic concepts is actually constituted by the relation concepts have to what they refer to ('lock to') in the mind-external world. For example, the concept CAT (and the word *cat*) gets its semantic 'content' from its relation to cats in the mind-external world. More specifically, semantic 'content' is 'individuated' via the relation a concept has to the properties things in the world possess. It is in this sense that Fodor can be said to espouse an externalist semantics.

It is important to emphasise, then, that for Fodor semantic 'content' is not an internal *property* of concepts. It is a *relation*. The 'content' is constituted by (i.e. is) the relation between the concept and what it applies to in the mind-external world – semantic 'content' such construed consists in the referential relation. In that sense (i.e. as a relation between a concept and something other than a concept) 'content' is actually external to the concept. This gives rise to the question of whether there is anything in Fodor's theory that constrains the 'content'-constitutive relation in a way which makes a concept like CAT get its 'content' from a relation to cats and not any other external objects.

Indeed, as observed by Burton-Roberts (ms), if a structured concept like BLACK CAT refers to black cats it does so *non-arbitrarily*. I assume (with Burton-Roberts) that Fodor would agree that the relation is non-arbitrary because of the semantic content of BLACK CAT. This semantic content is compositional, i.e. compositionally derived from its constituent concepts (and thus internally). In respect of this example, it seems reasonable to say that the referential value of the concept (assuming concepts have referential value) is *determined* by its internal compositional content. This implies a distinction between conceptual *content* and *reference*.

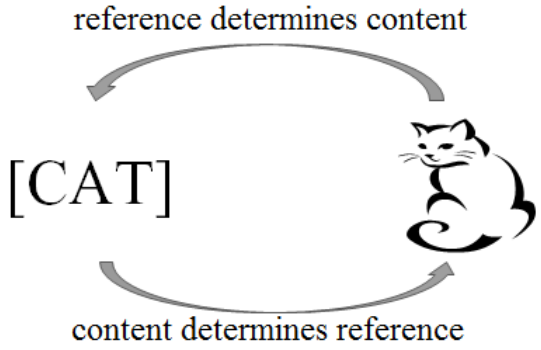
If concepts do actually lock onto things in the world, we want to say that they do so non-arbitrarily – in other words, that there is something about the concept itself (some property of the concept, which I am calling its internal content) that determines that it locks onto the things it does lock onto and not anything else. The question is then: should we align 'semantics' with (internal) content or with (external) reference? Having

so distinguished between content and reference, it seems reasonable to say that content is metaphysically prior to and a precondition for reference. Insofar as 'semantics' is referential at all, such semanticity derives from, is parasitic on, internal conceptual content. It is arguable, then, that it is internal content that is fundamentally 'semantic'.

Fodor's treatment of structured concepts is an argument for such non-referential semantics – it is the internal compositional content of BLACK CAT that determines its referent. And if that is true for phrases ('phrasal concepts'), the same should obtain for words (atomic concepts). Yet, for Fodor, when it comes to primitive concepts such as BLACK and CAT themselves, 'content' is determined – indeed consists in – the external relation of a concept to a referent. For Fodor, the 'content' of CAT or BLACK resides in a relation to mind-external things/properties – their semantics is referential and non-compositional. Hence, Fodor has no means of saying how, for example, BLACK *non-arbitrarily* locks onto black things. Why shouldn't the concepts RED, ROUND or LATE lock onto black things? The question is: what is it about any particular concept in and of itself that makes it lock onto the things it does lock onto and not other things? The problem is that if primitive concepts are to be in a non-arbitrary relation to what they 'lock to', then they must have logically prior internal content independent of, but determinative of, the 'locking' relation just like complex concepts.

However, an objection to that line of reasoning might go as follows. Fodor does not hold that BLACK CAT refers to black cats in virtue of some property internal to it. If reference composes, as it does for Fodor (2007: 18), then it is not any internal property of a concept that composes. Semantic 'content' for Fodor is, we remember, not a property of concepts but a relation to what the concept locks to. For Fodor, BLACK CAT refers to black cats in virtue of the referential content of its constituent concepts. In other words, BLACK CAT refers to black cats because BLACK refers to black things and CAT refers to cats. But notice that even if such an account of compositionality can be taken to explain why complex concepts refer to what they do, it cannot explain why atomic concepts refer to what they refer to. Figure 1.1 illustrates this point.

Figure 1.1: Fodor's circularity problem



For Fodor, the referential 'content' of CAT cannot determine its reference (the bottom arrow) because it is the referential relation that determines – in fact, is – the 'content' (the top arrow); it is in virtue of the referential relation that a concept gets the content that it does. Thus, to say that CAT refers to cats in virtue of its referential content would simply be circular. Unless there is something which is internal to atomic concepts and which determines the referential relation, the referential relation just is arbitrary and not explained.

Clearly, Fodor would not welcome this conclusion, even though it would seem to follow from a referential account of conceptual content. This is because of the previously mentioned requirement that Fodor sets for his theory – the constraint that a theory of concepts should be able to explain content identity amongst the species (1998: 30-34). Yet, if content is acquired in experience, then its identity cannot be guaranteed. Concept acquisition can proceed along different experiential paths for different individuals, *unless* there is some internal constraint on the content-constitutive locking relation. So how does Fodor resolve this problem?

According to Fodor, we acquire concepts like CAT and DOORKNOB from experiences with typical cats and doorknobs, not anything else; clearly then, for Fodor, the relation between concepts and referents is non-arbitrary. So what other than internal compositional content can determine the locking relation? Consider Quote 1 again, repeated here for convenience.

> **Quote 1**: […] it used to seem to me that atomism about concepts means that DOORKNOB is innate. But now I think that you **can trade a certain amount of innateness for a certain amount of mind-dependence**. *Being a doorknob* is just: striking our kinds of minds the way that doorknobs do. So, what you need to acquire the concept DOORKNOB "from experience" is just: the kind of mind that experience causes to be struck that way by doorknobs. The price of making this trade of innateness for mind-dependence is, however, a touch of Wotan's problem. It turns out that **much of what we find in the world is indeed "only**

> **ourselves"**. It turns out, in lots of cases, that we *make things be of a kind* by being disposed to *take them to be of a kind.* (Fodor 1998: 162, my emphasis in bold)

For Fodor (see also 1998: 148-150), the concept DOORKNOB is not innate, *just* mind-dependent – it is possible to acquire it because the properties we ascribe to doorknobs are in the mind prior to concept acquisition. Fodor posits such mind-dependent properties in order to 'naturalise' reference; if the property of *being a doorknob* is mind-dependent, then, Fodor (1998: 147) argues, laws about doorknobs are really laws about our kinds of mind. These mind-dependent properties are what constrains concept acquisition and what makes the concept-referent relation non-arbitrary.[7] I will refer to this as Fodor's 'mind-dependence thesis'.

There are two related issues here. First, the mind-dependence thesis allows that the external relation of a concept to a thing is internally determined – as in the account of structured concepts like BLACK CAT discussed above. In the light of Fodor's purported externalism about conceptual content, his mind-dependence thesis is at least surprising. I maintain that by exchanging innateness for mind-dependence Fodor is just reinstating in other words what he claims to reject – namely, that internal content determines the referential relation. In short, Fodor's mind-dependence thesis arguably amounts to the innateness of internal (i.e. non-relational/non-referential) conceptual content.

Furthermore, notice that the notion of reference that Fodor develops is not reference as traditionally understood. Canonically, for reference to occur there needs to be something to refer to. For Fodor, however, the properties that concepts refer (or lock)

---

[7] Fodor (2007: 142) agrees with Margolis that beliefs etc. may mediate the content constitutive locking relation.

> Margolis notes (rightly, I think) that the thesis that concept possession is atomistic does *not* preclude the possibility that beliefs, hypotheses, theories, and the like may mediate the causal/nomic relation between a mental representation and the property it's locked to. The stress here, however, is on the contrast between *mediation* and *constitution*. It may be that tokenings of beliefs about dogs are (sometimes? always?) links in the reference-making causal chain that connects token dogs to DOG tokens. (Hears bark; thinks: *Where there's a bark, there's a dog; therefore, there's a dog*.) It wouldn't follow that having that belief is *constitutive* of having that concept.

This is highly problematic in the light of Fodor's (e.g. 1998: 124, 2003: 3) claim that one cannot think about something unless one has a concept for it. In order to have a thought '*Where there's a bark, there's a dog; therefore, there's a dog'* one needs to have the DOG concept. This means that, on the pain of circularity, a thought about dogs cannot, even non-constitutively, mediate the acquisition of the content of the concept DOG. One can imagine the following objection to what I'm saying here. Of course, you cannot think about dogs unless you have the DOG concept. What Fodor means is that the causal/nomic relation between the concept DOG and the property of *being a dog* is about getting the *content* of the concept. Well, even if it were so, Fodor would still have to face the question of what it is that makes such a content-less concept lock onto dogs and not anything else.

to are *not* possessed by things out there in the world independently of our minds. The property of *doorknobhood* that we ascribe to doorknobs is not a fact about doorknobs, but a fact about our minds. This means that there are no doorknobs such as to refer to in any sense that supports externalism. In fact, when Fodor writes that 'much of what we find in the world is indeed "only ourselves"' (Quote 1), the following question arises: can one get more internalist than that?

Fodor's mind-dependence thesis has serious implications for his account of compositionality. Fodor (2007: 18) argues that the only semantic property that composes is reference. However, if I am right in arguing that Fodor's notion of reference is circular without presupposing the mind-dependent properties, it is not reference that composes (or if it does, it does so derivatively). For one thing, there is nothing out there in the mind-external world that a concept can refer to because things out there in the world do not possess, independently of our minds, the properties our minds ascribe to them. Secondly, what seems to compose are precisely such mind-dependent properties, which, I have argued, constitute the internal compositional content of a concept.

In summary, not only does Fodor need internalist semantics in order to get content identity, but he re-introduces it through the back door in the guise of mind-dependent properties. Fodor's referential content is circular without internalist content, and once internalist content is re-introduced, it is not in fact reference itself that composes.

My rejection of Fodor's referentialism and, relatedly, his atomicity claim, is also my rejection of lexical-conceptual isomorphism as explanatory of the word-concept relation. Fodor's mind-dependence thesis, I have argued, allows for internalist, compositional content of conceptual correlates of words. Further problems come to light when the isomorphism is considered in a cross-linguistic context. The next section addresses this.

1.1.2 *Fodor's lexical-conceptual isomorphism*

Fodor's lexical-conceptual isomorphism can be interpreted in two ways. On the first interpretation, atomic concepts are different for speakers of different languages. On the second interpretation, an atomic concept is such that there exists a word for it in some language. On both these interpretations, word-concept isomorphism is problematic.

The first interpretation is problematic for the assumption that we all share the same conceptual primitives, i.e. Fodor's atoms. In other words, it does not satisfy

Fodor's publicity constraint on concepts (1998: 28, 34) in cross-linguistic terms. Generally speaking, the isomorphism on the first interpretation is inconsistent with Fodor's view of LOT as invariant across the species.

Furthermore, the very existence of referential equivalents like English *shallow* and French *peu profond* (literal translation: *little deep*) is a problem for a referential account. Presumably, the concepts SHALLOW and PEU PROFOND have the same referential content – they apply to the same things in the world. But if the referential content of PEU PROFOND is the sum of the referential contents of its constituent parts and yet the referential content of PEU PROFOND is the same as the referential content of SHALLOW, then there is no reason why SHALLOW should not, in principle, have a compositional content – the content expressed by the French with the phrase *peu profond*. This undermines non-compositionality of conceptual correlates of words.

If isomorphism is a claim that atomic concepts are such that there exists a word for it in some language, then, in the *shallow/peu profond* example, both the English and the French have an atomic concept that for some reason is differently 'expressed' in the two languages. But now, having assumed lexical-conceptual isomorphism, one cannot explain why the French do not have a word to 'express' this atomic concept. More generally, one cannot explain why there is no one-to-one match between words and concepts cross-linguistically. But when taken to its logical conclusion, isn't that what lexical-conceptual isomorphism should predict? To put it bluntly, the problem is that lexical-conceptual isomorphism, when treated seriously, raises the question of why different languages have different vocabularies.

But this is not the end of the problems for the second interpretation. Take the English expression *a person who asks too many questions 'why'*. I take it that Fodor would agree that this expression corresponds, to use a vague term, to a complex concept in LOT. There seems to be nothing particularly controversial about this assumption, and there seems to be nothing particularly problematic for Fodor about it. Until we learn that Russians actually have a word for it – *почемучка*[8] (*pochemuchka*). Now, if one is a Fodorian theorist and one were prepared to agree that the English translation corresponds to a complex concept, then one has a problem with the Russian data. There is nothing that prevents the complex concept corresponding to the English expression from constituting the compositional content of the concept corresponding to the Russian

[8] The word is not monomorphemic, but neither is *doorknob*, which Fodor takes to correspond to an atomic concept.

word. This means that, at least methodologically, the isomorphism has to face the *shallow/peu profond* problem even on the second interpretation.

There are, obviously, many examples like that. The English word *doorknob* translates into Turkish as *kapı tokmağı*. The English expression *the day after tomorrow* translates into Polish as *pojutrze* and into Italian as *dopodomani*. The noun *palomino* translates into Romanian as *cal de culoare aurie*. Now, if a Fodorian analyst were a Romanian, especially one that is not interested in or does not care about horses, they would presumably argue that the Romanian expression corresponds to a complex concept in LOT. It would seem inconceivable to such an analyst (as a speaker of Romanian) to think that there should be a word for it. This does not mean that such an analyst would never construct a word for *palomino* or borrow it from another language.

Consider now Fodor's claim that the concept BLACK CAT is complex. What if some language developed a word for it – *blat*, for example? (Note that the word *blat* seems as implausible for us as *palomino* does for at least some speakers of Romanian.) If *blat* translates as *black cat*, we have to assume that the expressions *blat* and *black cat* 'express' concepts with the same referential content. This brings us again to the *shallow/peu profond* problem, this time with the conclusion that if Fodor is serious about his lexical-conceptual isomorphism, he has to face what is for him the problem of the creativity and productivity of language users.[9]

One final issue pertaining to lexical-conceptual isomorphism is that it is incompatible with Fodor's belief that thinking without language is possible and that there is no reason to suppose that thinking is language-dependent (2008: 218, 220). In this respect Fodor seems to be sympathetic to Chomsky's views. Consider the following quote from Chomsky (2000b: 76).

> **Quote 2**: Now what seems to me obvious by introspection is that I can think without language. In fact, very often, I seem to be thinking and finding it hard to articulate what I am thinking. It is a very common experience at least for me and I suppose for everybody to try to express something, to say it and to realise that is not what I meant and then to try to say it some other way and maybe come closer to what you meant; then somebody helps you out and you say it in yet another way. That is a fairly common experience and it is pretty hard to make sense of that experience without assuming that you think without language. You think and then you try to find a way to articulate what you think and sometimes you can't do it at all; you just can't explain to somebody what you think. Sometimes you make judgements about things very fast, unconsciously. If somebody asks you how you made the judgement, it is often extremely hard to

[9] More generally, Fodor's theory needs to face the 'problem' of accidental and socio-cultural factors involved in the development of language-particular vocabularies.

> explain. Experiences like that seem to indicate that we can and do think without language and, if you are thinking, then presumably there's some kind of conceptual structure there. The question of how this is related to language is just another research topic which, at this point, can barely be touched; but it is potentially important and interesting.

My own experience (and I believe others') chimes with what Chomsky reports here. Quote 2 suggests that thinking (at least one kind of it) and language are not one and the same phenomenon; thinking *can* and *does* take place without language. This in itself suggests that the relation between words and concepts is more complex than what isomorphism – and thus atomism – commits one to.

As mentioned, Fodor's rejection of compositional content for conceptual correlates of words follows from his rejection of the analytic-synthetic distinction on the grounds that it may be difficult, even impossible, to specify what such compositional content could be. But notice that it is impossible to specify such content *by using words*. The problem is the following. If at least one kind of thought – the kind that Chomsky reports on – is ontologically language-independent and pre-linguistic in the sense that, as Chomsky (2000a: 61) puts it, in language acquisition children 'label' pre-existing concepts, then the fact that speakers cannot 'articulate' compositional content in a particular language does not preclude the existence of such content. Given (a) Fodor's mind-dependence thesis, and therefore his internalism, and (b) Fodor's view that thinking is not necessarily language-dependent, Fodor seems closer to internalism than he overtly admits.

In the next section, I discuss Chomsky's internalist approach to semantics and reference.[10]

1.2 Chomsky's internalism

Chomsky's internalist pursuits have their basis in questions about the nature of the Faculty of Language. Therefore, before discussing Chomsky's internalist semantics specifically, I will provide a brief overview of his view of the architecture of the Faculty of Language. I will then discuss Chomsky's views on conceptual content and its relation

[10] As mentioned at the outset of this chapter, the aim of providing a critical overview of Fodor's theory was to justify my adoption of internalism. But Fodor's externalism is just one type of externalism (i.e. causal externalism). Another type is social externalism (e.g. Burge 1979, Putnam 1975), whose proponents also place themselves in opposition to internalism, and whose arguments Chomsky (2000a) rejects. I do not discuss social externalism in this chapter but return to it in some detail in chapter 3, where I look at the relation between philosophical accounts of mental content and the notion of linguistic semantics.

to natural language semantics. Here I will also identify a significant tension in Chomsky's thinking – the tension between Chomsky's view of linguistic expressions as objects doubly grounded in phonological and semantic properties (the double-interface view) and his individualistic approach to semantics. I will then discuss how Chomsky's double-interface view, but not his individualistic view of natural language semantics, has influenced Relevance Theory (e.g. Sperber & Wilson 1995, Carston 2002).

1.2.1 *Chomsky's Faculty of Language*

On Chomsky's theory (e.g. 2000a: 27, 117-118), the Faculty of Language (henceforth FL) has two components: a cognitive system that stores information and performance systems (articulatory-perceptual and conceptual-intentional) that make use of this information for articulation, interpretation, expression of beliefs and desires, talking about the world, etc. The cognitive system of FL is accessed by the performance systems, but is distinct from them. The cognitive system of FL is modified in response to linguistic experience; it changes states, from 'initial', which is innate and universal to human species, to 'I-language', which is internal (but not innate) and individual.

The initial state of FL includes general phonological and semantic principles, and the mature state (i.e. I-language) is a generative procedure that 'assigns structural descriptions to expressions and interacts with the motor and perceptual system and other cognitive systems of the mind/brain to yield semantic and phonetic interpretations of utterances' (Chomsky 2000a: 60).

I-language generates linguistic expressions (or structural descriptions, Chomsky 2000a: 26), which include instructions that performance systems access and use for interpreting thoughts on the meaning side and expressing thoughts on the sound side (Chomsky 2000a: 27-29). More specifically, I-language consists of a computational procedure (which is invariant) and a lexicon (where language variation lies). The lexicon is a set of lexical items – complexes of semantic and phonological properties (Chomsky 2000a: 175). The computational procedure selects items from the lexicon and maps them (i.e. forms an expression, a more complex array of such features) into phonetic form (PF) and logical form (LF) – i.e. it maps them to <PHON, SEM> pairs. All of this, for Chomsky, is pure syntax (Chomsky 2000a: 120, 125).

More recently, Chomsky (e.g. 2004, 2005a) has elucidated the relation between the cognitive system of FL and the performance systems in terms of the Strong Minimalist Thesis. The Strong Minimalist Thesis states that Universal Grammar – i.e. the innate and invariant part of FL – is perfectly designed to satisfy the interface

conditions, which are conditions imposed by the conceptual-intentional (C-I) and articulatory-perceptual (A-P) interfaces. Interface conditions must be satisfied '[i]f language is to be usable at all' (Chomsky 2004: 106).

This Minimalist Faculty of Language consists of NARROW SYNTAX, a computational procedure which maps a lexical array to a derivation; the PHONOLOGICAL COMPONENT, which maps a derivation to PHON; and the SEMANTIC COMPONENT, which maps a derivation to SEM. The computation maps a lexical array to a <PHON, SEM> syntactic object piece by piece, cyclically. That is, syntactic objects are constructed by narrow syntax in stages (or in 'phases'). The phonological component and the semantic component apply to units constructed by narrow syntax; the operation TRANSFER transfers a derivation to the phonological and semantic components, where they are converted into PHON and SEM, respectively (Chomsky 2004: 107-110).

In line with the Strong Minimalist Thesis, the A-P interface places a condition on narrow syntax that PHON indicates temporal order. An interface condition placed on narrow syntax by the C-I interface is that SEM anticipates a duality of semantic interpretation at C-I. One kind of such properties has to do with argument structure (theta-theoretic properties). The other kind involves everything else, namely scopal and discourse-related properties (Chomsky 2004: 110).

Chomsky (e.g. 2004, 2005a, 2005b) argues that there are three factors involved in the determination of I-languages: 1) genetic endowment, 2) experience and 3) principles that are language- or even organism-independent. Investigating the third factor allows a principled explanation of the properties of language – explanation in terms that reach beyond the linguistic. Chomsky's quest for principled explanation of FL in terms of its interaction with the interfacing systems has given rise to the view of FL as a *non*-autonomous module of the mind.

What is of particular interest to me now is the extent to which the semantic component of FL is determined by the conceptual-intentional system and the way it interacts with it. This brings us to the next sub-section, where I consider Chomsky's views on concepts (i.e. elements of the conceptual-intentional system) as well as the relation and interaction between them and the SEM element of the Chomskyan <PHON, SEM> object.

1.2.2 *Two kinds of semantic content?*

When discussing Chomsky's stance on semantics, it is important to note his (2007: 14) asymmetry assumption, which is the view that language is primarily optimised to the

conceptual-intentional interface. Mapping to the sensory-motor (i.e. articulatory-perceptual) interface is merely an 'ancillary procedure'; it is used only for externalisation and involves morpho-phonology[11], which is there to satisfy the 'linking condition'. The linking condition, which can be traced back to Aristotle, states that language links sound and meaning (Chomsky 2005a: 10) and has given rise to Chomsky's view of linguistic expressions as syntactic objects constituted by both semantic and phonological properties (i.e. double-interface objects). The double-interface idea will be discussed in more detail in chapter 2. Now I will concentrate on the interaction of semantic properties of Chomsky's linguistic expressions and their relation to the conceptual-intentional system. Consider quote 3 below.

> **Quote 3**: Generation of expressions to satisfy the semantic interface **yields a "language of thought."** If the assumption of asymmetry is correct, then the earliest stage of language would have been just that: a language of thought, used internally. It has been argued that an independent language of thought must be postulated. I think there are reasons for skepticism, but that would take us too far afield. (2007: 14, my emphasis)

This quote is revealing in several ways. If language is primarily optimised to the conceptual-intentional performance system (and since mapping to PHON is a 'secondary process', 2005a: 4), then, for Chomsky, language is primarily (in evolutionary, but also in functional terms) employed for thinking. We can see then why Chomsky is so concerned with the nature of concepts, i.e. constituents of thought. However, the view of language as primarily used for thinking gives rise to the following question: given that Chomsky distinguishes at least two kinds of thinking, namely language-independent thinking and thinking which involves language, how are the two types of thinking different (in the sense of Quote 2)? In other words, how does natural language semantics differ from the semantics of LOT? Answering this question becomes difficult in the light of the following quote:

> **Quote 4**: […] the core theory of language—Universal Grammar (UG)—must provide, first, a structured inventory of possible lexical items that are **related to** or perhaps **identical with** the concepts that are the elements of the ''cognoscitive powers,'' sometimes now regarded as a ''language of thought'' along lines developed by Jerry Fodor (1975)[12]; and second, means to construct from these lexical items the infinite variety of internal structures that enter into thought, interpretation, planning, and other human mental acts, and that are

[11] For Chomsky (e.g. 2007: 8) phonology includes morphology.
[12] Prior to his externalist commitments in Fodor (1998) and (2008), Fodor argued that conceptual correlates of words were innate.

> sometimes put to use in action, including the externalization that is a secondary process if the speculations just reviewed turn out to be correct. (2005a: 4, my emphasis)

It is clear from Quote 4 that the primitive elements of natural language semantics are strongly associated with the semantics of thought: they are 'related to' or perhaps even 'identical with' them. But if we are to make any sense of that, we need to know how the two kinds of semantics – that in LOT and that in UG – are different if they are merely related or, if they are identical, whether it is conceptually necessary to duplicate them, i.e. to posit two loci of the same semantic elements.[13]

If natural language semantics and the semantics of thought are 'related', then presumably they are grounded in the same conceptual-intentional modality (i.e. they both deal with concepts), but are structurally distinct.

Relevant here is the analytic-synthetic distinction. Fodor, as we have seen in 1.1.1, rejects the analytic-synthetic distinction and is therefore committed to atomic (i.e. non-structured) and therefore purely referential content. Chomsky, however, endorses the analytic-synthetic distinction and hence *possibly* also the notion of molecular semantic content for conceptual correlates of words. I used the word 'possibly' for two reasons; the first is that the existence and endorsement of the analytic-synthetic distinction does not necessarily amount to the existence and endorsement of molecular linguistic semantics and the second is that Chomsky is actually not unsympathetic to meaning holism. I explore the first of these points in the remainder of this section, and the second one in 1.2.3.

Chomsky (e.g. 2000a: 61-67) argues that there is enough linguistic evidence to support the analytic-synthetic distinction. Consider words like *chase* or *persuade*, which, as Chomsky observes, involve reference to human intention:

> **Quote 5**: To chase Jones is not only to follow him, but to follow him with the intent of staying on his path, perhaps to catch him. To persuade Smith to do something is to cause him to decide or intend to do it; if he never decides or

[13] Following from this last comment is the question about the rationale for positing the C-I interface at all. If the semantic elements of UG are 'identical to' the semantic elements of LOT, then, arguably, the semantic component of UG just is the language of thought. The following quotes from Chomsky are relevant here:

> The simplest thesis is that an expression E has no existence apart from its properties at the interface levels, PHON(E) and SEM(E) (**if these exist**). (2000a: 175, my emphasis)

> On very weak assumptions about efficient computation, there should be no linguistic levels beyond those that are imposed by interface conditions: that is, the interface levels themselves. In fact, **it is not even obvious that these exist**. (2005a: 16, my emphasis)

> intends to do it, we have not succeeded in persuading him. Furthermore, he must decide or intend by his own volition, not under duress […] (Chomsky 2000a: 62)

What follows from this observation is the distinction between 'truths of meaning' (i.e. analytic truths) and 'truths of fact' (i.e. synthetic truths). If Tom chased Jones, then he necessarily followed him intending (at least) to stay on his path. If Mark persuaded Smith to join the Chester-le-Street amateur rowing club, then at some point in time Smith decided or intended to join the Chester-le-Street amateur rowing club. In a similar vein, if John killed Bill, then Bill is dead. These are all 'truths of meaning' – we arrive at them just by virtue of knowing language, i.e. without having to consider or know what the world is like. This, for Chomsky, is a good indication that 'there are necessary connections among concepts, **reflected** in connections of meaning among words' (2000a: 63, my emphasis). The word 'reflected' here is crucial because it clearly indicates that when Chomsky says that there is no reason why semantic connections should not be 'completely fixed and stable as a matter of biological endowment', he is talking about connections among concepts, and *not* linguistic expressions. The latter merely 'reflect' such connections.

This distinction comes out even more clearly when Chomsky (2000a: 66) writes that in language acquisition the child's task is to 'discover labels' for the innate stock of concepts. The figures below will help me illustrate this point.

Figure 1.2: *blimpft*

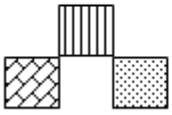


Figure 1.3: *molont*

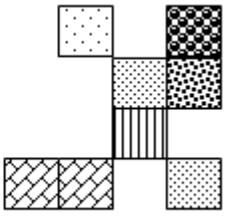


Figure 1.2 and figure 1.3 share the same configuration of three blocks. We can even say that the configuration of blocks in figure 1.2 is contained in the configuration presented in figure 1.3. Now, if we assume for the sake of this illustration, that the individual blocks represent primitive concepts and that the patterns formed out of them represent structured concepts, we can say that the structured concept in figure 1.2 is entailed by the structured concept in figure 1.3 because it is contained within it. In other words, there is an analytic relation between the configurations/structured concepts in the two figures. Let me now label the structure in figure 1.2 as *blimpft* and the one in figure 1.3 as *molont*. This allows us to say – by proxy – that the word *blimpft* is contained within

the word *molont*, that the word *molont* entails the word *blimpft*, and that there is an analytic relation between them. But notice the importance of 'by proxy' here; we know that it is not the word *blimpft* that is contained in the word *molont*, rather it is the structure which I have labelled *blimpft* that is contained within the structure that I have called *molont*.

It seems to me that when Chomsky talks about analytic relations he makes a similar distinction – the analytic relations concern concepts (i.e. SEM) and not their morpho-phonological labels (i.e. PHON) or linguistic expressions/words (i.e. <SEM, PHON>) as such. Let me explain this.

There are two aspects of Chomsky's linguistics that are relevant here. The first is his view of linguistic expressions as double-interface objects constituted by morpho-phonological and semantic properties. The second is the notion of I-language, with the emphasis on the individualistic nature of it. The first point to observe is that even though, for Chomsky, morpho-phonological and semantic properties together constitute a syntactic object, the fact remains that the relation between them is arbitrary, i.e. non-natural (it is the locus of Saussurean arbitrariness). The implications are the following. In my blocks scenario (figure 1.2 and figure 1.3), I have argued that the analytic relation does not hold between *blimpft* and *molont* but between the structures that these nonce words merely label. This argument can be further extended to say that the analytic relation does not hold between (i) the object constituted by the structure that the label *blimpft* labels and the label itself on the one hand, and (ii) another object constituted by the structure that the label *molont* labels and the label itself, on the other. The analytic relation does not hold between such doubly constituted objects because the relation between the structure and the label is arbitrary.

To explain this, let us imagine that I was wrong in calling those structures *blimpft* and *molont* respectively – I confused them with some other labels. The structures presented in figure 1.2 and figure 1.3 should actually be called *plumtug* and *gufhan*. The fact that the labels can change, while leaving the relation between the conceptual structures intact, clearly shows that the analytic relation holds between the conceptual structures only, rather than between labels or objects constituted by structures and labels. In other words, logical relations do not care about labels. This argument is necessarily extendable to natural languages because the relation between morpho-phonological and semantic properties is arbitrary, i.e. non-natural. Evidence for this comes from cross-linguistic, cross-dialectal and diachronic considerations: the same structure is labelled *chase* in English, *gonić* in Polish and *inseguire* in Italian; the same

thing is called *plimsoll* or *gym shoe* or *sandshoe* in English; no-one uses the Middle English label *þou* anymore. Analytic relations thus clearly do not hold between labels or linguistic expressions (i.e. double-interface objects). Chomsky seems to mean precisely that when he writes that it is 'the *a priori* framework of human thought, within which language is acquired, [that] provides necessary connections between concepts' (2000a: 62-63).

I also mentioned that the view of analyticity as pertaining solely to concepts follows from Chomsky's insistence on the individualistic nature of I-language. It is well known that the word *disinterested* is synonymous with *impartial* for many, but not all, speakers of English. It follows then that many, but not all, speakers of English will use the two words interchangeably. They will do so because in their minds the semantic structures labelled by the two words are similar, if not identical. That the analytic relation exists only between the semantic structures labelled is clear from the empirical fact that other speakers do not use the two words interchangeably. For those that do not, there is no analytic relation between the semantic structure labelled *disinterested* and the semantic structure labelled *impartial* (but there is such relation for them between the structure labelled *disinterested* and the structure labelled *uninterested*). This, and other more idiosyncratic uses like that of *pineapple* instead of *pinnacle* (Burton-Roberts 2012), demonstrates that there is a decisively individualistic element implicated in assigning and using labels. It also demonstrates that analyticity judgments are an individualistic matter.

These individualistic insights into analyticity judgements highlight a problem in Fodor's approach. Fodor's rejection of the analytic-synthetic distinction follows from his identification of concepts with word meanings/semantics (lexical-conceptual isomorphism). If concepts are to be shared amongst the human species but are also taken to equal word meanings/semantics, then such concepts cannot be internally compositional and shared precisely because internal compositionality of a concept which corresponds to a word is an individualistic matter. For Fodor, atomism follows. However, I have argued that Fodor's identification of concepts with word meanings/semantics is misguided. Once a distinction is made between concepts as such and the way a concept functions as a word's semantics, it is not necessary to posit the anyway problematic atomism to maintain the nativist claims about concepts.[14]

[14] I shortly return to the question of whether such an unambiguous distinction is actually made by Chomsky himself.

Now, if Chomsky is right in his I-assumptions – i.e. internalist and individualistic assumptions – about word meaning/semantics, then there is no interesting (i.e. cross-speaker and/or cross-context stable) notion of linguistic semantics. But before I discuss this point in more detail, let me investigate another reason behind Chomsky's I-assumptions, i.e. his observations about reference.

1.2.3 *Chomsky on reference*

What clearly distinguishes Chomsky from Fodor are their different notions of content. Fodor, as mentioned, endorses a notion of a concept's 'content' that is constituted by a relation to something external to the concept. Such 'content' is not a property of a concept; it is a relation. Fodorian 'content' is experientially acquired. Chomsky, by contrast, endorses a notion of content as a property of a concept. Chomsky has an innate, internalist, compositional notion of content. In principle, this contrast between Fodor and Chomsky is very clear, even though Fodor's mind-dependence thesis undermines his externalism. As argued in section 1.1, externalism about conceptual content is problematic if only because it is circular and not explained without presupposing internalist content.

Fodor (e.g. 2008: 16) argues that internalism is problematic because it leads to solipsism. Solipsism, in turn, is wrong because, to use Fodor's (2008: 53, fn 4) words, 'we use 'tables' and 'chairs' to talk about tables and chairs (respectively)'. I am re-iterating these issues here because internalism is not an uncontroversial position and is explicitly an undesirable one for referentialists. Fodor writes:

> **Quote 6**: Patently, the referentialist view that semantics is about a relation of reference that holds between symbols and the world is incompatible both with 'ontological' solipsism (the view that there isn't anything in the world) and 'methodological' solipsism (the view that it doesn't matter to semantics whether there is anything in the world). (Fodor 2008: 16 fn 29)

I maintain that even though the purpose of Fodor's referentialist enterprise was to come up with a non-solipsist account, he has failed to provide it. And even though he argues that no idealism follows from his theory because 'doorknobs are real but mind-dependent' (Fodor 1998: 150), I hope to have successfully shown in section 1.1 that, despite his proclaimed externalism, Fodor himself is a methodological solipsist precisely by virtue of his mind-dependence thesis. This suggests that there is no escape from methodological solipsism, which, at the end of the day, is the claim that there are mind-external entities, but that we can only experience them through the 'perspectives'

or ‘lenses’, to use Chomsky’s (e.g. 2000a: 36, 42) terms, allowed by our minds. In other words, we cannot experience the world (e.g. doorknobs) independently of our minds. This seems even pre-theoretically uncontroversial, and theoretically, I have argued, it is the only place to start from.

Indeed, Chomsky (e.g. 2000a: 127-129) argues that reference is much more fine-grained than object reference allows for. The example that he famously discusses in this context is that of the name *London*. Chomsky (2000a: 37) observes that we can think and talk about London in terms of ‘a location or area, people who sometimes live there, the air above it (but not too high), buildings, institutions, etc., in various combinations’. Moreover, he notes that ‘under some circumstances, it [London] could be completely destroyed and rebuilt somewhere else, years or even millennia later, still being London, that same city.’ For Chomsky then, ‘there neither are nor are believed to be things-in-the-world with the properties of the intricate modes of reference that a […] name encapsulates’ (2000a: 37). Hence, what we ever refer to is the world as seen through ‘the perspectives made available by the resources of the mind’ (Chomsky 2000a: 16).

Ludlow (2003) remarks that Chomsky’s concerns about referential semantics are, at their root, concerns about what he calls ‘the language/world isomorphism’. This isomorphism is a correlation between something in the world (an object, an event) and its mental representation. What Chomsky’s London example shows is that there is no isomorphism between an entity in the world and the linguistic expression *London* that we use when we talk about London. More precisely, there is nothing in the mind-external world such that it corresponds to the semantic properties ‘encapsulated’ by the term *London*.

Chomsky’s arguments against the language-world isomorphism can be grouped into three inter-related categories. Ludlow (2003) calls them the Implausible Commitments Argument, the “Type Mismatch” Argument and the Misbehaving Object Argument. I will discuss them now and then, in the following sub-section, I will try to identify some insights that these arguments offer on Chomsky’s views on the relation between natural language semantics and the semantics of the C-I system.

The Implausible Commitments Argument concerns reference to fictional entities such as unicorns or Santa Claus, but also, and perhaps more fundamentally, it is related to Chomsky’s discussion of terms like *London*, i.e. terms which *prima facie* could be granted an external referent. The argument emphasises the fine-grained nature of reference and, following from this, the implausibility of there being in the mind-external world any entities that correspond to what we are talking about when using language.

Other expressions that Chomsky famously discusses in this context are *the flaw in the argument* and *the average man*. Chomsky observes:

> **Quote 7**: If I say "the flaw in the argument is obvious, but it escaped John's attention," I am not committed to the absurd view that among things in the world are flaws, one of them in the argument in question. Nevertheless, the NP *the flaw in the argument* behaves in all relevant respects in the manner of the truly referential expression *the coat in the closet* – for example, it can be the antecedent of it and serves as an argument, taking a θ-role. (Chomsky 1981: 324)

Just as there are no things in the mind-external world such that correspond to terms like *London* and *the flaw in the argument*, neither are there entities corresponding to expressions like *the average man* (consider '*The average man has 2.3 children'*). For Chomsky neither the world nor the cogniser's mental model is constituted by entities that speakers describe as things that concern them (e.g. Chomsky 2000a: 135-136).

The "Type Mismatch" Argument concerns a disparity between the type of individuation that objects and substances intuitively have and the type of individuation that a referential semantics provides. For explanatory purposes, Ludlow (2003: 149) makes a distinction between a P-substance and an I-substance. In the context of the expression *water*, the P-substance would be H_2O, i.e. that sort of matter that would play part in a physical theory, whereas the I-substance would be that which we intuitively talk about when we are using language. The problem, Ludlow (2003) argues, is that, in terms of referential semantics, *water* would be assigned the P-substance (H_2O) as its semantic value, but the I-substance we are talking about when we use the term *water* is often something else. For example, what we find in rivers is often called *water*, but it would rarely be considered H_2O. Furthermore, it appears that there are things that have chemical composition closer than river water has to H_2O, but that we do not call *water*. Tea is one such example. The general argument then is that P-substances and I-substances do not match up.

The third argument against referential semantics, the Misbehaving Object Argument, is closely related to the first two. The "Type Mismatch" Argument was illustrated by the usage of the term *water*, but what the case of *water* also illustrates is that the I-substance talked about when using this word is, as Ludlow (2003: 15) puts it, 'a most ill-behaved sort of substance'. Chomsky imagines the following scenario:

> **Quote 8**: Suppose cup_1 is filled from the tap. It is a cup of water, but if a tea bag is dipped into it, that is no longer the case. It is now a cup of tea, something different. Suppose cup_2 is filled from a tap connected to a reservoir in which tea

> has been dumped (say, as a new kind of purifier). What is in cup_2 is water, not tea, even if a chemist could not distinguish it from the present contents of cup_1. The cups contain the same thing from one point of view, different things from another; but in either case cup_2 contains only water and cup_1 only tea. (Chomsky 2000a: 128)

The point that Chomsky is making here is that whether we regard something as water or not depends on the intricate interaction of human interests, concerns, beliefs and intentions. The fact that I-substances, or what we may call I-meanings, are so 'unruly' follows naturally from the assumption that meaning/semantics of linguistic expressions is determined by internalist-individualistic psychology. Chomsky (e.g. 1975, 2002) argues that even apparently simple concepts like that of object do not correspond to what we may describe as P-substances in any interesting way. He observes that spatiotemporal contiguity, a *prima facie* good candidate for identifying P-substances, is not sufficient; a wing of an aeroplane is an object, but its left half (though equally continuous) is not, a herd of cattle is not an object, but a picket fence with breaks is (though equally discontinuous). Whether something is called an object or not, Chomsky (1975: 203) concludes, is crucially determined by beliefs about human will and intention.

Even the simplest of words integrate information about such properties as material constitution, design, intended use, origin, causal properties, etc. (e.g. Chomsky 2002: 87). More interestingly, lexical items often provide conflicting perspectives on how to view the world. Books, for example, can be both abstract and concrete; if two copies of the same book are taken out from the library by two people, then in the abstract sense they have taken out the same book, but in the concrete sense a different one. Chomsky writes:

> **Quote 9**: […] the internal conditions on meaning are rich, complex, and unsuspected; in fact, barely known. The most elaborate dictionaries do not dream of such subtleties; they provide no more than hints that enable the intended concept to be identified by those who already have it (at least, in essential respects). (Chomsky 2000a: 36)

There are two points to be made about quote 9. On the one hand, the semantics, or the perspectives for viewing the world that lexical items afford, are genetically given. Chomsky claims that a child would not be able to understand the richness and complexity of meaning if it were not – in some 'essential respects' – innately pre-determined (see also 2000a: 61). It seems then that such innate semantics should, in principle, be subject to a naturalistic inquiry, like that of Chomsky's. But it is not; the

question about the meaning/semantics of linguistic expressions falls outside Chomsky's naturalistic enquiry into the nature of FL (e.g. Chomsky 2000a: 20, 36). The reason for this is that, even though perspectives – or let us call them conceptual primitives – are innate and hence species-invariant, the associations between morpho-phonological labels and conceptual structures are dependent on context and individualistic psychology. If we combine the innateness thesis with the individualistic nature of language, Chomsky's point seems to be that even though semantic content (i.e. conceptual content) is innate, the role it plays in language use – or the way it is involved in meaning/semantics of linguistic expressions – is context-variable and speaker-variable.

Another thorny issue that Chomsky hints at in discussing natural language semantics concerns the intricacy and often paradoxicality of the perspectives that a linguistic expression can provide; the abstract and concrete aspects of a book discussed above are a case in point. But let us take a closer look at a word like *cow*. If we wanted to argue that natural language semantics was subject to a naturalistic inquiry, we would have to find a level of semantic representation that corresponds to the word *cow* and that is species-invariant. Presumably, such semantic representation would have to be abstract enough as to underdetermine all uses of the word *cow*. The intuitive, pre-theoretical problem is that a word *cow* can provide such conflicting perspectives as [animate] as well as [inanimate] ('*Would you like a cow or a dog?*', asked by a mother to a child at a toy shop), [human] ('*She's a cow*', spoken derogatorily of a woman) alongside [non-human], [male], [female], [white], [brown], [purple], [dotty], [dead], [alive], [fleshy], [fluffy], [made of clay] among millions of other possible perspectives. It is not surprising then that Chomsky (2000a: 132) regards a naturalistic inquiry into natural language semantics as ill-fated.

The consequences are profound: from Chomsky's internalism-individualism it follows that the meaning/semantics of linguistic expressions is neither stable nor shared in any interesting way, i.e. it is not shared in a way that supports naturalistic (or social externalist) accounts of linguistic semantics. This is the essence of I-language. Fundamentally, Chomsky (e.g. 2000a: 137) argues that a radically individualistic approach to the study of natural language meaning/semantics – i.e. a holistic approach, which involves the study of beliefs mediating between us and things in the world – seems to be a 'fruitful' direction to pursue.[15]

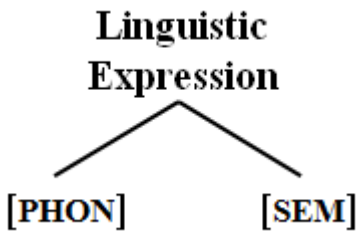
[15] In this context, Chomsky (2000a: 137) makes reference to Bilgrami's (1992) radically contextualist thesis about unity and locality of semantic content. Bilgrami's approach is discussed in chapter 3.

However, this sympathy of Chomsky's for holism and individualism is in tension with his view of language, and thus of linguistic expressions, as doubly constituted by phonological and semantic properties. The next section is concerned with this tension.

1.2.4 *Chomsky's legacy in research into semantics and pragmatics*

As discussed earlier, Chomsky (e.g. 2005a: 10) reconstructs Aristotle's idea – that language links sound and meaning – by attributing both phonology and semantics to linguistic expressions. On this view, linguistic expressions are taken to be double-interface syntactic objects constituted by phonological and semantic properties (<PHON, SEM>). The following figure illustrates this.

Figure 1.4: The double-interface object



Following from this double-interface assumption is the view that (i) words, and linguistic expressions generally, have meaning in virtue of encoding linguistic semantics *as their constitutive property* and that (ii) it is in virtue of having meaning as their constitutive property that words guide interpretation to specific regions of the conceptual space. For example, in Relevance Theory (e.g. Sperber & Wilson 1995, Carston 2002), which considers itself an adjunct to (/an extension of) Chomskyan tradition (Wedgewood 2007: 679), the encoded linguistic semantics is necessarily and always decoded by the language module and delivered as logical form to the pragmatic processor.

The point that I am making here is this. The view that there exists a specifically linguistic, i.e. non-pragmatic, semantic level follows from the view that linguistic expressions are double-interface objects. Indeed, in the light of the double-interface view of language, Chomsky's (2005a: 4) claim that natural language semantics may be 'related' to the semantics of thought and Chomsky's (e.g. 2000a: 62) endorsement of the analytic-synthetic distinction may be interpreted in favour of some notion of linguistic semantics. On the other hand, Chomsky's (2005a: 4) claim that natural

language semantics may be 'identical' with – i.e. the same as – the semantics of thought, and Chomsky's (e.g. 2000a: 137) endorsement of holism and individualism (as well as a careful reading of Chomsky's writings on the analytic-synthetic distinction) are incompatible with the notion of linguistic semantics. Indeed, Chomsky concedes:

> **Quote 10**: […] that any such thing as I-meaning ("semantic representation," "narrow content") even exists is now commonly denied. Comparable questions about I-sound have rarely been raised. The empirical disciplines seem to me to study them in much the same way: in particular, to assume that both involve invariant universal features of which LIs [linguistic expressions] are constituted (and hence are not radically holistic). I will **tentatively** assume that postulation of I-sound and I-meaning is legitimate, returning to reasons for denying it. (Chomsky 2000a: 170, my emphasis)

The problem, for Chomsky, is that if there is no interesting (i.e. cross-speaker and/or cross-context shared) notion of linguistic semantics (i.e. no interesting notion of SEM) – as implied by his I-assumptions – it is not clear in what sense a linguistic expression can be a double-interface <PHON, SEM> object. This tension in Chomsky's thinking – that there is SEM because language deals with double-interface objects and that there cannot be SEM because of the I-nature of words' meanings/semantics – is the leitmotif of this thesis.

1.3 Conclusion

This chapter has engaged with two related questions: about the nature of semantics and about the relation between words and concepts. As for the first question, I argued that Fodor's notion of referential semantics does not offer a plausible answer to the question of how we think about the mind-external world as it presupposes innate, non-referential, compositional content in the guise of mind-dependent properties (e.g. Fodor 1998: 141, 162). I showed that the account of referential semantics is circular unless it presupposes such innate non-referential content. Fodor's mind-dependence thesis, I concluded, is actually a re-introduction of the Chomskyan notion of conceptual content, i.e. content which is an innate, constitutive, compositional property of concepts.

What motivates Fodor's account of referential semantics, I suggested, is his lexical-conceptual isomorphism, i.e. a view that primitive concepts correspond to word meanings/semantics. I have argued that this isomorphism yields an inappropriate account of the word-concept relation as it is incompatible with cross-linguistic evidence and with Fodor's own view that thought is language-independent.

The problems with Fodor's theory provide ample justification for the adoption of Chomsky's internalist account of conceptual content, which was introduced in the second part of the chapter. The aspects of Chomsky's theory that I endorse are the following:

a) **Innateness**: conceptual primitives, and structures defined over them, are innate and hence species-invariant
b) **Compositionality**: conceptual correlates of words are structured, i.e. compositionally constituted by conceptual primitives
c) **Context-variability**: words often 'encapsulate' conflicting perspectives, i.e. their meaning/semantics is highly context-variable
d) **Individualism**: the meaning/semantics of words is individualistic, i.e. the conceptual structures 'encapsulated' by words vary from person to person

In the light of the problems with Fodor's account, assumptions (a) and (b) are rather uncontroversial: from (a) and (b) it follows that a conceptual structure associated with a word like *carburettor* is innate in the sense that the conceptual primitives by which it is compositionally constituted are innate, as is the generative computation that allows for such composition. Problems start when we consider assumptions (c) and (d) in detail. As mentioned, Chomsky's I-assumptions, if correct, have serious implications not only for theories postulating the existence of some form of linguistic semantics, but also for the double-interface view of language as such.

In the next chapter, I criticise Relevance Theory's notion of linguistic semantics and argue that such content is implausible and redundant in accounting for meaning in language. I introduce the Representational Hypothesis (e.g. Burton-Roberts 2012), a conceptual programme which rejects the double-interface view of language and, along with it, the assumption that words have meaning as a constitutive property. In line with the Representational Hypothesis, I argue for a semiotic, wholly inferential, account of utterance interpretation, and explain the mechanics of such an account in terms of Hintzman's (e.g. 1986) multiple-trace theory of memory.

Chapter 2. The locus of semantics and the decoding-inferring distinction

2.0 Introduction

The aim of chapter 1 was to show that Fodor's referential notion of semantic content does not offer a plausible answer to the question of how we think and talk about the mind-external world. One reason is that it presupposes (and is circular without presupposing) internalist, compositional content and the other is that it is based on a methodologically unsound assumption of lexical-conceptual isomorphism. The conclusions were that the nativist epistemology is necessary in explaining human cognition and that the relation between words and concepts is more complex than Fodor's lexical-conceptual isomorphism predicts.

The discussion of problems with Fodor's account brought me to Chomsky's internalism about conceptual/semantic content. I identified four aspects of Chomsky's thinking that I endorse: (a) innateness of conceptual primitives (and structures defined over them), (b) internal compositionality of a word's semantics, (c) its context-variability and (d) its individualistic nature. However, I argued that there is a tension in Chomsky's account between his internalist-*individualistic* assumptions and the double-interface view of language and linguistic expressions.

As discussed in 1.2.1, for Chomsky (e.g. 1995) Language (i.e. FL) is subject to a realist/naturalistic enquiry in terms of necessary principles – it is 'a real object of the natural world'. By hypothesis, Language is 'perfect', which means that it is free of any form of variation – there is nothing in it that is not conceptually necessary. Crucially, for Chomsky (e.g. 2007), the two systems that Language interfaces with, C-I and A-P – and hence SEM and PHON components of Language – are indeed conceptually necessary. Given Chomsky's realist/naturalistic concerns, it should follow that since SEM is a necessary component of Language, which is natural, SEM too should be subject to a formal explanation in terms of necessary principles. Thus, positing SEM, i.e. some notion of linguistic/lexical semantics which is cross-speaker and cross-context shared follows from Chomsky's realist/naturalistic inquiry into Language as the double-interface system.

However, as mentioned in chapter 1, Chomsky's I-assumptions are incompatible with positing some such form of SEM. Indeed, Chomsky (2000a: 170) concedes that he

only 'tentatively' assumes that it exists.[1] The problem is that Chomsky's acknowledgement of the variability (individualistic and contextual) in the association between a word form and semantic properties suggests that, for him, word 'meanings' cannot be shared in any interesting sense, and yet the double-interface idea depends on the existence of SEM (and PHON). This gives rise to the question of whether or not it is possible and necessary to posit SEM.

In this chapter, I address (i) the question of whether the notion of shared semantic content is theoretically sound, and if not (ii) whether it is possible to account for mutual understanding between interlocutors in communication without positing the level of shared semantic content.

As mentioned in 1.2.4, Relevance Theory (e.g. Sperber & Wilson 1995, Carston 2002) places itself within the Chomskyan tradition. For this reason, I answer the first question (i) by looking at Relevance Theory's (e.g. Carston 2002, 2010) notion of linguistic semantics. In 2.1, I argue that Relevance Theory does not actually offer an account of linguistic semantics which is shared among the members of the same speech community or across contexts. The appeal to linguistic semantics is thus not doing the work it was devised to do in the first place. The second problem is that, even in Relevance Theory's terms, the process of deterministic decoding of linguistic semantics is redundant; depending on the interpretation of Relevance Theory, linguistic content of a word is either decoded only to be 'dropped and replaced' or it is anyway pragmatically inferable. Therefore, I argue that the notion of specifically linguistic semantics should be rejected.

In 2.2, I discuss Burton-Roberts' (e.g. 2009) criticism of the double-interface tradition. I then introduce the Representational Hypothesis (e.g. Burton-Roberts 2000, 2012; Burton-Roberts & Poole 2006a; Chng 1999), a conceptual programme which rejects the double-interface view of linguistic expressions (<PHON, SEM>) as problematic and anyway unnecessary to account for linguistic 'sound with a meaning'. In rejecting the double-interface tradition, the Representational Hypothesis rejects the notion of linguistic semantics (i.e. it attributes no semantic properties to linguistic expressions) and offers a semiotic, wholly inferential, account of meaning in language.

Finally, in 2.2.3, I argue that the mechanics of the representational account can be implemented in terms of Hintzman's (e.g. 1986) multiple-trace theory of memory. Like the Representational Hypothesis, Hintzman's model predicts that the

[1] Similar claims follow, and similar problems arise, on the PHON side of the double-interface object (see e.g. Carr 2000). This issue, however, goes beyond the scope of this thesis.

understanding of speaker-intended meaning is not mediated by the cognitive process of deterministic decoding of specifically linguistic concepts. Rather, it is a wholly pragmatic, inferential process. I then discuss some advantages of a wholly inferential model over Relevance Theory's two-step (i.e. 'decode and infer') model.

2.1 Linguistic semantics in Relevance Theory

In the double-interface tradition, linguistic signs have meaning in virtue of encoding linguistic semantic properties: the word *cat* is meaningful because it constitutively encodes (i.e. has as its property) the concept CAT. This effectively equates having meaning with having semantic properties.

A consequence of this assumption is that to account for meaning in language we need to posit two kinds of semantics – that of words and, on the assumption that thoughts too have semantics, that of thoughts. In Relevance Theory (e.g. Carston 2002), this idea is developed in terms of the distinction between linguistic semantics and real semantics, respectively. I explain this distinction below.

For Relevance Theory (e.g. Carston 2002, henceforth RT), the linguistic system 'maps' a phonetic representation into a semantic representation, which is 'a structured (presumably conceptual) mental representation' (Carston 2002: 9). This semantic representation, or 'encoded linguistic meaning', is to be understood in terms of 'some appropriate notion of logical form, computed by linguistic decoding system' (*ibid*.).

One of Carston's (2002) aims is to work out the details of the relation between the assumed linguistic meaning decoded by the language faculty and occasion-specific interpretations of utterances (i.e. propositions, or thoughts, communicated by utterances). This concern of Carston's reflects RT's distinction between linguistic meaning and speaker meaning. LINGUISTIC MEANING is context-independent and refers to 'relatively stable meanings in a linguistic system, meanings which are widely shared across a community of users of the system' (Carston 2002: 19-20). In contrast, SPEAKER MEANING refers to a thought communicated by a particular speaker on a particular occasion. Hence, speaker meaning is to a large extent individualistic and context-dependent. The semantic content of the logical form en/decoded by the language faculty is referred to as LINGUISTIC SEMANTICS, whereas the semantic content of the thought/proposition communicated by the speaker is referred to as REAL SEMANTICS.

Whereas linguistic semantics, in virtue of being context-independent, does not have truth-theoretic properties (i.e. it is not true of false with respect to any state of affairs in the world), real semantics has truth-theoretic properties (i.e. it is true or false

with respect to some state of affairs in the world). Fundamentally, thus, the distinction between real and linguistic semantics is made in terms of truth-theoretic properties or lack of them, respectively. Importantly, Carston (2002: 11) also explains the linguistic-real distinction in terms of two types of cognitive processes: DECODING and INFERRING. It is linguistic meaning that is decoded and it is speaker meaning that is inferred.

Related to the claim about the non-truth-theoretic nature of linguistic semantics is RT's claim that linguistic meaning UNDERDETERMINES utterance meaning, i.e. it underdetermines the thought/proposition communicated by a speaker. It is because linguistic meaning is non-truth-theoretic that speaker meaning cannot be determined by linguistic decoding alone but has to be derived by pragmatic inference (from context). RT's linguistic UNDERDETERMINACY THESIS pertains to all levels: 'While sentences encode thought/proposition templates, words encode concept templates; it's linguistic underdeterminacy all the way down' (Carston 2002: 360).

One of the key insights of Relevance Theory (e.g. Sperber & Wilson 1987, Wilson & Sperber 2004) is that the inferring of speaker meaning is guided by the pragmatic principle of RELEVANCE. A stimulus (e.g. an utterance) is taken to be relevant if it yields a POSITIVE COGNITIVE EFFECT. In other words, a stimulus is relevant if it makes a difference to an individual's representation of the world by improving the individual's knowledge, confirming a suspicion, settling a doubt, etc. The central type of positive cognitive effect caused by the processing of an utterance is a CONTEXTUAL IMPLICATION. This is a conclusion deducible from the conjunction of the linguistic input and the context. In general terms, the greater the positive cognitive effect is, the greater the relevance of the stimulus.

However, construing relevance just in terms of positive cognitive effects does not suffice. In principle, there is no limit as to how many and what sort of contextual implications an individual may draw; there is nothing that tells the hearer which of the many potentially relevant contextual implications is *the most* relevant. What counterbalances this undesirable result is the importance of PROCESSING EFFORT, which is the required use of perception, memory and inference, in establishing the most relevant conclusion. Generally, the less the processing effort is, the greater the relevance of the stimulus. For Relevance Theory thus relevance is calculated in terms of positive cognitive effects and processing effort. Consider the following example, taken from Wilson & Sperber (1993).

(1) Peter took out his key and opened the door.

In (1), the fact that the speaker conjoined two pieces of information (that Peter took out his key and that he opened the door) suggests that the speaker intends these two pieces of information to be processed together (i.e. as relevant to each other). Wilson & Sperber (1993) argue that because encyclopaedic knowledge that to open a door one normally uses a key is easily accessible, the interpretation that Peter used the key he took out (referred to in the first verb phrase) to open the door (referred to in the second verb phrase) comes readily to mind. This interpretation yields adequate cognitive effects (i.e. it explains why Peter took out his keys or how he opened the door) with the least cognitive effort (because the knowledge of how possessing keys is relevant to opening a door is easily accessible) and is thus taken by the hearer to be the intended speaker meaning.

In the next section, I discuss problems in RT's distinction between linguistic and real kinds of semantic content.

2.1.1 *Relevance Theory's linguistic-real distinction*

Parallel to the distinction between linguistic semantics and real semantics is RT's distinction between LEXICAL and AD HOC[2] concepts. Lexical concepts are elements of the logical form decoded by the language module, whereas ad hoc concepts are elements of the pragmatically inferable thought. By definition, the content of lexical concepts (just like the content of the logical form which is composed by lexical concepts) does not have truth-theoretic properties. In contrast, the content of ad hoc concepts (just like the content of the thought/proposition which is composed by ad hoc concepts) has truth-theoretic properties. Thus, just as the decoded logical form underdetermines the communicated thought/proposition at the sentence level, so the decoded lexical concept underdetermines the communicated ad hoc concept at the word level. Carston (2010: 250) suggests that the most plausible assumption is that both lexical and ad hoc concepts are atomic in the sense of Fodor (1998).[3]

In the spirit of RT's underdeterminacy thesis, Carston (e.g. 1998: 2, 2002: 360) entertains the idea that words encode 'concept schemas' or 'pointers' to a conceptual space. As such, they are 'templates' for the construction of a fully propositional conceptual structure. On this proposal, schematic content of concepts encoded by

[2] The notion of ad hoc memory categories was first proposed by Barsalou (1983), for whom ad hoc categories, unlike common categories (like *birds*, *furniture*), are not well established in memory.
[3] I return to this issue in 2.1.3.1.

linguistic expressions is juxtaposed with the real (i.e. truth-theoretic) nature of the ad hoc concepts and thoughts/propositions they compose.

However, this radical understanding of linguistic underdeterminacy – in terms of truth-theoretic value or lack of it – has been consistently undermined in RT. Already in (2002: 362), Carston allows that at least some lexical concepts – for example, natural kind concepts like CAT – are 'full-fledged'. By this, Carston means that some lexical concepts have truth-theoretic properties. This already undermines the distinction between two kinds of semantic content.

In fact, this problem goes deeper. Burton-Roberts (2005) observes that Carston (2002) consistently qualifies her claim that sentences do not encode propositions, i.e. truth-theoretic objects, by 'seldom if ever' or 'almost never'. He also notes that Sperber & Wilson (1986/95) allow that the decoded logical form enters into logical relations of contradiction and therefore implication. Consider example (2) below (discussed in Burton-Roberts 2005: 395-396).

(2) She carried it.

According to Sperber & Wilson (1986/95), the logical form of sentence (2) (i.e. SOME FEMALE ENTITY AT SOME POINT IN THE PAST CARRIED SOMETHING), enters into the relation of contradiction with NO ONE EVER CARRIED ANYTHING and thus into the relation of implication with SOMEONE AT SOME POINT CARRIED SOMETHING. The problem with this, Burton-Roberts (2005) argues, is that only *propositional*, i.e. truth-theoretic, forms can ever enter into logical relations. In other words, by allowing that the decoded logical form enters into logical relations of entailment and contradiction, Sperber & Wilson must allow that the decoded logical form has truth-theoretic properties. This leads Burton-Roberts (*ibid*.) to argue that RT's concept schemas, which are elements of the logical form, actually are *general concepts*.

Burton-Roberts' argument is further supported by Carston's (2010: 235) observation that a 'hearer of 'She's happy' who retrieves just the **proposition** SOME FEMALE PERSON IS HAPPY has not fully grasped the proposition expressed' [my emphasis]. What is grasped – i.e. the linguistically encoded, 'existentially closed' formula – is given the status of a proposition by Carston, contrary to RT's underdeterminacy thesis.

Burton-Roberts' point is also strengthened by various claims in Carston (2010), in particular by Carston's (2010: 242) claim that inferred ad hoc concepts 'may be more specific or more general than the encoded concepts'. The problem with it is that if

concepts can be compared for specificity, they must have the same kind of semantics. Indeed, Carston (2010: 245) makes an unambiguous statement when she writes that (full-fledged) lexical concepts are 'basic element[s] of the language of thought'.

This shows that RT's distinction between ad hoc concepts and lexical concepts in terms of truth-theoretic value or lack of it has actually collapsed. This leads to several problems for RT. The first is that if truth theoretic value or lack of it is no longer defining of the linguistic-real distinction, it is not clear how RT still subscribes to the linguistic underdeterminacy thesis. The second problem is that if some lexical concepts are full-fledged, it is not clear that lexical concepts can actually provide linguistic 'evidence' for the speaker-intended interpretation. I explain this problem in 2.1.2. The third problem is that with the truth-theoretic versus non-truth-theoretic distinction undermined, there is no longer a principled distinction between RT's lexical concepts and ad hoc concepts. I discuss this problem in 2.1.3.

2.1.2 *The problem of linguistic 'evidence'*

As Carston (2002: 365) puts it, encoded linguistic semantics (henceforth 'encoded semantics' or 'linguistic semantics') provides 'evidence (often rich and detailed evidence, but never a complete encoding, never a proof) of the thoughts being communicated'. Wedgewood (2007: 666) phrases it in terms of shared content: 'RT notion of encoded meaning […] provides the logically necessary level of some infallibly shared content'.

There are two ways in which linguistic semantics is supposed to be shared in RT. The first pertains to context-independence – encoded meaning is shared across contexts. I refer to it as cross-context shareability. The second refers to shareability among speakers – in RT linguistic meaning is (widely) shared amongst members of the same speech community. I refer to it as cross-speaker shareability.

Cross-context shareability allows that the (shared) encoded content be deterministically (i.e. necessarily, always) decoded by the linguistic system. Without the assumption of cross-context shareability, positing the process of deterministic decoding would be implausible. As for cross-speaker shareability, it is supposed to help explain how mutual understanding between interlocutors, against their backgrounds of holistic, idiosyncratic beliefs, is possible. Explaining mutual understanding is a huge task and encoded semantics is supposed to aid it by providing a (relatively) stable mutual core of meaning shared by interlocutors in communication. Thus, the theoretical significance of cross-speaker shareability falls on the assumption that it is a reliable – because shared

by (almost) all speakers of a given language – constraint on pragmatic search for the speaker intended meaning. It is in this sense that encoded semantics functions as the linguistic 'evidence'.

However, I argue that neither concept schemas nor full-fledged lexical concepts can (or are necessary to) provide evidence for intended speaker meaning.

2.1.2.1 *Concept schemas*

Recanati (1998: 630) makes reference to Hintzman's (1986) multiple-trace theory of memory, which predicts that '[w]ords, as expression-types, do not have "meanings", over and above the collection of token-experiences they are associated with. The only meaning which words have is that which emerges in context'. Carston (2002: 375) also makes reference to Hintzman's radical contextualism, but only to reject it in favour of a more 'conservative' view 'on which words do encode something, albeit something very schematic, which simply sends the system off to a particular region in long-term memory'. The schemas that Carston talks about are, as discussed above, one proposed type of encoded semantics. But there are problems.

Firstly, there are fundamental questions about the cognitive function of abstract/schematic lexical concepts and how they are acquired. Cross-context shareability demands that encoded semantics is a context-invariant, stable 'meaning' of a word, which underlies all uses of a particular word. For example, for a word like *take*, encoded semantics has to underlie such uses as *take money from someone=remove something without permission, take 4 from 15=subtract, take credit cards=accept, take the bus=travel somewhere using a particular form of transport, take linguistics at university=study*, etc.[4] This brief survey of different uses of the word *take* shows that for encoded semantics to underlie all uses of a given word it has to be very abstract or schematic.[5]

The abstract nature of lexical semantics is actually the first problem that we encounter: Carston (2002: 364-365) herself concedes that schemas are so abstract that they are (almost) never employed in thinking. She admits that it is quite difficult to see how they are ever acquired if they do not play a role in thoughts and how they become the meanings of lexical expression types. The first concern reveals an interesting thing:

[4] These uses were taken(!) from Cambridge Dictionaries Online.
[5] Young (2005) also criticises the notion of a concept schema. Young (2006) illustrates the 'abstractness' problem in particular with a thorough cross-linguistic comparison of the form-concept relations for the English word *open*.

since it is only *concepts*, not concept 'templates', that are entertainable in (i.e. play a role in, are elements of) thought, it seems that even for Carston, schemas should preferably be *general concepts* if they are ever to be acquired. Despite these problems, Carston assumes the following.

> **Quote 1**: There must be some process of abstraction, or extraction, from the particular concepts associated with the phonological form /open/ to the more general 'meaning', which then functions as a gateway both to the existing concepts of opening and to the materials needed to make new OPEN* concepts which may arise in the understanding of subsequent utterances. (Carston 2002: 364)

However, as noted by Burton-Roberts (2007), there is a problem with the acquisition of concept schemas. Linguistic semantics is supposed to be acquired in experience and, in Carston's terms, it provides a 'gateway' to the understanding of the meaning of a word in context. It guides the hearer to the intended context-dependent meaning of a word. However, linguistic semantics is abstracted from, and hence presupposes prior understanding of utterances. In other words, the acquisition of linguistic semantics is post hoc. But if the acquisition of linguistic semantics is post hoc and presupposes prior understanding of utterances, then, in principle, decoding linguistic semantics cannot be a necessary step in utterance interpretation.

In a nutshell, the problem is that positing concept schemas as encoded semantics amounts to positing constructs which are either (a) impossible to acquire, since it is impossible to employ them in thinking or (b) possible to acquire, but not necessary in understanding an utterance. If (a), concept schemas cannot function as cross-context shared content. If (b), their post hoc acquisition undermines the condition that they be necessarily (i.e. deterministically) accessed.

I will later argue that concept schemas (by which I mean general concepts) – even *when* they exist – are not linguistically decoded, but pragmatically inferred. In the meantime, I look at full-fledged lexical concepts.

2.1.2.2 *Full-fledged lexical concepts*

Carston (2002, 2010) asserts that RT follows Fodor (1998, 2008) in assuming that conceptual correlates of words are atomic, i.e. not compositionally constituted. As discussed in chapter 1, Fodor's atomism equals, roughly, the assumption that all *apparent* logical relations and other information which is associated with a given concept do not constitute the content of this concept. The content of a concept is

constituted by a causal relation a concept bears to things in the mind-external world. For example, the content of an atomic concept like CAT is constituted by its relation to properties that cats in the mind-external world possess.[6]

Likewise, in RT, every word is taken to encode an atomic concept. For example, the word *cat* encodes an atomic concept CAT. By assumption, all the bits of information that in a cogniser's memory may be associated with the word *cat* – for instance, that cats are animals, that they are quadrupled felines, that they meow, have whiskers, that my neighbour's cat is called Smudge, that Smudge scratched my neighbour last week, etc. – do not constitute the content of this concept. In other words, all associated (logical and encyclopaedic) information is, by assumption, *non-constitutively* attached to a given concept.

That said, Carston (e.g. 2002: 214, fn 31) emphasises the following difference between Fodor's account and RT. Fodor, she argues, endorses a code model of communication whereby the concepts communicated are the same as the concepts en/decoded. In RT, however, there is only a partial mapping between words and concepts since large part of the conceptual repertoire is not lexicalised – i.e. large part of the conceptual repertoire is populated by ad hoc concepts.[7]

Importantly, Carston (2010: 247) argues that ad hoc concepts are contextually derived from such a (non-content constitutive) pool of logical and individualistic/holistic information. This means that ad hoc concepts are constructed through personal inference and shows that Carston's LOT (unlike Fodor's LOT) is populated by pragmatically (i.e. individualistically/holistically) derived concepts. Thus, Carston's conceptual content (unlike Fodor's content) is *not* free from individualism/holism. This gives rise to another problem for Carston's argument that lexical concepts provide 'evidence' for the speaker intended meaning.

As mentioned at the outset of this section, full-fledged lexical concepts are ex-ad hoc concepts which underwent lexicalisation and became encoded as linguistic meaning (Carston 2010: 244). The problem with linguistic 'evidence' in the case of full-fledged lexical concepts is that if full-fledged lexical concepts are ex-ad hoc concepts, then they

[6] Fodor writes:

> […] concepts have both referents and a congeries of beliefs (etc.) in which they are embedded. It's just that, whereas the former has to do with the content of the concept, the latter has to do with its (e.g. inferential) role in mental processes. The distinction between these is independently motivated; *content is what composes*, and inferential roles and the like do not. (Fodor 2008: 87-88)

[7] A similar claim is made by Lalumera (2009).

have been constructed through personal inference. If that is the case, full-fledged lexical concepts cannot, in principle, constitute adequate cross-speaker 'evidence'.

It may be argued that the problems I have discussed in this section are not really problems for RT. After all, Carston (2002: 365, 19-20) makes it clear that linguistic semantics is 'never a proof' and that it is only 'widely' shared among members of the same speech community. However, what I have discussed here does raise the question of whether it is true that (A) or (B):

(A) Hearer H converges (to a sufficient degree) on what speaker S communicates because S and H share linguistic semantic content which is encoded and therefore deterministically decoded in utterance interpretation and which therefore constrains utterance interpretation
(B) Hearer H converges (to a sufficient degree) on what speaker S communicates *despite* the non-existence of such shared linguistic semantic content

In this section, I have argued that linguistic semantic content is either impossible to acquire or if possible, then either unnecessary in utterance interpretation (because post hoc) or not cross-speaker shared (because constructed through personal inference). The status quo is this. RT's distinction between real and linguistic semantics can no longer be made in terms of truth-theoretic value or lack of it (as argued in 2.1.1). Neither can it be made in terms of the distinction between holistic/individualistic content versus shared content, respectively (as argued in this section). However, there is one more way of defining the distinction between real and linguistic semantics – namely, in terms of the distinction between the cognitive processes of decoding versus inferring (as argued by Carston 2002: 11). This distinction, however, relies on RT's conservative assumption that the cognitive process of deterministic decoding of specifically linguistic semantics is a necessary step in utterance interpretation. In what follows, I argue that the process of decoding is actually redundant.

2.1.3 *Redundancy of decoding*

In this section, I argue that decoding is redundant in cases of loose use, cases of so-called 'concept narrowing'[8] and where the communicated concept is the same as the purportedly encoded concept, i.e. in pretty much all cases. My argument holds

[8] In RT (Carston 1996, 2002), concept narrowing and broadening are both accounted for by the single process of concept adjustment/modulation. I use the term "narrowing" for expository purposes.

regardless of whether concepts are thought to be atomic á la Fodor (1998, 2008) or compositionally constituted.

2.1.3.1 *Atomic content or compositional content?*

As discussed earlier, Carston (2002, 2010) asserts that RT follows Fodor (1998, 2008) in assuming that conceptual correlates of words are atomic, i.e. not compositionally constituted. In practice, however, conceptual correlates of words are often decomposed into analytic properties in RT. Burton-Roberts (2005: 399) observes that Carston (2002) sometimes decomposes lexical concepts; for example, RAW is decomposed into NOT + COOKED. Relatedly, consider example (3) below (taken from Carston 2002: 27).

(3) The steak is *raw*.

In RT, the interpretation of the utterance of (3) proceeds is two steps. First, the linguistic system deterministically decodes the lexical concept RAW[9], which we can understand as not cooked (though, by RT's assumption, this concept is not compositionally constituted). From such linguistically decoded input, the pragmatics module derives the ad hoc concept RAW*, which is the concept communicated in this particular context. Here we can understand the concept RAW* as 'cooked but not long enough', but again, by assumption, this concept is atomic.

The use of the word *raw* in (3) exemplifies the so-called 'loose use' (/'concept broadening'), the term which subsumes metaphorical uses, hyperbolic uses or cases in between. In such cases, the encoded lexical concept is 'relaxed' and a more general ad hoc concept is pragmatically derived on its basis. Now, Carston (2002: 28) observes that the idea that the ad hoc concept is pragmatically derived from a lexical concept is controversial in cases like (3). This is because, the ad hoc concept RAW* is not logically related to the lexical concept RAW. Carston's point is that RAW* cannot be derived from RAW because RAW* does not 'analytically imply [uncookedness]'.

There are two observations to be made here. First, if RAW* cannot be derived from the lexical concept RAW, it follows that the lexical concept is only decoded to be 'dropped and replaced' with a logically unrelated ad hoc concept. This is acknowledged by Carston (2002: 28, 341, 358), who herself talks of the 'drop and replace' scenario in the case of loose use. The problem is that if the lexical concept is only decoded to be 'dropped and replaced', it is *needlessly* decoded.

[9] By convention, in RT lexical concepts are represented by using capitalised letters, whereas ad hoc concepts are accompanied by asterisks (*).

The second observation is this. Carston (Carston 2002: 28) argues that the idea that ad hoc concepts are pragmatically derived from lexical concepts is controversial in the case of loose use like (3) precisely because the communicated concept does not 'analytically imply' the lexical concept. From this we can deduce that there is nothing controversial in an ad hoc concept being pragmatically derived from a lexical concept if the ad hoc concept 'analytically implies' the lexical concept. The problem for Carston is that if the *ad hoc* concept 'analytically implies' the lexical concept, then it is, in part, compositionally constituted by it. Otherwise, it would not 'analytically imply' it.

This further shows that in RT there is an instability about whether concepts are atomic or compositional. A bigger problem for RT, however, is that the cognitive process of decoding of lexical concepts is redundant regardless of whether concepts are taken to be atomic or compositionally constituted. This is what I turn to now.

2.1.3.2 *The redundancy problem*

In RT (e.g. Carston 2002, 2010), the process involved in the derivation of ad hoc concepts is the pragmatic process of concept MODULATION or ADJUSTMENT. For example, in the case of loose use the lexical concept is modulated/adjusted so as to derive a more general concept. As discussed, it is acknowledged in RT that in cases of loose use the idea of pragmatic derivation of an ad hoc concept from a lexical concept is controversial.

Let us now apply RT's 'decoding plus inferring' model to the analysis of (4) and (5) below (taken from a random BNC[10] search).

(4) When Mr Cerezo, before his election, stood up against successive military regimes the rightists tried to *kill* him.
(5) […] one of the mechanisms by which phagocytes *kill* intercellular organisms is the production of reactive oxygen intermediates […].

On the assumption that RT's model does apply here, the interpretation of the word *kill* in (4) and (5) proceeds in two steps. First, a lexical concept KILL is deterministically decoded (as part of the logical form) in both (4) and (5). The second step involves the pragmatic process of modulation/adjustment of the lexical concept to a more specific ad hoc concept – KILL* in (4) and KILL** in (5). In cases of concept narrowing, the idea

[10] Data cited herein have been extracted from the British National Corpus Online service, managed by Oxford University Computing Services on behalf of the BNC Consortium. All rights in the texts cited are reserved.

of pragmatic derivation of the ad hoc concept on the basis of the lexical concept does not seem controversial because the pragmatically communicated ad hoc concepts do entail the encoded lexical concept – i.e. the concept KILL in both examples.

However, Burton-Roberts (2007) and Groefsema (2007) argue that the very notion of concept adjustment or modulation is incompatible with the notion of *atomic* conceptual content; if lexical concepts really are atomic, then they cannot in principle be adjusted or modulated. Now, when RT theorists argue that ad hoc concepts are formed from lexical concepts, what they mean is that ad hoc concepts are formed from logical and encyclopaedic entries attached to lexical concepts (as discussed earlier). Since, by definition, lexical concepts are atomic and since logical and encyclopaedic entries are in RT non-content-constitutive, i.e. they are not part of the lexical concept, ad hoc concept formation cannot amount to the modulation or any other sort of manipulation of the lexical concept. It can only amount to the *replacement* of the posited lexical concept with another one.

The problem for RT is that if concepts are – as Carston argues – atomic, Burton-Roberts'/Groefsema's criticism applies to *all* cases of utterance interpretation. If that is the case, lexical concepts are always decoded only to be dropped and replaced with ad hoc concepts, i.e. they are *needlessly* decoded.

However, given RT's instability regarding whether concepts are atomic or not, Burton-Roberts'/Groefsema's criticism might be dismissed on the grounds that it is irrelevant because RT actually allows that concepts are compositionally constituted. I argue that the cognitive process of deterministic decoding of lexical concepts is redundant even if concepts are taken to be compositionally constituted.

With loose uses like (3) (*'The steak is raw*') the redundancy of decoding seems quite clear. If concepts are compositionally constituted, then the concept encoded by the word *raw* is that of not being cooked, whereas the concept communicated by the use of this word in the context of (3) is that of being cooked but not enough. In line with Carston's own argument, the communicated concept cannot be pragmatically derived from the encoded lexical concept because, if something has been cooked even for a short period of time, then it is not the case that it has not been cooked. Thus, even on a compositional analysis, the concept of not being cooked is needlessly decoded only to be dropped and replaced with an ad hoc concept.

The situation with concept narrowing seems less straightforward but is not less problematic for RT. Consider (4) and (5) again. Given the availability of a compositional analysis in RT, RT could argue that in (4) and (5) the communicated ad

hoc concepts are pragmatically derived from lexical concepts because they entail the lexical concept of causing death.[11] In (4) the concept of causing death by an intentionally acting human agent entails the concept of causing death, and so does the concept of causing death by a chemical reaction at a cellular level in (5). Translating this into RT's metalanguage, the fact that the concepts KILL* and KILL** entail KILL suggests that both can be seen as pragmatically derived from the lexical concept KILL. Prima facie, on a compositional analysis, instances of concept narrowing make a case for the existence of deterministically decodable linguistic semantic content which is then pragmatically modulated to yield a more specific concept.

However, the problem for RT is that if KILL* is compositionally constituted by the primitive concepts that KILL is composed by, the primitive concepts that compose KILL have to be pragmatically derived in deriving KILL*. This means that the structured concept KILL must be pragmatically derived (as compositional part of KILL*), after it has been deterministically decoded. In other words, in RT the concept of causing death has to be activated twice by two distinct cognitive processes. Arguably, this makes the activation by the process of decoding redundant (especially in the light of problems discussed in 2.1.1 and 2.1.2). The point is that if the primitive concepts that compositionally constitute a complex general concept of causing death have to be pragmatically derived, it means that they *can* be pragmatically derived; and if they can be pragmatically derived, then they do not have to be deterministically decoded.

Consider now a case where the communicated concept is not assumed to be more general or more specific than the purported lexical concept, but rather the same. In RT, concept adjustment is a 'free' pragmatic process. This means that there is nothing in the linguistic form such as to indicate that it must be carried out (Carston 2010: 242-243). Carston (*ibid*.) illustrates the 'free' nature of concept adjustment with the use of the word *dance* in the following example.

(6) Children in most cultures *dance* spontaneously.

According to Carston (*ibid*.), the use of *dance* in (6) communicates a general concept of dancing (i.e. a concept devoid of any particularities such as specific movements). As

[11] Two brief comments are in order here. Fodor (1970) objects to the idea that the word *kill* can be compositionally defined in terms of *cause to die*. The first point is that I assume that RT could compositionally analyse the word *kill* in (4) and (5) because compositional analysis is allowed in RT (as discussed in 2.1.3.1). The second point is that Fodor's (1998, 2008) claim that the content of conceptual correlates of words is atomic and referential cannot be maintained for the reasons discussed in chapter 1.

dance also happens to encode this general concept, there is no need for 'adjustment' here.

The problem for RT is that in cases like (6), the linguistic system first deterministically decodes the encoded concept and then, on the basis of contextually available information, the pragmatics module infers that this and no other concept is being communicated. The point is that concept adjustment/modulation may not be necessary, but perceiving that no further adjustment is required is itself a pragmatic inference (from context). Thus, in cases like (6) one and the same concept is activated twice by two distinct processes, decoding and inferring. Because the relevant concept has to be pragmatically inferred anyway, and because there are so many problems with linguistic semantics, I argue that it is the process of decoding which is redundant.

2.1.4 *Interim summary*

The point I have been making so far is this. The notion of linguistic semantics is posited in aid of saying that there is some constant semantic property of a word which constrains the search for speaker-intended meaning. But there are compelling arguments that linguistic semantics is a poorly defined construct (section 2.1.1), which is impossible to acquire (section 2.1.2) or if possible to acquire, it is unnecessary in utterance interpretation because post hoc and, in principle, not shared (section 2.1.2). Furthermore, the process of deterministic decoding of such content is redundant regardless of the strength of RT's commitment to Fodor's (1998, 2008) atomism (2.1.3). In cases of loose use like (3) decoding is redundant; if the ad hoc concept is logically unrelated to the lexical concept, the lexical concept cannot constrain the search for the communicated ad hoc concept – a point acknowledged by Carston (2002: 28). In cases of the so-called 'concept narrowing' and where the purported lexical concept is the same as the communicated concept, decoding is redundant because the concept that is decoded has to be pragmatically inferred anyway.

As mentioned, RT's insistence on the existence of linguistic semantics appears to rest on cases where the communicated concept is assumed to be more specific that the lexical concept from which it is allegedly derived. In cases like this, *if* linguistic semantics existed it *could* constrain the search for ad hoc concepts. However, the intuition that the ad hoc concept is logically related to a lexical concept arises only when the a priori assumption (A) is made. In other words, an ad hoc concept like KILL* can be seen as logically related to a **lexical** concept KILL only if one assumes that a **lexical** concept KILL exists.

Now, it is undeniable that, for example, a specific concept of causing death by an intentionally acting human agent (as in (4)) is logically related to a more general concept of causing death. However, we cannot and should not (in the light of the problems with the notion of linguistic semantics discussed so far) extrapolate from this fact to the existence of specifically linguistic semantic level. Similarly, the undeniable fact that the word *kill* in (4) and (5) is used to communicate partially overlapping concepts, should not lead us to conclude that this overlap (the intersection) constitutes evidence for the existence of linguistic semantics.

I argue that there is no linguistic semantics (and no lexical concepts) and that utterance interpretation is a wholly pragmatic, inferential enterprise. But if I want to argue that, I need to face two lines of criticism; (a) If words do not have meaning in virtue of encoding linguistic semantics as their constitutive property, then how do they mean?; (b) If there is no linguistic semantics constraining word use and interpretation, then how do we ever communicate successfully? In the next section, I argue that the Representational Hypothesis' (e.g. Burton-Roberts 2012) definition of *meaning-as-relation* invalidates criticism (a) and that Hintzman's (1986) multiple-trace theory of memory and information retrieval invalidates (b).

2.2 Utterance interpretation without linguistic semantics

In this section, I first introduce the notion of words as linguistic 'pointers' to conceptual space and discuss the incompatibility of the 'pointer' metaphor with the double-interface view of language. Along with the Representational Hypothesis, I reject the double-interface view and adopt the semiotic (in the sense of Peirce) notion of a pointer. I introduce the Representational Hypothesis and endorse its definition of meaning-as-relation (to semantic content) as well as its stance on the nature of semantics. Finally, I argue that the mechanics of the representational account can be implemented in terms of Hintzman's (e.g. 1986) multiple-trace theory of memory. I discuss some advantages of a wholly pragmatic, inferential account of utterance interpretation over a two-step (i.e. decoding and inferring) model like that of Relevance Theory's.

2.2.1 *Saussurean/Chomskyan linguistics*

As discussed earlier, RT's distinction between deterministically (i.e. necessarily, always) decoded lexical concepts and ad hoc concepts is motivated by the traditional

double-interface assumption whereby linguistic signs have meaning in virtue of being partly constituted by semantic properties. Because in RT it is in virtue of the decoding-inferring distinction that linguistic signs are argued to have some specifically linguistic meaning, the decoding-inferring distinction is also dictated by the double-interface view.

However, Burton-Roberts (e.g. 2009, 2011) argues that the double-interface idea itself is problematic and redundant. The next section places the problems with linguistic semantics and lexical concepts in the context of the double-interface view of language.

2.2.1.1 *Linguistic semantics and the notion of a pointer*

Fodor's lexical-conceptual isomorphism (discussed in 1.1) is a very strong version of the claim that natural languages 'inherit' their semantics from the semantics of thought (e.g. Fodor 1998: 9). However, as argued in 1.1.2, even a brief consideration of cross-linguistic data provides strong evidence that Fodor's lexical-conceptual isomorphism is untenable (e.g. the existence of referential equivalents such as English *shallow* and French *peu profond*). Furthermore, isomorphism is incompatible with the following claims of Fodor's:

> **Quote 2**: English inherits its semantics from the contents of the beliefs, desires, intentions, and so forth that it's used to express, as per Grice and his followers. Or, if you prefer (as I think, on balance, I do), *English has no semantics*. Learning English isn't learning a theory about what its sentences mean, it's learning how to associate its sentences with the corresponding thoughts. (Fodor 1998: 9)
>
> **Quote 3**: […] only thought *has* content, strictly speaking. (Fodor 2001: 2)
>
> **Quote 4**: Quite possibly English has no semantics, some appearances to the contrary notwithstanding. (Fodor 2008: 99)

Strictly speaking, the claim that natural language expressions 'inherit' their semantics from the semantics of thought is incompatible with the claim that natural language (e.g. English) expressions have no semantics at all. One has to make a choice between these two contrary claims. In the light of problems with Fodor's lexical-conceptual isomorphism (which supply an argument against the 'inheritance' account), it becomes necessary to consider carefully the assumption that linguistic expressions have no semantic properties.

It is also important to notice that the same tension – regarding the ascription of semantic properties to linguistic expressions – is present in Relevance Theory (e.g.

Sperber & Wilson 1998, Carston 2002: 360). Despite positing linguistic semantics as mediating between linguistic forms and ad hoc concepts, Carston (e.g. 2002: 330, 360) occasionally speaks of linguistic encodings as 'pointers' to conceptual space. This is interesting because in the case of 'pointers' we are not dealing with linguistic semantic content, but with pragmatic constraints on the processing of an utterance. This is made clear in Carston's discussion of Blakemore's account of discourse markers, quoted below.

> **Quote 5**: Blakemore found that **rather than making conceptual contributions** to some other level (implicatures), they [discourse markers] appear to function more like **filters on, or pointers to, the pragmatic inferences the hearer is to carry out**. (Carston 2002: 160-161, my emphasis)

If pointers do not make 'conceptual contributions' to utterance interpretation, it means that they are not constituted by semantic/conceptual properties. In fact, even though for Blakemore (1992) discourse markers *encode* procedural semantics (but not conceptual semantics), it is not clear that Carston uses the pointer metaphor to talk of any sort of encoding. According to Carston (2002: 360-363), it may be that a word (p.360) or even the lexical form (p.361) 'points to a conceptual region or maps to an address (or node, or gateway, or whatever) in memory' and that 'this pointing or mapping provides access to certain bundles of information'. But to entertain the idea that it is the lexical form which 'points' to specific regions in memory is to entertain the idea that linguistic expressions 'have meaning' without having linguistic semantics as their constitutive property. This is to deny the rationale for positing linguistic semantics. As was the case with Fodor, the problems with Relevance Theory's notion of linguistic semantics suggest that the notion of a linguistic expression as a pointer needs to be seriously considered. Interestingly, Carston (2002: 363) thinks that this choice 'doesn't make much difference' and she in fact continues to attribute conceptual properties to words.

Significantly, Chomsky (2003: 303) too has expressed a view of linguistic expressions as 'pointers' to conceptual space. Occasionally, Chomsky (e.g. 2000a: 61) has also argued that in language acquisition children 'label' already available concepts, 'with much or all of their intricacy and structure predetermined'. Whereas the notion of words as 'pointers to' or 'labels for' concepts seems appealing (especially in the light of the problems with Fodor's account and Relevance Theory), there does seem to be a problem with its application on Chomskyan account.

As observed by Burton-Roberts (2007), pointers or labels, in principle, do not have the properties of what they point to or what they label. When I point with my

finger at a tree, my finger does not partake the properties of that tree. Similarly, a label on a bottle of wine does not have the properties of that bottle, nor of the wine. A price label on a supermarket shelf is not the price itself, it just points at (communicates) what the price is. It does not have the properties of the price either; the properties that a price label has can be: made of paper, made of plastic, made of printer or ball-pen ink, perhaps even self-adhesive. Prices, on the other hand, are clearly not made of plastic, etc., nor can they ever be self-adhesive. To have a pointer or a label, then, is to have something other than what they point to or label.

The problem, already signposted in 1.2.4, is that if words are pointers to conceptual space, then it is at least conceptually unnecessary to attribute to them the properties that they point to, i.e. semantic/conceptual properties. The view of words as pointers is thus incompatible with the double-interface view of linguistic expressions as constituted by phonological and semantic properties (<PNON, SEM>). One has to make a choice between them.

I have already suggested that the problems with Fodor's lexical-conceptual isomorphism and the problems with Relevance Theory's notions of linguistic semantics and lexical concepts require that the view of words as pointers be carefully considered. Chomsky's double-interface view of language is relevant here because, as argued earlier, it is the double-interface claim itself that imposes the claim about the existence of encoded semantics. Therefore, if it can be shown that the double-interface idea itself is problematic and redundant, we have compelling arguments to reject it along with its implications.

In fact, Burton-Roberts (e.g. 2007, 2009, 2011, 2012; Burton-Roberts & Poole 2006a, 2006b) argues that the double-interface view of language is:

- unnecessary on semiotic grounds
- insufficient on conceptual grounds, and
- impossible on sortal grounds (i.e. it constitutes a category mistake)

I will now discuss Burton-Roberts' arguments against the double-interface view of language.

2.2.1.2 *Problems with the double-interface view of language*

As observed by Burton-Roberts (2011), the Chomskyan double-interface view of language is closely tied up with the traditional, Saussurean, view of a linguistic sign.

For Saussure, a linguistic sign is composed of two parts: a sound image (phonology) on the sound side and a concept (semantics) on the meaning side. For Chomsky:

> **Quote 6**: […] there are sensorimotor systems that access one aspect of an expression and there are conceptual-intentional systems that access another aspect of an expression, which means that **an expression has to have two kinds of symbolic objects as its parts**. These objects can be regarded as a kind of an interface between the language faculty and other systems of the mind-brain. (Chomsky 2000b: 9, my emphasis)

The relation that a sound image and a concept have with respect to one another and with respect to the sign that they together make up is a part-part (i.e. mereological) relation. The nature of the relation between the sound image and the concept is arbitrary. This is why the concept CAT, for example, is 'expressed' by different words in different languages – *cat* in English, *gato* in Spanish and *kissa* in Finnish.

Burton-Roberts (2012) points out that the mereological (part-part) idea is conceptually *in*sufficient to account for meaning; there is nothing in the mereological relation itself between a sound image and a concept that distinguishes it from the mereological relation between a bottle and a cork or a table leg and a table top. Crucially, the mereological relation does not explain why a concept (i.e. semantics) should be related to a sound image – it has nothing to say about meaning. Therefore, Burton-Roberts (*ibid*.) argues, it was necessary for Saussure to supplement the mereological idea with the semiotic (signifier-signified) idea; in a semiotic relation, the sound image is a signifier and a concept is a signified.

Burton-Roberts also argues that these two ways – mereological and semiotic – of thinking about the relation between the sound image and the concept are incompatible. Whereas the mereological, part-part relation is a symmetric (two-way) relation (the sound image is as much part of the sign as the concept is; they are co-parts), the semiotic relation is antisymmetric and it goes from the sound image to the concept.

Relatedly, Burton-Roberts (2009, 2011) argues that the mereological idea is *im*possible for several reasons. First of all, parts in a mereological relation constitute a separate object – just as table legs and a table top constitute a table, so a sound image and a concept are taken to constitute a linguistic sign. In this sense, Saussurean sign is partly constituted by what it is a sign of, i.e. a concept. However, a semiotic relation is different in that it treats the relation between a sound image and a concept in purely relational terms – there is no separate object constituted by the two relata. It follows

from the semiotic relation then, Burton-Roberts argues, that there is no object constituted by the sound properties and the meaning properties.

Indeed, the mereological idea of a sign is problematic if linguistic signs are to be 'realised' – i.e. produced – in speech. Whereas it is plausible for a sound image to be 'realised' as a sound in speech, it seems inconceivable that a concept or a combination of a sound image and concept could be realised as a sound. For example, it is possible to 'realise' as a sound (i.e. to utter) a phonologically constituted object like [not], but it is impossible to 'realise' as a sound (i.e. to utter) the logical functor of negation (Burton-Roberts 2007).

The reason why it is possible to produce in speech the former but not the latter is their sortally different natures. Whereas phonology (a sound image) is grounded exclusively in phonetics (i.e. it can be interpreted in articulatory-perceptual terms), semantics (a concept) is grounded exclusively in conceptual-intentional terms. It is these different groundings and their sortal (i.e. categorical) incompatibility that make it impossible for the sound image and the concept to be combined into one object. Burton-Roberts (2011) argues that Saussurean arbitrariness follows from the sortal incompatibility of a sound image and a concept – any relation that holds between sortally incompatible properties must be arbitrary (i.e. non-natural).

Now, Chomsky (e.g. 2000b) reconstructs the Saussurean view of a sign by adopting the mereological idea; lexical items are taken to be double-interface syntactic objects constituted by phonological and semantic properties, but no semiotic relation is posited between the two kinds of properties. However, Burton-Roberts (2009, 2011) emphasises that the sortal incompatibility of phonological and semantic properties is actually acknowledged in Chomskyan linguistics by the principle of Full Interpretation. This principle predicts that semantic properties are only interpreted at LF, an interface between UG and the conceptual-intentional system, and phonological properties are interpreted only at PF, an interface between UG and sensorimotor systems. As observed by Burton-Roberts (*ibid*.), double-interface syntactic objects, i.e. objects constituted by phonological and semantic properties are not, however, interpretable at any of the two interfaces. Even though double-interface objects enter the syntactic computation, the two kinds of properties are separated at Spell-Out to satisfy the principle of Full Interpretation. This means that words, or linguistic items, as reconstructed in Chomskyan linguistics, i.e. as objects mereologically constituted by phonological and semantic properties, are not interpretable. The question that Burton-Roberts (*ibid*.) thus asks is why such double-interface objects should be posited if it is acknowledged that

the two properties are sortally incompatible and hence mutually un-interpretable. From a methodological perspective, the question is: why posit double-interface objects only to split them?

The discussed problems with the mereological relation raise an important issue. If words cannot be constituted by both sound and meaning properties, what are words and what is the relation between the two kinds of properties? The answer, as argued by Burton-Roberts (e.g. 2007), lies in the semiotic relation – an issue I turn to now.

2.2.2 *Conventional representation: meaning as relation to semantics*

In the previous section, I discussed Burton-Roberts' criticism of the double-interface idea on the grounds that it is (i) insufficient to account for meaning in language as it needs to be supplemented by the semiotic relation and (ii) impossible on sortal grounds – a fact which is acknowledged in Minimalism by its need to split phonological and semantic properties. A further argument of Burton-Roberts' is that the double-interface idea is actually unnecessary to account for meaning in language. In order to account for meaning in language, the mereological relation has to be supplemented (as it is for Saussure) by an arbitrary, i.e. non-natural, conventional, *semiotic* relation, whereby the sound properties function as a *signifier* and the meaning properties (concepts) function as a *signified*. Burton-Roberts argues that the explanatory power of the semiotic relation renders the anyway problematic mereological relation conceptually unnecessary. In what follows, I concentrate on this, perhaps the most important, criticism of the double-interface idea – the claim that it is unnecessary. I introduce the Representational Hypothesis, a framework which is motivated by the problems with the double-interface tradition and which develops the semiotic idea.

Let us consider first how meaning in general arises. The first thing to observe is that many things other than words are meaningful to us without our wanting to say that they have semantic content. For example, a footprint on the sand 'is meaningful' to Robinson Crusoe; it means to him that someone has been there. It is an indexical sign, i.e. a natural sign which gives rise to meaning independently of (communicative) intention. Even though we take the footprint to be meaningful, we would not like to attribute any semantic content to it. The same goes for symbolic signs, like the one in figure 2.1.

Figure 2.1: National speed limit applies

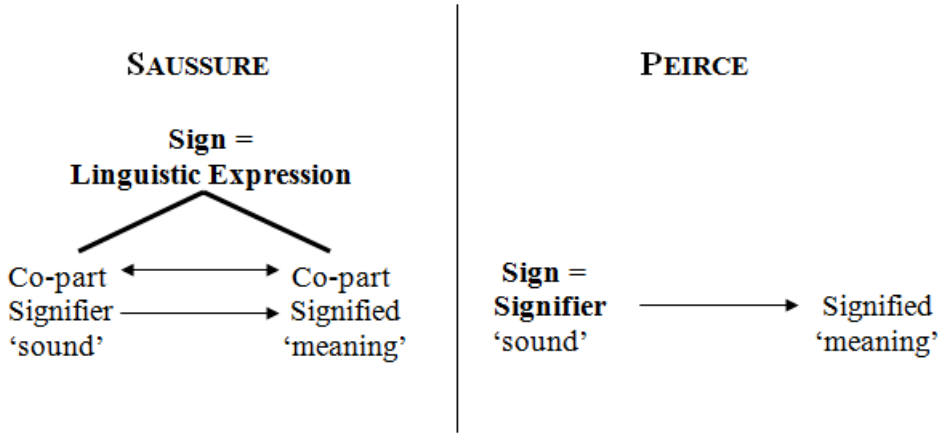
The symbolic sign ‘national speed limit applies’ does not have conceptual/semantic properties. It ‘has meaning’ for a person A because when A is driving a passenger car on a motorway and A sees it placed next to the road, this sign gives rise to the thought in A’s mind that A can speed up to 70 m/h. Meaning here arises in virtue of a semiotic relation between a physical sign with no semantic content and a semantically/conceptually constituted thought.

I opened this chapter with the discussion of the notion of a pointer. This notion is relevant here because the view of words as pointers to conceptual content suggests a view of words as symbolic signs. The Representational Hypothesis (e.g. Burton-Roberts & Poole 2006a, 2006b; Burton-Roberts 2012) is a development of the idea that utterances of linguistic expressions – i.e. sounds utilised by speakers in communicating thoughts – are symbolic signs.

2.2.2.1 *The Representational Hypothesis*

The Representational Hypothesis (henceforth RH) rejects Saussure’s notion of linguistic sign and adopts C.S. Peirce’s notion of linguistic sign. The two notions of linguistic sign are compared in the figure below.

Figure 2.2: Saussure’s and Peirce’s notions of linguistic sign



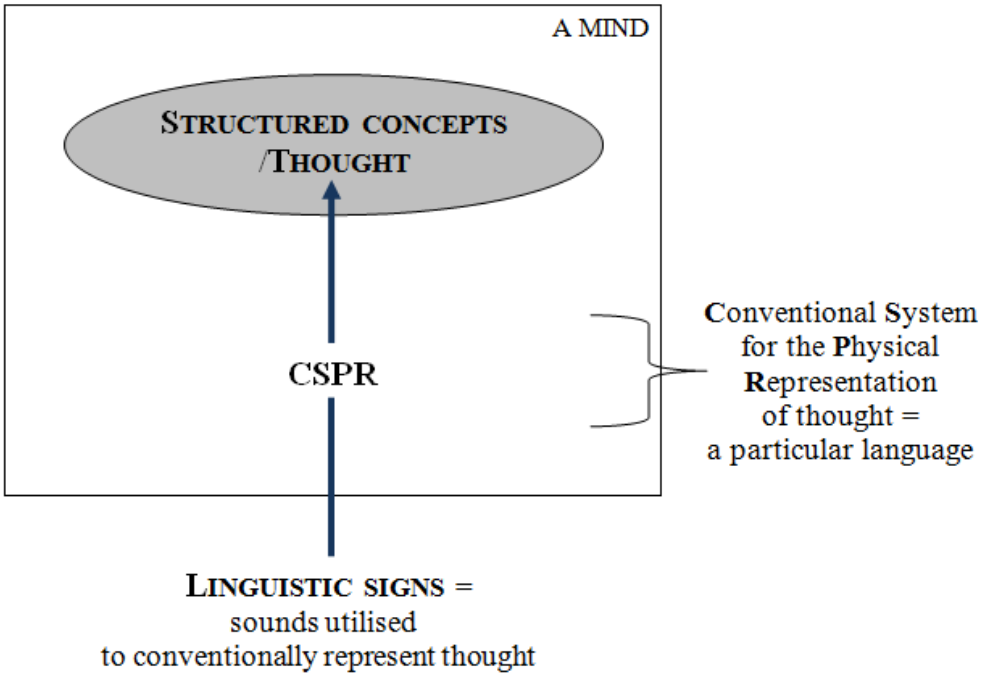
As discussed earlier, for Saussure (and for Chomsky) the linguistic sign is partly constituted by what is a sign of – i.e. semantics. This notion motivates the assumption that linguistic expressions have meaning as their constitutive property (i.e. linguistic semantics) and gives rise to the problems discussed in 2.1 and 2.2.1.2. The RH develops C.S. Peirce's notion of linguistic sign where the signifier and the signified are in a purely semiotic relation and do not constitute (as they do in the Saussurean and Chomskyan traditions) a separate superstructure. For Peirce and for the RH, the linguistic sign is *other* than what it signifies – it is the sound utilised for communication which is the linguistic sign (Burton-Roberts 2007, 2012).

In the RH specifically, sounds of particular languages function as symbolic signs which are used to conventionally REPRESENT internalist conceptual/semantic content (i.e. structured thought). One of the central claims of the RH is that natural languages are Conventional Systems for the Physical Representation of thought (henceforth CSPRs). Significantly, the RH (e.g. Burton-Roberts & Poole 2006b) distinguishes between the object represented (*representatum*) – i.e. structured thought – and the physical phenomena that it is represented by (*representans*). The representatum (*x*) is generated by the Language of Thought (LOT), whereas CSPRs define what counts as a (morpho-phonetically) well formed representans (R(*x*)). The relation between *x* and R(*x*) is that of conventional representation, where $R(x) \neq x$.

In the RH, the representatum (*x*) is innate and invariant across the species[12]. By contrast, how it is represented (defined by particular languages/CSPRs) involves massive variation. It is precisely because the representational relation between R(*x*) and *x* is non-natural and conventional that different CSPRs are constituted by different representational conventions. The following figure is a sketch of the representational architecture of mind.

[12] The representatum (*x*) can only be innate and invariant across the species in the sense that it is generated by innate and invariant LOT (more on this shortly). I will argue in chapter 3 that thought itself (representatum) can only ever be holistic/individualistic.

Figure 2.3: Representational architecture of mind



In the RH, particular languages (i.e. CSPRs) are constituted by representational conventions[13] which define what counts as a well formed linguistic sign/representans (they specify what counts as an appropriate morpho-phonetic form) and mediate the relation between a morpho-phonetic form and a structured concept/representatum (they specify what counts as an appropriate representans for a given concept).[14] Thus, speakers abide by representational conventions in doing two things: (i) constructing a symbolic representation and (ii) representing structured concepts (thought). It is this second (ii) application of representational conventions which is fundamentally involved in meaning and which I am concerned with.

2.2.2.2 *Meaning-as-relation*

The RH's semiotic account of the relation between sounds and conceptual structures brings with it a radical change in the understanding of what it is for a word to mean something. In the RH, having meaning does not equal having semantic properties. Words clearly are meaningful but do not have meaning or semantics as a constitutive property; words themselves have no conceptual attributes. Only thought has conceptual/semantic properties.

[13] Convention governs any relation which is not physically, logically, or in any sense naturally, necessary.
[14] I later argue that crucial to the definition of the RH's convention is Hintzman's (1986) multiple-trace theory of memory, which I discuss in 2.2.3. For that reason, I cannot define RH's convention before introducing Hintzman's model. In the meantime, conventions can be thought of in terms of what they are for.

The argument of the RH is that uttered words are symbolic signs. Like other symbolic signs (e.g. the 'national speed limit applies' sign discussed earlier), and like signs in general, words are meaningful without having semantic properties attributed to them. In the RH (Burton-Roberts 2007, 2012), *meaning* and *semantics* are different. However, they are related in the following way:

> ***Meaning* is not a property of anything, but a *relation* for someone between X (anything, including phonetic phenomena) and what has *semantic* content Y (thought and only thought)**

There are three important points in this definition. First, linguistic signs (like other signs) in and of themselves are not meaningful. They only "have"[15] meaning *for someone*. Second, thought is *the only* locus of semantic content. Linguistic signs, like all other signs, do not have semantic properties. Third, a linguistic sign is meaningful for the hearer (H) if and only if it leads H to have a thought. Meaning of any sign lies in the relation it has to a thought in a cogniser's mental world.

An uttered word leads H to have a thought if and only if H recognises the communicative intention of the speaker and H knows the representational conventions. For illustrative purposes, let us consider again the symbolic sign 'national speed limit applies' (figure 2.1). If a person A, who is well versed in the Highway Code, sees a pile of 'national speed limit applies' road signs stack up on the side of the road, A will not think that A can speed up to 70 m/h (or 60m/h if A is driving a vehicle with a trailer) because A does not recognize a communicative intention – A knows that the pile of signs was left on the side of the road by a road maintanance crew rather than intentionally used to communicate something. Conversely, recognizing a communicative intention (in virtue of understanding that a sign, which is placed on a pole on the side of the road, has been placed there to signify something) will not lead a person B to have a thought that B can speed up to 70 m/h (or 60m/h) if B has not learnt the Highway Code (i.e. the conventions). Thus, for a 'national speed limit applies' sign to "have" meaning, three conditions must be met: (i) the sign must be used with a communicative intention (here to speed up to 60 or 70 m/h, depending on the context), (ii) the communicative intention must be recognised by a person, and (iii) the person who recognizes the communicative intention must know the relevant convention.

[15] I am using inverted commas here to emphasise that in the RH meaning is not equivalent to the encoding of semantic properties. In the RH, linguistic signs "have" meaning insofar as they lead cognisers to have a thought.

Coming back to linguistic signs, in the RH, linguistic signs are purely phonetic/physical phenomena – they do not have any conceptual/semantic content. Phonetic phenomena "have" meaning ***for someone, on a given occasion of use*** because they give rise to particular semantically constituted thoughts (in virtue of recognising communicative intention and invoking a representational convention).[16]

In 2.1.4, I suggested that my argument that there is no linguistic semantics (and no lexical concepts) and that utterance interpretation is a wholly pragmatic, inferential process needs to face two potential criticisms. The first was that if words do not have meaning in virtue of encoding linguistic semantics as their constitutive property, then words cannot have meaning. I argue that the Representational Hypothesis' definition of *meaning-as-relation* invalidates this criticism. In making the unambiguous distinction between semantics (i.e. the content of thought and only thought) and meaning (i.e. a relation to semantics), the RH offers a way of saying that linguistic signs do not have/encode semantic properties, reconciling this with the fact that uttered words are meaningful. The RH explains how linguistic signs, like other signs, "have" meaning, i.e. lead cognisers to have a (semantically constituted) thought, without attributing the anyway problematic linguistic semantic properties to linguistic signs.

In rejecting the notion of linguistic semantics, I reject the notion of meaning-as-property (henceforth 'meaning/semantics') which follows from the double-interface tradition, and endorse Burton-Roberts' notion of meaning-as-relation (henceforth 'meaning').

2.2.2.3 *Semantics in the Representational Hypothesis*

In the RH, meaning in language is a relation for someone between a phonetic phenomenon and semantic content (of thought); sounds are used in aid of conventionally representing semantic content generated by a radically internalist, innate system which the RH identifies as LOT.[17] In the RH, talking about semantics equals talking about thoughts and the concepts they are composed of.

The RH (Burton-Roberts 2011) assumes that humans from birth have some kind of direct but deeply subconscious access to concepts (I shall call this PRIMARY ACCESS). It seems plausible to assume that Chomsky (2000b: 76) reports on primary access to

[16] The distinction between meaning and semantics is independently motivated. Burton-Roberts (2012) argues that a thought itself may have meaning (i.e. may lead a cogniser to a further thought) and that the meaning of a thought is different from the thought's semantic content. I return to this point in chapter 5.
[17] The RH's position is that there is LOT, but, in contrast to claims made in Fodor (1998, 2008), its concepts have a compositional, internalist content. The arguments presented in chapter 1 support the RH's position.

thought when he talks about thinking without language, thinking which is 'hard to articulate' (Quote 2, chapter 1). The RH argues that what is acquired in acquiring a particular language (CSPR) is another, more conscious, though indirect (because mediated by a particular language) kind of access to concepts (SECONDARY ACCESS); the acquisition of a particular language allows for the *activation* (to be contrasted with Fodor's acquisition) of concepts in a more or less conscious mental life.

In the RH (2012), concepts represented in communication are taken to be structured. It is helpful, I think, to identify the primitive components of such a conceptual structure as Jackendoff's (2002) 'quarks' – i.e. concepts 'ineffable' in isolation. Such conceptual quarks and structures defined over them (generated by LOT) are not acquired but innate.

Burton-Roberts (2011, fn 16) argues that concepts are prior to and hence independent of language acquisition during which they are activated and accessed by proxy (i.e. labelled by a representational label). Note that, in the RH, conceptual 'quarks' are posited on independent principled grounds: if sound is to function as a sign, there must exist something it is a sign of, namely a signified, independently of the fact of signification. In other words, that there be a concept to signify is a precondition for signification (Burton-Roberts 2012).

The existence of this conceptually necessary precondition supplies another argument against the double-interface view of language and in favour of the account proposed by the RH. Burton-Roberts (2012) observes that for Saussure concepts exist only as constituents of linguistic signs (i.e. double-interface objects) and, therefore, that thought is couched only in the signs of a natural language. The problem is that if concepts exist only as parts of linguistic signs, the set of concepts must be arbitrary since signs are arbitrary. This is an extreme version of the Sapir-Whorf hypothesis – on the Saussurean double-interface view of language, thought is completely determined by particular languages. The RH's representational relation between linguistic signs and concepts, in particular its unambiguous distinction between linguistic signs and concepts, predicts that the set of entertainable concepts, and so thought, is delimited by human nature. It is the secondary access to such pre-linguistic thought that is mediated by and hence delimited by the conventions of a particular language.[18]

[18] Kjøll (2009) also argues for an unambiguous distinction between words (i.e. linguistic signs) and concepts. Kjøll's point, however, is not that it is only concepts which have semantic content. Kjøll argues that whereas concepts have externalist semantic content á la Fodor, words have internalist semantic content.

In acquiring secondary access to pre-linguistic thought, one acquires a system for conventional representation (CSPR) of thought. Acquiring such a system allows for thinking/computing by proxy.[19] This is possible because acquiring a particular language allows for the labelling[20] of the activated conceptual structures (i.e. making an association between a sound and a particular structure). In this sense, a label 'packages' an intricate conceptual structure and gives us a short-cut to it (Burton-Roberts 2011). Sounds utilised by a particular language give us access to such a 'package' if constituent concepts are associated with another sound in that particular language (if they, too, are labelled). However, if there is no such association for a constituent concept, we are dealing with conceptual 'quarks'. 'Quarks', thus, are 'ineffable' in the sense that they are not labelled.

Now, compositional theories of concepts have often been criticised for not being able to specify what such compositional content might be (e.g. Fodor 1998, Laurence & Margolis 1999). The RH's distinction between primary (subconscious) and secondary (more conscious) access to thought offers an explanation of why speakers of particular languages are (often) unable to 'define', or represent, compositionally constituted semantic/conceptual content. In essence, the fact that particular languages do not give us a (more conscious) access to conceptual quarks does not mean that quarks do not exist – it only means that speakers cannot represent compositional content in a particular language (c.f. Chomsky 2000b: 76). But this fact does not necessarily preclude (as it does for Fodor) the existence of such content. From the RH's perspective, inability to define is an almost unavoidable consequence of the advantage conferred by lexical 'packaging'.

I opened this chapter with the discussion of the notion of a pointer. The RH's overall contribution to the discussion about linguistic pointers is the following. The idea that linguistic signs are pointers to a conceptual space is plausible if and only if we assume that pointing can be understood in terms of the semiotic (in the sense of Peirce) relation of conventional representation.

The consequences of adopting the RH's account of representation and, consequently, its claim that words do not have meaning as a constitutive property are very radical. The way in which CSPRs mediate between sounds and conceptual

[19] Thinking/computing by proxy equals using 'the representations themselves as proxies for what they are representations of' (Burton-Roberts 2011). For illustration, Burton-Roberts offers a comparison with an arithmetical calculation with the use of pen and paper, where the calculation is performed over conventional symbols rather than arithmetic concepts.
[20] I will say more about labelling in 2.2.3.2.

structures (figure 2.3) does not involve positing double-interface objects and, therefore, it does not involve positing a distinction between deterministic decoding and contextually-constrained inferring. This gives rise to the question of how CSPRs mediate between acoustic events and conceptual structures. It also invites the second criticism of the argument that there is no linguistic semantics, namely the worry that without linguistic semantics to constrain word use and interpretation, it is difficult to explain how we ever communicate successfully. If there is no linguistic semantics to mediate between acoustic events and conceptual structures – no 'common core of meaning' – cannot words, in principle, mean anything? Doesn't utterance interpretation without linguistic semantics uninvitingly look like a Humpty-Dumpty enterprise?

In the next section, I argue that the underlying mechanism for how CSPRs mediate between sounds and structured concepts can be cashed out in terms of a multiple-trace theory of memory (e.g. Hintzman 1986, 1988, 2008). In the light of Hintzman's (1986) model, I will argue that the Humpty-Dumpty worry is unwarranted.

2.2.3 *Hintzman's multiple-trace theory of memory*

Hintzman's (1984, 1986, 1988, 2008) multiple-trace theory offers a model of memory and information retrieval from memory. It is relevant to theories of utterance interpretation insofar as utterance interpretation is an activity which involves the retrieval of (relevant) information from memory. The overarching question that Hintzman is concerned with is how abstract (generic) knowledge is related to specific (episodic) experience. Hintzman disagrees with a view (e.g. Tulving's 2002) that abstract, unitary representations of a category are stored in a functionally separate generic memory system. On Hintzman's view, generic knowledge does not have a special status and is not stored in a functionally separate memory system, but can be retrieved *on-line* from a pool of episodic memory traces. Hintzman's model has serious implications for linguistic theory as, like the RH, it dispenses with the decoding-inferring distinction. I explain this in the remainder of this section.

2.2.3.1 *Memory traces and echo retrieval*

On Hintzman's theory, each experience, including linguistic experience is stored as a separate memory trace. As for the question of what memory traces consist in, Hintzman assumes that experiences are internally represented as an active configuration of primitive properties. By 'primitive properties' Hintzman means anything from

modality-specific sensory features (e.g. basic colours and odours), simple emotional tones, properties accessible by more than one modality (e.g. intermittency, spatial location) to primitive abstract relations (e.g. before, same as).[21] Similar experiences will share certain configurations of properties. Hintzman (1986: 412) argues that such primitive properties are distinct from the ability to 'label' them in that they are not acquired through experience. Given that for Hintzman (1986: 412) primitive properties that constitute memory traces are innately specified, I think it is not misguided to think of Hintzman's primitive properties as primitive concepts, or Jackendoff's (2002) 'conceptual quarks' (i.e. concepts ineffable in isolation) and of Hintzman's configurations of primitive properties as 'structured concepts'.

In a linguistic context, Hintzman's model predicts that every communicative event to which a person attends – such as hearing an uttered word – will leave a new memory trace (an association between a mental representation of an acoustic event – a word form – and a structured concept). Such a trace will co-exist in memory with other occurrences of the same word form and associated structured concepts. Using the 'label' metaphor mentioned above, we can say that there will be aggregates of memory traces storing information about associations between a given word form, which we may call an acoustic 'label'[22], and structured concepts.

Hintzman (1986) distinguishes between PRIMARY MEMORY (PM) and SECONDARY MEMORY (SM). PM is the active representation of (a record of) the current experience and SM is a pool of largely dormant memory traces. PM and SM interact in the following way.

The active configuration of primitive properties in PM (i.e. current experience) constitutes a RETRIEVAL CUE or PROBE which is sent to all traces in SM, which it activates according to their similarity to the probe (where similarity depends on the extent to which the traces in SM and the probe share the primitive properties). PM then receives a single reply or *echo* from SM. The echo that emanates back from SM is a pattern of most strongly activated properties and it is the echo that, for Hintzman, corresponds to the interpretation of a word on a particular occasion of use.[23] Depending

[21] It is not clear whether Hintzman thinks of the properties as multi-modal or whether he considers them to be some kind of a unicode, able to access and bind different modality representations. My interpretation is that of a unicode, or multi-dimensional code, of the sort assumed by Baddeley (2000) to be involved in integrating information in episodic long-term memory and episodic buffer.
[22] 'Labelling' is discussed in more detail 2.2.3.2.
[23] Hintzman (1986) actually uses the term 'meaning' to refer to occasion-specific interpretations (retrieved echoes). To avoid confusing Hintzman's meaning-qua-echo with meaning/semantics (or indeed meaning-as-relation), I do not use the term 'meaning' in the sense of Hintzman's echo.

on the structure of the probe, the information retrieved from SM can be of different degrees of abstractness.

Hintzman (1986, 2008) argues that the process of echo retrieval can retrieve 'the essence' of what, for example, dogs are from individual memory traces. When cued (e.g. when asked to think of a definition of a word, or on hearing a generic statement), a generic concept (i.e. a schema) of a dog can be retrieved on-line by cumulative activating of all traces and cancelling out the properties that are not shared by the traces. On this view a schema is 'a temporary, dynamic structure that springs into being when a retrieval cue occurs' (Hintzman 1986: 424). The crucial point is that the retrieval of such abstract/schematic echo – like all instances of echo retrieval – is necessarily context-dependent. In other words, echo retrieval can yield 'different nuances, different levels of abstraction, or entirely different meanings of a word by addressing different subsets of stored contextual features' (2008: 25). This means that on a generic use of a word the individuating properties of traces, such as temporal and spatial location properties, will be cancelled out. The experience, Hintzman (1984: 241) remarks, will be abstract and devoid of specific details. The crux of Hintzman's theory is that the retrieval of such an abstract/schematic concept does not, in any sense, happen by some default – the process underlying it is the same as the process underlying the retrieval of more specific concepts.

This context-sensitivity follows from the nature of the probe. The probe, Hintzman (1986: 420) emphasises, consists not only of the mental representation of a relevant acoustic event but also of its context. The echo retrieved by such context-sensitive probes, i.e. a function of the particular subset of episodic traces activated by the probe, is thus necessarily context-sensitive too, whether schematic or not.

In Hintzman's model then there is no room for context-independent conceptual schemas/general concepts necessarily mediating utterance understanding. The general mechanism of echo retrieval handles cases where general concepts are retrieved and cases where they are not retrieved.[24] On Hintzman's model, the existence of general concepts (as memory traces) does not amount to the existence of linguistic semantics –

[24] More recently, similar claims have been made by Barsalou (e.g. 2005, 2012), who argues that abstractions are dynamically derived temporary online constructions. Indeed, Barsalou (2005: 417) goes on to suggest that abstraction is a skill rather than a structure; he argues that 'what develops permanently is not a fixed summary representation, but a skill for interpreting instances effectively and efficiently'. The Hintzman/recent Barsalou view is in opposition to the traditional view (e.g. Barsalou 1982, Smith & Medin 1999) that categorisation of memory traces is impossible without some sort of abstract 'summary representation' (which corresponds to the notion of a lexical concept).

memory traces do not 'resolve' themselves into a schema which is then deterministically accessed in utterance interpretation (contrary to claims made by Carston (2002: 364-365)).

The idea that both general concepts and more specific conceptual structures are retrieved by the very same *contextually-constrained* process is particularly important in the context of Relevance Theory. As mentioned, RT's distinction between linguistic and real semantics is cashed out in terms of two types of cognitive processes – deterministic, context-independent decoding of linguistic meaning (whether schematic or full-fledged) and pragmatic inferring of speaker meaning. Hintzman's model effectively undermines this distinction, making the process of utterance interpretation radically contextualist and wholly pragmatic/inferential.

Before I illustrate Hintzman's account with some examples, I discuss his notion of 'labelling' in the light of the Representational Hypothesis.

2.2.3.2 *A note on labels*

Hintzman (1986: 412) argues that what happens in so-called 'word meaning acquisition' is establishing of a relation between an acoustic event on the one hand and a configuration of primitive properties on the other. What is acquired is the ability to 'label' primitive properties. Now, Hintzman (*ibid*.) does not say much about the process of acquiring this 'labelling' ability, but, since I intend to apply his model to utterance interpretation, it is important to make a brief comment on that.

According to Burton-Roberts (2012), there are three conditions for an acoustic event to be recognised as a linguistic signifier for a conceptual structure: (a) convention, (b) semiotic intention, and (c) inferentially derived recognition of (a) and (b). These conditions are fundamental to the process by which an acoustic event comes to function for a hearer (H) as a semiotic label in the following way.

If in a linguistic context, a relation is to be established between a given acoustic event and some conceptual structure, H must *infer* that there is a relation between the acoustic stimulus and the conceptual structure which is being currently entertained in H's mind. More specifically, H must infer that the speaker (S) produced the acoustic stimulus with the intention of leading H to entertain a particular conceptual structure.

For illustrative purposes, let us imagine that an infant learns to associate the acoustic event [p^{h}en] with a mental representation of an enclosure in which it is often put, complex emotions it gives rise to (i.e. sometimes boredom, sometimes happiness, sometimes fear) etc. Presumably, for the infant to learn that association, the parents

must have made the sound [p^{h}en] more than once in the context in which the infant could make that association, and the infant must have recognised the parents' communicative intention. Now, being able to discern the similarity of new experiences of uses of [p^{h}en] with old experiences means that the infant has recognised/relies on a communicative convention. The point is that an acoustic event becomes a semiotic label for a given cogniser in virtue of the cogniser's recognition of a communicative intention and convention experientially associated with a given acoustic event. Once the convention is recognised (more on convention shortly), similar memory traces will be stored under the same morpho-phonetic label. At this stage in our scenario, a morpho-phonetic label [p^{h}en] has multiple memory records associated with it.

Now, records of experiences in a linguistic context are not stored under a particular morpho-phonetic label based exclusively on the morpho-phonetic properties of the relevant acoustic event. For example, in our scenario there is an infant in whose mind the morpho-phonetic label [p^{h}en] has been established and is associated with an aggregate of records of experiences with an enclosure (or enclosures). This association, which is bound up with the infant's recognition (even if below the level of consciousness) of communicative intention and convention, will influence the interpretation of new experiences of uses of the morpho-phonetic label [p^{h}en]. In the initial stage thus, if someone is holding a writing instrument in their hand and says to the infant that it is a pen, the infant will interpret the communicative intention in accordance with the already established associations. For example, the infant may think that the person wants the infant to put the object that the person is holding in their hand in the enclosure.

Presumably, at some point the infant will come to notice inconsistency/ contradiction in uses of [p^{h}en]; mental representations that certain uses of [p^{h}en] give rise to will not be congruent (in terms of similarity of primitive properties they are composed by) to those stored under the already established morpho-phonetic label [p^{h}en]. In simple terms, some utterances of [p^{h}en] will not be associated with experiences of an enclosure but with experiences of the action of writing or doodling. At some stage, the infant will recognise that the communicative intention associated with the utterances of [p^{h}en] is different in different contexts. In recognising the difference in communicative intention, the infant will recognise the difference in communicative convention and, arguably, a separate aggregate of memory traces will be established – this aggregate will share the morpho-phonetic label [p^{h}en] with the old

one, but will be separate precisely because the memory records are not compatible in terms of the primitive properties they are composed of.

This simple scenario illustrates the role of the recognition of communicative intention and convention in the acquisition of what we may call the 'labelling' ability. It seems to me that the importance of the labelling ability is best appreciated when considered in relation to Burton-Roberts' (2012) argument (discussed in 2.2.2.3) that what is acquired in so-called language acquisition is no more and no less than a particular kind of *access* to conceptual structures. On this account, the establishing of a morpho-phonetic label consists in the 'packaging' of various intricate conceptual structures in one aggregate of multiple traces and gives a short-cut to them, allowing for faster and more efficient processing. In 2.2.3.4, I say more about how the notion of communicative conventions construed in terms of a multiple-trace theory of memory is different from standard models (e.g. RT) and why, I believe, it is better. In the meantime, I discuss some advantages of a wholly inferential account and illustrate the application of Hintzman's wholly inferential model to utterance interpretation with some examples.

2.2.3.3 *Some advantages of a wholly inferential account*

Perhaps the most straightforward advantage of a wholly inferential account is the fact that it does not posit deterministic decoding, which is most obviously unnecessary in cases like (3) (*The steak is raw*). Given that there is no logical relation between the alleged lexical concept of not being cooked and the communicated concept of being cooked but not long enough, even RT must admit that here successful communication relies solely on the power of pragmatic inference. Consider now example (7), uttered during a phone conversation between two friends who together attended a comedy show last week.

(7) That comedian killed me.

The information communicated by the utterance of (7) is that the comedian was hilarious or very entertaining. Hintzman's model predicts that the utterance of (7) may activate the concept of being very entertaining without first activating the concept of causing death.[25] This is possible because the noun *comedian* (co-text) constrains the

[25] This is compatible with Gibbs' (2002) and Hamblin & Gibbs' (2003) observation that, given the right context, even novel non-literal forms may be processed as quickly as or more quickly than their literal counterparts. I return to this issue in chapter 5.

cognitive search in that it activates the theme of entertainment[26], whereas the demonstrative *that* (co-text) further narrows it down to an experience which is familiar to both the speaker and the hearer. Quite plausibly, the relevant cognitive context is such that the hearer already knows that the speaker enjoyed the show (by watching her reactions at the show). Furthermore, the presence of the pronoun *me* in the co-text and the perceptual presence of the speaker in the situational context exclude the possibility of the speaker having been murdered by the comedian.

It needs to be emphasised that I am arguing against deterministic decoding of a concept which is, purportedly, associated by default (i.e. invariably) with a given morpho-phonetic form. In other words, I allow that for *some hearers* the utterance of (7) *may* initially activate the concept of causing death, and that the utterance of (3) *may* initially activate the concept of not being cooked. Whether it does or does not depends on many (contextual) factors. For example, it is undeniable that some conceptual structures may be more easily accessible than others due to frequency effects (e.g. Dąbrowska 2004: 25-27). However, the existence of frequency biased interpretations – sometimes referred to as 'dominant meanings' (e.g. Gibbs 2002) or 'attractors' (e.g. Barsalou 2005) – does not amount to the existence of linguistic semantics. Firstly, from the existence of attractors, it does not follow that it is language, and not thought, which hosts them, or that it is the linguistic processes, and not the pragmatic processes, which access them. Secondly, *deterministic* activation of such attractors would not always be an efficient interpretive strategy (as I will shortly show). Indeed, Barsalou (2005: 416) argues that various factors (e.g. contextual evidence) may inhibit attractors and facilitate other interpretive strategies. Consequently, the interpretation process remains highly dynamic even if there are frequency-based attractors. Let me illustrate this with example (3) (*The steak is raw*).

If the conversation of which (3) is part takes place among two people who often dine together and where the hearer knows that the speaker likes his steak well done, the utterance of (3) will most likely activate the concept of being cooked not long enough. If, however, the hearer of (3) is a vegetarian who has heard stories about people eating raw meat (e.g. steak tartare) but does not know what such dish looks like, the utterance of (3) may initially (i.e. before clarification or further evaluation of the context) activate the concept of not being cooked.

Consider now a scenario such that (3) is uttered in a context where on a plate is not an English steak but steak tartare, and where the hearer, well accustomed to English

[26] This would normally contrast with the theme activated by the expression such as *the mafia boss*.

steaks, is unaware of the fact that some people do eat, and some restaurants do serve, meat that has not been cooked. Also in this situation, the concept which is initially activated in the mind of the hearer is likely to be that of not being cooked for long enough. If, on the other hand, the hearer is familiar with steak tartar, the concept activated in his mind on hearing (3) in such context will most probably be that of not being cooked.

The point is that a wholly inferential account, on which utterances are interpreted in relation to a *contextually constrained* background of individualistic beliefs and associations, allows us to explain such variations. Furthermore, it offers a more efficient account of processing – it does not require that the *irrelevant* concept NOT COOKED be decoded where there is clear contextual evidence to the contrary. At the same time, the contextually constrained process of echo retrieval accounts for why very often what the speaker communicates and what the hearer takes her to communicate 'converge' to a sufficient[27] degree. Because the process of echo retrieval is contextually constrained, not all memory records are involved in the derivation of the echo, only those relevant to a **particular conversation** in a **particular context** between **particular interlocutors**.[28] Hintzman's (e.g. 1986) contextually constrained process of echo retrieval, combined with Burton-Roberts' (2012) definition of meaning-as-relation, make assumption (B) – that H converges (to a sufficient degree) on what S communicates *despite* there being no shared linguistic semantic content – plausible.

Needless to say, a wholly inferential account abides by Occam's Razor in cases where the alleged lexical concept is the same as the communicated concept. As mentioned, RT's problem with utterances like (6) (*Children in most cultures dance spontaneously*) is that the very same concept of dancing is activated by two distinct cognitive processes – it is deterministically decoded and then it has to be pragmatically inferred anyway. In contrast, Hintzman's model predicts that, on hearing the utterance of (6), a generic concept of dancing is retrieved because the context is such as not to trigger the activation of a more specific concept – relevant here is the presence of bare nominal *children* and the quantifier *most* in the co-text. Such generic concept does not have to be deterministically decoded first as there are enough clues in the context to pragmatically derive it.

[27] Convergence has to be sufficient enough to co-ordinate the interlocutors' actions, including the very action of communicating.

[28] This radical contextualism is compatible with arguments put forward by Bilgrami (1992). Bilgrami's account will be discussed in chapter 3. Horton & Gerrig (2005) is an important development of the point that the knowledge of one's interlocutor places immediate constraints on utterance interpretation. This is discussed in detail in chapter 5.

Similarly, a wholly inferential model has an advantage over a two-step model of utterance interpretation like RT in cases like (4) (*When Mr Cerezo, before his election, stood up against successive military regimes the rightists tried to kill him*) and (5) (*… one of the mechanisms by which phagocytes kill intercellular organisms is the production of reactive oxygen intermediates*), where the alleged lexical concept is more general than the ad hoc concept purportedly derived from it. As mentioned, the problem for RT is that if KILL* analytically implies, i.e. is compositionally constituted by KILL, the concept KILL has to be retrieved by two distinct cognitive processes.

In contrast, a wholly inferential model predicts that the concept of causing death is pragmatically inferred as part of the concepts communicated by (4) and (5) without being first decoded. The process of pragmatic inference is sufficient to derive the subtleties of different concepts which are in part constituted by the concept of causing death. These subtleties are inferable from the context of (4) and (5). In (4) there are several aspects relevant to arriving at the intended interpretation. First of all, the verb *kill* here is interpreted in the context of the co-text such as *stood up against, military regimes* and *the rightists*. The co-text thus restricts the conceptual search to the hearer's knowledge of extreme political/military regimes and their treatment of opposition. In this context, the concept of causing death by a human agent intentionally acting upon a patient is highly relevant and thus easily accessible. In (5), the concept of causing death does not involve the concept of an intentional act of a human agent, but of biochemical reactions at a cellular level. This is inferable from the co-text (e.g. *phagocytes*, *intercellular organisms*), but the relevant conceptual space may have also been activated by preceding discourse. This, again, satisfies Occam's Razor.

However, what is controversial about cases of the so-called 'concept narrowing' is that the uses of the word *kill* in (4) and (5), and in fact many other uses of this word, do undeniably share 'a common core of meaning' – the concept of causing death. From an RT perspective, this seems to be almost tangible evidence for the existence of encoded semantics (if we disregard all the problems with this notion I have discussed so far). But is it?

I have already mentioned that instances like (4) and (5) can only be considered as evidence for linguistic semantics if one assumes that there exists linguistic semantics (assumption (A)). If no such assumption is made, (4) and (5) illustrate a situation where two uses communicate distinct specific concepts which are compositionally constituted by the same generic concept. In terms of a multiple-trace memory model, mental records of such uses (i.e. memory traces) are constituted by partially overlapping

properties. Thus, the intuition that (4) and (5) share some 'common core of meaning' arises from the partial overlap of conceptual structures; it can be explained without making the anyway problematic claim that such an intersection is functionally independent (i.e. deterministically decoded by the language module). Apart from all the problems discussed earlier, why else should we not extrapolate from the existence of partially overlapping properties to the existence of linguistic semantics? Consider (4) again, repeated here for convenience.

(4) When Mr Cerezo, before his election, stood up against successive military regimes the rightists tried to *kill* him.

RT tells us that before pragmatic processes go to work, not only does the language module access lexical semantic content of the words, but it also compositionally combines them into a logical form. Now, a wholly inferential account predicts that the preceding discourse, situational context as well as co-text provide sufficient evidence to constrain utterance interpretation so that by the time the speaker utters the word *kill*, the cognitive space of military regimes and their ways of dealing with opponents is activated in the mind of the hearer, allowing for a specific interpretation of this particular use. On a two-step model like RT, however, the interpreter is assumed, by the terms of the theory, to ignore all that until the logical form is deterministically decoded. On RT's terms, the preceding discourse, situational context and co-text come out as merely secondary 'evidence' with respect to the primary evidential function of the allegedly encoded semantics.[29]

Importantly, a wholly inferential account is also supported by the consideration of compositionality, a principle which is supposed to explain productivity (i.e. the infinite expressive power) of a particular language. In the context of RT's distinction between lexical and ad hoc concepts, the compositionality principle should operate at the linguistic semantic level if it is to explain productivity of language. Accordingly, RT's compositionality principle operates to combine the encoded semantic content of linguistic expressions into a logical form. The resulting structurally complex logical form delivered by the linguistic module serves as an input to pragmatic processing. However, there is evidence which shows that compositionality does not take place at any context-independent level.

[29] In this sense, RT weakens its own claim (e.g. Carston 2002: 99-100) that pragmatics is much more substantially involved in meaning in language than Grice (1989) thought was the case. I return to this point in chapter 4.

For example, Recanati (2005) argues that the compositionality principle applies after the pragmatic processes have done their work. In other words, pragmatic processes do not operate 'globally' on a compositionally constituted 'output of the grammar'. The order is actually reverse to what Relevance Theory predicts – pragmatic processes are at play before the compositionality process applies. Consider (8) and (9) below (taken from Recanati 2005).

(8) There's a lion in the courtyard.

(9) There's a stone lion in the courtyard.

Let us assume that (8) is to be understood as communicating that what is in the courtyard is not a real animal but a representation, or statue, of a lion. The pragmatic inference that leads us to interpret the word *lion* in the sense of a statue is often referred to as REFERENCE TRANSFER. In the case of (9) too, we are dealing with reference transfer – we understand that what is said to be made of stone is a representation, or statue, of a lion, not a real animal. This fact, Recanati (*ibid*.) argues, shows that reference transfer must take place before the composition rule applies to the noun-noun (*stone lion*) construction. The evidence is this. The interpretation that we get for (9) is: *(a representation of a lion) that is made of stone*. But this interpretation is only possible if reference transfer occurs before the two expressions, *stone* and *lion*, are combined. If, however, reference transfer applied globally, i.e. after the compositionality process was applied at the linguistic semantic level, the interpretation we would get is: *a representation of (a lion that is made of stone)*. The absurdity of the result, Recanati argues, seriously undermines the view that compositionality applies at the linguistic semantic level.

It may be argued that this criticism is not applicable to RT given RT's underdeterminacy thesis and its emphasis on the pragmatic contribution to grasping the proposition explicitly expressed. However, in the light of the discussed RT's instability about the propositional nature of logical form and truth-theoretic properties of lexical concepts (section 2.1), the underdeterminacy thesis cannot be used as an argument against Recanati's criticism. In my opinion, what Recanati's argument challenges is not only the issue of compositionality at the linguistic semantic level, but, more fundamentally, the utility of the linguistic semantic level as such.

Indeed, Recanati's argument suggests that a much more radical underdeterminacy thesis is in order. As discussed earlier, in RT, communicated thoughts are underdetermined by linguistic signs in that linguistic semantic content by

which such signs are partly constituted needs to be pragmatically 'enriched' into real semantic content. A more radical underdeterminacy thesis is offered by the RH; in the RH, communicated thoughts are underdetermined by linguistic signs in that linguistic signs do not have semantic properties at all; pragmatic processing (along the lines of Hintzman, as argued earlier) gets an interpreter from a linguistic sign, which does not have any conceptual/semantic properties, to thought – the only locus of semantic properties.[30]

In the final section, I discuss the difference between assumptions (A) (that H converges on what S communicates because they share linguistic semantic content) and (B) (that H and S converge despite there being no such linguistic semantic content) with respect to the question of what constitutes evidence in utterance interpretation.

2.2.3.4 *The role of acquisition context and communicative context in utterance interpretation*

The difference between assumption (A) (advocated by RT) and assumption (B) (advocated here) amounts to the difference in what is taken to count as 'evidence' in a communicative act. As discussed, in RT the evidential role that context plays in utterance interpretation is secondary in the sense of being post-decoding. I argue that the secondary role ascribed to contextual information follows from RT's stand on what is acquired in so-called 'word meaning' acquisition. Let me explain this.

On the assumption that the product of 'word meaning' acquisition is a context-independent concept/schema, all contextually available information must be stripped out in the process of acquisition. But if what is acquired is context-independent and if the role of what is acquired is to guide future utterance interpretation (as in RT), it follows that utterance interpretation is fundamentally constrained by the context-independent product of acquisition and that contextual information comes out as secondary evidence. On such an account, the relevance of contextual information (both

[30] This radical underdeterminacy thesis is related to the RH's definition of parsing. In the RH (e.g. Burton-Roberts & Poole 2006a, 2006b), parsing consists in putting something (e.g. an acoustic event) which lacks conceptual/semantic content and structure (i.e. representans) into correspondence with something that has conceptual/semantic content and structure (i.e. thought), on the assumption that the relevant acoustic event was produced with the intention of conventionally representing a conceptual/semantic structure. Evidence for this conception of parsing comes from considering so-called 'structural ambiguity' (Burton-Roberts & Poole 2006b); because anything that has or is a structure can only ever have one structure (e.g. it is impossible for the Eiffel Tower to have two structures at the same time), the possibility of 'structural ambiguity' (as in '*Flying planes can be dangerous*') shows that the linguistic sign we are dealing with does not have/is not a structure. In the RH, 'structural ambiguity' arises because a CSPR allows one to parse the structure-less representans in two different ways.

incoming and that which is stored in memory) is conditional upon (i.e. preceded by and constrained by) the activation of the context-insensitive product of acquisition. In arguing against the notion of linguistic semantics and the deterministic process of decoding of such content, I am arguing against the claim that contextual information plays a merely secondary, i.e. post-decoding, evidential role.

On the account advocated here, contextual information (both incoming and stored in memory) has a primary function in utterance interpretation. This is possible because what is acquired is not stripped of contextual information – on the contrary, the product of acquisition is the relation between a morpho-phonetic label and an aggregate of multiple memory traces, which are themselves context-sensitive. On this account, the interpretation of acoustic stimuli is a wholly pragmatic, inferential process of accessing a conceptual structure (or Hintzman's echo) via the matching of the incoming contextual information (Hintzman's probe) with that which is stored in memory traces. The role of the morpho-phonetic label in the process of utterance interpretation is to constrain the process of echo retrieval so that the incoming stimuli are matched not with *any* memory traces but with those already associated with the relevant morpho-phonetic label.[31] Thus, it is the existence of the morpho-phonetic label (and not of some problematic notion of lexical concept) that allows for fast and efficient, contextually constrained, processing.

In 2.2.3.2, I argued, in line with Burton-Roberts (2012), that the recognition of communicative intention and convention is essentially involved in the process whereby a mental representation of an acoustic event becomes a label for an aggregate of conceptual structures. Acquiring the relation between a morpho-phonetic label and an aggregate of conceptual structures consists, I argue, in acquiring a communicative convention. Such a construal of convention is necessarily individualistic because one of the relata is the record of individualistic experiences, but its acquisition is constrained by a cogniser's observation of the communicative behaviour of others in particular contexts. Furthermore, the convention is also constrained by the H's assessment of the convergence of the H's interpretation with that intended by S – presumably, the convergence is assessed based on the degree of success of action co-ordination between S and H. The end effect is individualistic and hence divergent associations, along the lines of Chomsky's individualism (discussed in 1.2), which play a primary evidential role (together with the incoming information) in the derivation of interpretation which is sufficiently convergent *in a particular conversational situation* with that intended by S.

[31] Presumably, the process is further constrained by the active cognitive background.

Burton-Roberts (2012) argues that conventions (i.e. CSPRs) LICENSE the use of sounds as symbolic (representational) signs. In this sense, a word is defined by Burton-Roberts as a symbolic license/rule for the use of sounds as symbolic signs. In the light of the arguments presented in 2.2.3, I argue that a word as a license/rule/convention is to be understood as a multiple-trace aggregate in the sense of Hintzman (1986).

2.3 Conclusion

I ended chapter 1 by saying that the tension between Chomsky's double-interface claims (which make room for encoded semantics) and his I-assumptions (which make no room for encoded semantics) is the leitmotif of this thesis. The overall conclusion of chapter 2 is that this tension should be resolved by rejecting the double-interface view and endorsing the individualistic direction.

In particular, I have argued that the process of utterance interpretation is a wholly pragmatic inferential process. First, I discussed the problems with Relevance Theory's notion of linguistic semantics and lexical concepts. I argued that the introduction of full-fledged lexical concepts (i.e. concepts with truth-theoretic properties) leaves RT with no principled distinction between lexical and ad hoc concepts. Secondly, I argued that even if acquired, lexical concepts are either unnecessary in utterance interpretation (because post hoc) or cannot constitute adequate evidence for the intended interpretation (because individualistic). Thirdly, I argued that the process of deterministic decoding of linguistic semantics/lexical concepts is redundant in cases of loose use, cases of so-called concept narrowing and when the communicated concept is the same as the purported lexical concept.

I then argued, in line with Burton-Roberts' Representational Hypothesis (e.g. 2012), that linguistic signs can "have" meaning even if one rejects the traditional double-interface view of linguistic signs and with it the traditional assumption that linguistic signs have meaning in virtue of encoding linguistic semantic content. I endorsed Burton-Roberts' (*ibid*.) view of meaning-as-relation to semantic content of thought and his representational account of the relation between sounds used by speakers and conceptual structures represented by their use.

I argued that utterance interpretation can be successfully constrained (without evoking the anyway problematic notion of linguistic semantics) by adopting Hintzman's (e.g. 1986) multiple-trace theory of memory and information retrieval. Finally, I suggested that a multiple-trace model correctly predicts that contextual information (both incoming and stored in memory) places immediate constraints on utterance

interpretation. This allowed me to define a representational convention in terms of an association between a morpho-phonetic label and an aggregate of multiple memory traces.

The last point I want to make concerns a recent claim (Urquiza 2011) that Relevance Theory is actually compatible with Hintzman's multiple-trace theory of memory (contrary to arguments in Carston 2002: 375). Urquiza (2011) argues that Hintzman's model can be used in RT as a model for the pragmatic process of lexical concept adjustment/modulation. There are several observations to be made here. Firstly, Hintzman's model is a radically contextualist model of utterance interpretation – thus, to argue that it applies post-decoding contradicts the very rationale for adopting a model like that. The point is that to argue that a radically contextualist model works, and works well, is to argue that deterministic decoding is redundant. Thus, whereas I share with Urquiza (2011) my enthusiasm for Hintzman's model, her proposal is simply inconsistent in the light of her commitment to linguistic semantics/lexical concepts and thus to traditional (and standard RT's) notion of 'word meaning' (i.e. meaning-as-property).

I argue that assumption (B) (that there *is no* linguistic semantics) is much more plausible than (A) (that there *is* linguistic semantics) and that the information available in the context of the acquisition of a representational convention and in the context of utterance interpretation constitutes the primary (the only) evidence for utterance interpretation. As argued in this chapter, linguistic semantics is a highly problematic and redundant notion. On the view endorsed here utterance interpretation does not proceed from what is (controversially) shared to what sufficiently converges, but from what diverges to what sufficiently converges.

In the next chapter, I look at philosophical legacy in thinking about linguistic semantics. I discuss narrow and wide kinds of mental content (e.g. Putnam 1975, Burge 1979) and their relation to RT's notion of linguistic semantics. In the light of my discussion of various philosophical proposals, I maintain the RH's claim that there is only one kind of mental/semantic content and argue that such content is necessarily holistic/individualistic. I argue that holism is not a problematic approach to meaning and semantics, contrary to claims made by its opponents (e.g. Fodor & Lepore 1992).

Chapter 3. Unleashing holism

3.0 Introduction

Most generally, the doctrine of semantic holism states that the meaning/semantics of an expression is relative to the entire system containing it (Lepore 1999). Put differently, semantic holism suggests that the meaning/semantics of an expression is determined by its place in the network of beliefs constituting entire theories or even a cogniser's entire belief system (Block 1998a, Pagin 2006). Consequently, a holistic meaning/semantics of an expression is dependent on 'all or most other' expressions (Block 1995), a point I elaborate shortly.

Semantic holism may, but does not have to, be a consequence of the assumption that there is no distinction between logical (i.e. analytic) and encyclopaedic (i.e. synthetic) inferential relations (Block 1995). Fodor (1998, 2008), for example, rejects both the analytic-synthetic distinction and semantic holism and argues for semantic atomism. However, as argued in 1.1, Fodor's atomism cannot be maintained.

Importantly, the rejection of both the analytic-synthetic distinction and atomism has holism as its consequence. If there are no logical constraints on inferences licensed by a given word, the meaning/semantics of a word is, in principle, the sum of unlimited inferences the word may give rise to. For example, on a holistic account of semantic content, the word *cat* may have the concept of my neighbour as part of its meaning/semantics if my neighbour happens to have a cat (let us call it inference level I). But if this is the case, then the meaning/semantics of the word *cat* is indirectly dependent on the inferential relations that my concept of my neighbour may enter into. For example, the meaning/semantics of *cat* may be indirectly dependent on the inferential relation between MY NEIGHBOUR and OPHTHALMOLOGIST if my neighbour is an ophthalmologist (let us call it inference level II). Given that the concept OPHTHALMOLOGIST may licence an inference to the concept HOSPITAL (inference level III), we see that the meaning/semantics of the word *cat* is indirectly linked to the concept of hospital. When critics say that, on holistic approaches, the meaning/semantics of a term is defined relative to entire theory or belief system, what they have in mind is that the meaning/semantics of a word depends on inference levels I, II and III as well as the conceptual relations which inference level III gives rise to, and so on and so forth. Let me refer to this cross-relational characteristic of holistic

meaning/semantics as the NETWORK EFFECT. In essence, the network effect manifests itself in that if meaning/semantics is dependent on a cogniser's so-called encyclopaedic knowledge, word meanings/semantics is interdependent on many ('all or most') other words' meaning/semantics.

Another consequence of rejecting the analytic-synthetic distinction (in conjunction with rejecting atomistic accounts á la Fodor) is the lack of any objective, cross-speaker shared semantic content. Since people's knowledge about what the world is like differs depending on their idiosyncratic experiences, it is very unlikely that the meaning/semantics of any single word will ever be the same for any pair of interlocutors. For example, the word *water* will be related to complex beliefs about its chemical composition in the mind of a person who is chemically informed, but not in the mind of a person who is chemically ignorant (Bilgrami 1992). On a holistic approach thus the meaning/semantics of any word relies on, because it is couched within, an individual's psychology. I will refer to this consequence of holism as the RADICAL INDIVIDUALISM[1]. Indeed, the network effect and radical individualism contribute to what I referred to in chapter 2 as the 'Humpty-Dumpty worry'; if word meaning/semantics is cross-relational and individualistic, how do speakers ever communicate successfully?

In this chapter, I argue that a holistic account of mental content is inescapable and, in fact, not as problematic as critics (especially Fodor & Lepore 1992) argue. First, I introduce (the complexities of) the ontological distinction between wide and narrow kinds of mental content in 3.1.1. Then, in 3.1.2, I discuss the relation between various philosophical notions of mental content and Relevance Theory's (RT) linguistic semantics. I argue that RT's preoccupation with the search for shared semantic content (in the sense of cross-context stability) has resulted in an account of mental content which is contradictory in philosophical terms. In 3.2, I argue that all philosophical notions of shared content, namely causal-externalist wide content, social-externalist wide content and non-truth-theoretic narrow content, are problematic and that holism is the only plausible thesis about mental content. In 3.3, I endorse Bilgrami's (holistic) thesis about the unity and locality of content. I discuss how Bigrami's philosophy fits in with the wholly inferential model of utterance interpretation I argued for in chapter 2. Finally, I discuss how, in the light of Bilgrami's thesis and the Representational

[1] The network effect and radical individualism are conceptually distinct but closely related aspects of holism.

Hypothesis, we can and should distinguish between a domain of concepts and a domain of associations between semiotic labels and concepts.

3.1 Kinds of mental content

In chapter 2, I discussed the question of whether it is possible to account for mutual understanding between interlocutors in communication without positing the level of linguistic semantics. This question concerned the existence of context-independent lexical concepts, purportedly shared among members of the same speech community (e.g. Carston 2002). Now, linguists' preoccupation with shared linguistic semantic content has intertwined with philosophers' interest in WIDE and NARROW kinds of conceptual content (e.g. Putnam 1975, Burge 1979). In this section, I introduce the distinction between wide and narrow content, discuss problems related to this distinction and their relevance to my argument that there is no linguistic semantics.

3.1.1 *Revisiting shareability: wide content and narrow content*

The broadest characterisation of the wide-narrow distinction is in terms of the externalist-internalist distinction: wide content is defined in terms of the relation that thoughts (and their components, i.e. concepts) bear to mind-external things and properties, and narrow content is defined solely in terms of mind-internal conceptual relations (e.g. Bach 1996).

The wide-narrow distinction was introduced by Putnam (e.g. 1975), who argued that the content of many natural kind concepts, like WATER, depends not only on a cogniser's internal, i.e. psychological state but also, and importantly, on what is going on in the mind-external environment. To show this, Putnam devised a thought experiment known as the Twin Earth experiment. Putnam imagines a place, called Twin Earth, where everything is exactly as it is on Earth except one thing. On Twin Earth the substance called *water* is not H_2O but XYZ, i.e. it is chemically composed by some other stuff. Despite this chemical difference, water and twin-water look superficially the same, both fall from the sky and fill up rivers and lakes, both quench thirst, both are called *water*, etc. Putnam argues that when a man from the Earth (Oscar) and his Twin Earth doppelganger (twin-Oscar) both think that water quenches thirst, they may be in identical psychological states (i.e. identical in the narrow sense) but, because their thoughts have different (wide) contents, they actually have different thoughts with different conditions of application: one is about water quenching thirst and the other is

about twin-water quenching thirst. Putnam's conclusion about the term *water* is that Oscar and twin-Oscar use *water* to express different concepts, whose contents are constituted by mind-external things and properties. This holds regardless of the fact that, for Oscar and twin-Oscar, water and twin water have the same phenomenological properties (i.e. they subjectively seem the same in terms of appearance, taste, smell, etc.).

There are several points about Putnam's argument which are relevant to my discussion. Putnam defines conceptual content in relational (referential) terms – the content of a concept is wide, i.e. it is determined by what it relates (refers) to in the mind-external world. In doing so Putnam rejects the doctrine of psychologism, i.e. he rejects the idea that 'knowing the meaning of a term is just a matter of being in a certain psychological state' (Putnam 1975: 219). Putnam (1975: 227) famously argues that ''meanings' just ain't in the *head'* because what is in the head – i.e. the psychologistic/narrow – fails to account for varying contexts (e.g. H_2O versus XYZ). Now, when Putnam (e.g. 1975: 221) talks of narrow psychological states, he means psychological states which are independent of the mind-external environment in a (methodological) solipsist sense (see section 1.1). Putnam's narrow-wide distinction is thus a distinction between internalist and externalist notions of conceptual content and as such its legacy is evident today in the Chomsky-Fodor debate (discussed in chapter 1).

As observed by Swiatek (2012), Putnam's criticism was directed at the privacy or non-publicity of word meaning/semantics; for Putnam, meaning/semantics cannot be 'in the head' precisely because it is (has to be) publicly available for the purposes of communication. The Twin Earth experiment was designed to show that the relational wide content, unlike the narrow content, is sufficiently fine-grained to account for context variability (i.e. it co-varies with the context). Relational wide content is supposed to be publicly available in that it is dependent on the mind-external objective reality.

Another idea introduced by Putnam (1975) with respect to the public availability of meaning/semantics was the socio-linguistic hypothesis about the division of linguistic labour. Putnam (1975: 228) argued that the meaning/semantics of words is crucially social in character – it is 'possessed by the collective linguistic body, even though [it is] not possessed by each individual member of the body'[2]. The 'division of

[2] Actually, the assumption that 'meanings' are not possessed by each individual shows that they are not collectively shared. I discuss this point further in section 3.2.3.

labour' manifests itself in the fact that in order to establish the meaning/semantics of a word we often (need to) rely on the judgement of 'expert' speakers, where expert speakers are those members of a speech community who have acquired 'correct' – in the expert's sense – conditions for the application of a given word. For example, a 'correct' condition for the application of the word *water* (on Earth) is that *water* applies to H_2O.

In this context, Putnam discusses his own incomplete grasp of the words *elm* and *beech*. Putnam observes that his concepts of *elm* and *beech* are exactly the same, i.e. what he knows about elms and beeches does not make it possible for him to distinguish between the two kinds of trees. Because – in fact – it is known that these two words have different meanings/semantics, Putnam (1975: 226) argues that what is in his mind cannot be identified as the meanings/semantics of *elm* and *beach* respectively. Based on this observation, Putnam concludes that meanings/semantics cannot be equated with someone's idiolect but are determined by the conventions of public language, which in turn are determined by deference to experts. This is another sense, on top of the relational sense discussed above, in which meaning/semantics is publicly (i.e. objectively) available and another sense in which content is thought to be wide, i.e. determined by factors which are independent of an individual mind.

Relational wide content, which is individuated by causal co-variances in the mind-external environment, can be referred to as CAUSAL-EXTERNALIST CONTENT, whereas social wide content, which is 'the sociolinguistic state of the collective linguistic body to which the speaker belongs' (Putnam 1975: 229) can be referred to as SOCIAL-EXTERNALIST CONTENT.[3]

Putnam's social-externalist conception of meaning/semantics was developed by Burge (1979), who argued that meaning/semantics is individuated accordingly with some community norms. To show this, Burge devised a thought experiment where a man called Al believes that the word *arthritis* refers to a disease affecting joints (which is 'correct' in the expert's sense) as well as muscles (which is 'incorrect', i.e. not what experts mean by this word) and that he has developed arthritis in his thigh. Now, Cal – Al's Twin Earth doppelganger – also believes that arthritis is a disease of joints and muscles, but unlike in Al's speech community, in Cal's Twin Earth speech community the word *arthritis* is in fact used by experts to refer to a disease affecting both joints and muscles. Burge argues that even though both Al and Cal believe that arthritis is a disease of muscles, Al's belief is false whereas Cal's belief is true. This shows, Burge

[3] This terminology is borrowed from Bilgrami (1992).

(1979: 106) continues, that what Al and Cal believe is different and therefore that the content of their beliefs is partly a matter of external social environment. The crux of Burge's argument is that the 'misuse' of the word *arthritis* by Al can only be properly explained when we assume the existence of meaning/semantics constituted by community norms, i.e. social-externalist wide content; Burge claims that without such a norm, it would be impossible to know that Al has misused the term but Cal applied it correctly.

The interesting difference between the two types of wide content is the following. The social-externalist enterprise, through its overt reference to a speech community, seems primarily interested in the stability of concepts qua linguistic objects and their role in the explanation of 'correct' (Cal's) and 'incorrect' (Al's) linguistic behaviour. The causal-externalist content, however, especially as presented by Fodor (1998, 2008), is a much stronger thesis which has been developed to allow for mental generalisations across human species and across slices of time in a cognitive development of one person.[4] Differences apart, we can nevertheless see that both types of wide content have been devised with *some* notion of shareability in mind (more on shareability shortly). Indeed, both types of wide content have been posited as linguistic semantics in Relevance Theory (Sperber & Wilson 1998, Carston 2002).

As for the notion of narrow content, it has been developed in two ways: it can be described as a function from context to wide content (as in Perry 1977) or it can be defined in terms of a conceptual role, i.e. in terms of a concept's inferential connections with other concepts (as in Block 1986).

On the first (Perry's) characterisation, the narrow-wide distinction parallels the distinction between content which has no truth-theoretic properties, and that which has truth-theoretic properties, respectively (Recanati 1993: 66). Accordingly, Perry (1977) argues that narrow content, which can be *entertained*, is to be characterised as an incomplete thought (or sense), whereas wide content, which can be *apprehended*, is a complete (and thus truth-evaluable) thought. Different thoughts (different wide contents) may be apprehended in different contexts by entertaining the same narrow content (or sense), and conversely, the same thought (the same wide content) may be apprehended by entertaining different senses. For example[5], the narrow content of '*My pants are on fire*', when entertained on a particular occasion by Paul, gives rise to the

[4] In fact, Putnam's narrow-wide distinction was designed as a thesis about the content of *linguistic expressions* and was only subsequently employed in theories of concepts as such (as argued by Brown 2011).

[5] These examples are taken from Recanati (1993: 66).

thought that Paul's pants are on fire. But the same narrow content entertained on a particular occasion by Fred gives rise to the thought that Fred's pants are on fire. Conversely, both Paul and Fred can apprehend the thought that Paul's pants are on fire by entertaining different narrow contents: '*My pants are on fire*' and '*His pants are on fire*', respectively. The way in which non-truth-theoretic narrow content plays a role in the commonsense explanation of behaviour is illustrated by Perry with the following scenario.

> **Quote 1**: When you and I entertain the sense of "A bear is about to attack me," we behave similarly. We both roll up in a ball and try to be as still as possible. Different thoughts apprehended, same sense entertained, same behavior. When you and I both apprehend the thought that I am about to be attacked by a bear, we behave differently. I roll up in a ball, you run to get help. Same thought apprehended, different sense entertained, different behavior. (Perry 1977)

A similar scenario can be envisaged for Recanati's example. The identity of narrow content of '*My pants are on fire*' makes Paul and Fred behave in the same way (both run around in frenzy), whereas the difference in narrow content ('*My pants are on fire*' apprehended by Paul and '*His pants are on fire*' apprehended by Fred) makes them behave in different ways (Paul runs around in frenzy, Fred throws a blanket over Paul). According to Perry (1977), it is the non-truth-theoretic narrow content (sense), and not the truth-theoretic thought apprehended, which is tied to human action and thus implicated in explaining and predicting actions.[6] Importantly, when narrow content is construed as a function from context to wide content, it is narrow content – rather than wide content – which is tied to linguistic expressions in a stable way. On this characterisation, wide content is unstable (or fine-grained) as it changes depending on contextual co-variances. Thus, the implications that Perry's distinction has for the notion of linguistic semantics contradict those made by Putnam's distinction. Indeed, Perry's characterisation of the narrow-wide distinction seems to underlie the standard Relevance Theory's distinction between shared (non-truth-theoretic) linguistic semantics and (truth-theoretic) real semantics (section 2.1).

Significantly, the non-truth-theoretic narrow content may also be seen as a schema. Recanati (1993: 209) observes that narrow content of a thought episode can be characterised as 'not a complete representation, but a schema whose contextual

[6] Perry's argument cannot be maintained. If I entertain the same narrow content 'His pants are on fire' but apprehend different thoughts about (a) Mark, who's my friend, and (b) Tom, who I hold a grudge against, I may rush to help Mark but laugh at Tom. It is Perry's wide content, thus, that explains my differing actions.

enrichment yields a complete representation'; narrow content is 'what we get when we abstract from the 'objective' component contributed by the environment'. As a schema, narrow content is non-truth-theoretic because it is stripped of contextual detail. In other words, it is not 'fully representational' (Recanati 1993: 210).[7] As discussed in 2.1.2.1, in line with linguistic underdeterminacy thesis, Relevance Theory (e.g. Carston 2002: 364-365) has also defined linguistic semantics in terms of schemas.

On the second (Block's) characterisation, narrow content is construed in terms of inferential roles that a concept has in a cogniser's mental life. Block (1998*b*) observes that an inferential role theorist has the option of constraining a set of inferences in which a given concept participates so as to distinguish between content-constitutive and non-content constitutive connections and argue that narrow content is molecular – i.e. it consists in logical (analytic) relations (e.g. Block 1986). However, the main challenge for the molecular approach to narrow content is to find a non-arbitrary way to distinguish a determinate narrow content for each concept (Bach 1996, Fodor 1998) – a criticism I agree with.

The other option, endorsed by Block (1995) is to argue that the narrow-wide distinction is a distinction between holistic (private, individualistic) content and shared (causal or social) externalist content, respectively. I return to this proposal in 3.2.1.

When discussing shareability of linguistic semantic content (chapter 2), I distinguished two ways in which such content is supposed to be shared: across speakers and across contexts. The kind of shareability invoked by the notions of social-externalist and causal-externalist wide content corresponds to this linguistic sense of shareability qua stability or sameness across speakers and across contexts. Crucially, however, it does not apply to Perry's notion of wide content, as for Perry wide content changes depending on contextual co-variances (e.g. '*My pants are on fire*' uttered by Paul versus Fred). This gives rise to the question of why Perry's wide content is actually considered to be wide and, relatedly, whether the notion of shareability qua speaker and context stability or sameness is defining of the wide content.

In fact, it is not. There is another – more fundamental – kind of shareability involved in the philosophical notion of wide content. It is the sense of shareability qua non-privacy or public availability of mental content to other individuals. Indeed, Putnam's (1975) wide-narrow distinction was devised in order to explain how

[7] Whereas schemas are context-independent, they do not seem to be independent of the mind-external environment since schemas are, presumably, acquired in experience (see 2.1.2.1). In this context, it is interesting to consider Bach's (1996) comment that one challenge for the proponents of narrow content is to 'specify narrow contents informatively, rather than by abstraction from wide contents'.

meaning/semantics is (has to be) *non-private but publicly available* for the purposes of communication. The notion of wide content is not as much about stability (though compatible with it), as it is about non-privacy afforded by what is mind-external. Causal-externalist content is supposed to be publicly available in that it is, by definition, constituted by relations to the mind-external objective reality. Social-externalist wide content is supposed to be publicly available because it is, by definition, determined by the conventions of *public* language, which in turn are determined by deference to experts. Perry's wide content is supposed to be publicly available because, by definition, it depends on (and thus varies with) contextual co-variances. What these three notions of wide content have in common is that, regardless of the issue of stability, they are all shared in the sense that they are publicly available because in one way or another dependent on the mind-external objective reality. This dependence of wide content on the mind-external reality is what makes it truth-theoretic (i.e. truth-evaluable). Therefore, it is the notion of shareability qua public availability, and not qua stability, that distinguishes wide content from narrow content.

The point that shareability qua stability is *not* defining of wide content is best illustrated by the following quote from Block (1998b):

> **Quote 2**: According to the external factor, 'Superman flies' and 'Clark Kent flies' are semantically the same since Superman = Clark Kent; the internal factor is what distinguishes them. But the internal factor counts 'Water is more greenish than bluish' as semantically the same in my mouth as in the mouth of my twin on Twin Earth. In this case, it is the external factor that distinguishes them.

Wide content in the Superman = Clark Kent scenario can be referred to as COARSE-GRAINED wide content, whereas wide content in the Twin Earth scenario can be referred to as FINE-GRAINED wide content (see e.g. Fodor & Lepore 1992: 169). Whereas coarse-grained wide content is stable, fine-grained wide content changes with contextual co-variances, of which the Twin Earth scenario is a rather extreme example.[8]

[8] Indeed, the coarse-grained versus fine-grained discrepancy in the wide content of the word *water* is apparent even here on Earth. Social- and/or causal-externalist wide content is supposed to enable theorists to say that whoever uses the term *water* on Earth to refer to H_2O does so 'correctly', and whoever uses this term to refer to XYZ does not. It might be argued that the notion of wide content explains successful communication (members of the same speech community use the term *water* to refer to H_2O) and miscommunication (application of the term *water* to a substance other than H_2O is a false application of this term). From a linguistic semantic perspective, to say that H_2O is the meaning/semantics of the word *water* is to say that the word *water* applies to all instances of H_2O regardless of any contextual co-variances – i.e. regardless of whether we are dealing with a rain droplet or a puddle of rain, water in a lake or in a cup, water flowing in a river or from a tap, etc. From this perspective, H_2O as the wide content of the word *water* is more coarse-grained than all such co-variances; it underlies all mind-external co-variances it is acquired from and is cross-context shared/stable. At the same time, however, wide content has to be fine-grained to account for all contextual co-variances and determine the correct

Indeed, the fact that narrow content can also be shared/stable (with the exception of holistic narrow content) shows that shareability qua stability cannot be defining of wide content.

In the next section, I look at the relation between various kinds of mental content posited by philosophers and RT's notions of linguistic semantics and real semantics.

3.1.2 *Philosophical legacy in Relevance Theory's notions of content*

The relation between different philosophical ways of thinking about mental content and the pursuit of linguistic semantics seems to rest crucially on the issue of shareability qua (speaker and context) stability. We have seen that all philosophical notions of mental content, except for holistic narrow content, can be characterised in terms of the notion of shareability qua stability. For this reason, it is perhaps no surprise that the notion of linguistic semantics has been characterised in terms of every kind of mental content which philosophers have argued to be shared/stable. In fact, when approached from a philosophical perspective, it transpires that Relevance Theory's notion of linguistic semantics is characterised in terms of (a) causal-externalist wide content, (b) social-externalist wide content and (c)-(d) two sub-types of non-truth-theoretic narrow content.

As already mentioned (section 2.1), Relevance Theory's distinction between linguistic semantics and real semantics was originally characterised in terms of a truth-theoretic value or lack of it. Now, at the level of a lexical concept, non-truth-theoretic linguistic semantic content may take the form of a logical entry, which is a set of internalist deductive rules[9] or a schema. The non-truth-theoretic linguistic content constituted by analytic relations is compatible with the notion of non-truth-theoretic narrow content. Such content is non-truth-theoretic because it is, by definition, not dependent on and hence non-representational (in Fodor's sense of *representational*, section 1.1) of the mind-external environment. As for schematic non-truth-theoretic narrow content, RT's lexical schemas perfectly match Recanati's (1993: 209) definition of narrow content in terms of on object 'whose contextual enrichment yields a complete representation'. Relevance Theory's characterisation of linguistic semantics in terms of

application of the word *water* in a particular context. For example, when a person A utters '*This water stinks*' when holding up a glass of water, the context is such as to determine that the word *water* applies to the water in the glass, rather than to water in the reservoir that can be seen through the window. The wide content of the thought expressed by this utterance must be much more fine-grained than the mere H_2O. Yet, both – the coarse-grained 'H_2O' and the fine-grained 'H_2O plus impurities in the glass which the speaker is holding up' – are called 'wide' content.

[9] This holds on the assumption of the analytic-synthetic distinction.

logical relations can be found in Carston (e.g. 2002: 9) and as schemas in Carston (e.g. 2002: 375).

However, as argued in 2.1, there is an instability in RT as to whether linguistic semantics is truth-theoretic or not. Thus, Carston (2010) proposes that many lexical concepts are ex full-fledged concepts – by assumption, concepts with Fodor's (1998, 2008) referential content (section 2.1.3.1). Carston's reference to Fodor is important because, if full-fledged lexical concepts are defined in terms of Fodor's concepts, then linguistic semantics is defined in yet another – this time – causal-externalist terms. Indeed, things get even more complicated because RT's, especially Sperber & Wilson's (1998), reference to 'meanings' which are 'encodable in the public language' and Carston's (e.g. 2002: 18) reference to concepts encoded in 'the public language system' point to the characterisation of linguistic semantics in a social-externalist wide sense.

This brief survey reveals that Relevance Theorists, in their project of defining linguistic semantics, are concerned with shareability qua stability, remaining undecided about the truth-theoretic value of such content. From this perspective, Perry's notion of wide content cannot constitute the linguistic content precisely because it is not stable across contexts. This is alarming – since wide content is defined mind-externally and not in terms of stability (as narrow content can also be stable), it follows that both social-externalist and causal-externalist kinds of wide content are compatible with Perry's definition of wide content. Philosophically speaking, Perry's definition applies to social-externalist wide content and causal-externalist wide content. It is only when one focuses on stability (i.e. tries to apply the notion of wide content to the study of linguistic meaning/semantics) that wide content as characterised by Perry is seen as different from social- and causal-externalist notions of content and that the coarse-grained and fine-grained distinction becomes significant.

Now, when theories of communication use wide content, they do so for two purposes. The first is that of playing part in the explanation of utterance interpretation and the second is that of having a role in the explanation of word meaning/semantics acquisition. In principle, the two processes, of word meaning/semantics acquisition and utterance interpretation, are inter-dependent – how one interprets utterances depends in part on what word meaning/semantics one has acquired. Ideally, thus, we should be able to explain the two processes with just one notion of content. There is, however, a methodological assumption that makes the two processes different – whereas both word meaning/semantics acquisition and utterance interpretation take place *in context*, on traditional approaches (e.g. Relevance Theory) the product of word meaning/semantics

acquisition is context-independent. In other words, whatever is acquired is not only acquired in a variety of circumstances (*in context*), but it is also an abstraction from this variety of circumstances (abstraction *from context*). Only coarse-grained content fits such an approach to word meaning/semantics acquisition.

However, as argued earlier, the discrepancy between coarse-grained and fine-grained perspectives on wide content is merely forced upon us by the methodological assumption that there exists some shared/stable content which is abstracted from contextual co-variances. This assumption gives rise to two problems.

The first problem – already discussed in 2.2.3.4, is that it imposes a non-uniform and paradoxical explanation of word meaning/semantics acquisition and utterance interpretation where contextual co-variances are ignored in acquisition only to become significant in utterance interpretation. Perry's definition of wide content, and thus the content's public availability in terms of context dependence is important because it draws attention to the 'evidential' role of (public availability in terms of) the context, which is variable and dynamic. It thus supports my argument (section 2.2.3.4) that contextual information (both incoming and stored in secondary memory) plays the primary evidential role in utterance interpretation.[10] I return to this point in 3.3.

The second problem concerning the discrepancy between coarse-grained and fine-grained perspectives on wide content relates to the way in which the narrow-wide distinction applies to Relevance Theory's shifting positions on mental content. This is what I turn to now.

The distinction between coarse-grained and fine-grained wide content becomes particularly important, and complex, in Relevance Theory. As mentioned, for philosophers, wide content is something which is objective, non-private and truth-evaluable. This holds for wide content as defined by causal-externalists, social-externalists and by Perry (1977). Wide content thus is contrasted with internalist, private and thus non-truth-evaluable content. The wide-narrow distinction is traceable in the original RT's distinction between linguistic and real semantics in terms of truth-theoretic value or lack of it. As long as linguistic semantics is maintained to be non-truth-theoretic and real semantics truth-theoretic, RT's distinction and the wide-narrow

[10] Even though I endorse the idea of public availability in terms of context-dependence (and in spite of its variability), I do not endorse the narrow-wide distinction, to which Perry is committed. I say more about these issues in 3.2 and 3.3.

distinction seem consistent.[11] In this scenario (interpretation 1), the relation between RT's distinction and the wide-narrow distinction is as follows.

INTERPRETATION 1

linguistic semantics ~ real semantics

non-truth-theoretic narrow content ~ causal-externalist wide content

On interpretation 1, RT's notion of linguistic semantics parallels the philosophical notion of non-truth-theoretic narrow content. As such, it may be seen as constituted by either analytic relations or schemas (3.1.1). Given Carston's (2002, 2010) frequent reference to Fodor (1998, 2008), it seems reasonable to assume that real semantics is constituted by Fodor's causal-externalist, i.e. referential content. But this interpretation raises a problem for RT. Whereas non-truth-theoretic narrow content is, by definition, cross-context shared/stable, it does not support truth-evaluation in the way required by philosophers. This is because non-truth-theoretic narrow content is internalist.[12] One may think that this is not a problem for RT because RT has always claimed that it is real semantics that has truth-theoretic value – and, on interpretation 1, this is satisfied by adopting Fodor's causal-externalist wide content. However, as discussed in 2.1, RT substantially deviates from Fodor's theory. Unlike Fodor, RT claims that concepts are more numerous than word meanings/semantics. Indeed, Carston (2010) argues that LOT hosts ad hoc concepts, i.e. concepts which are constructed through personal inference. Ultimately, RT's re-interpretation of Fodor's LOT is so extreme that RT's real semantics can no longer be said to host concepts constituted by Fodor's causal-externalist wide content. This is because RT's real semantics, despite references to Fodor, is not shared but individualistic/holistic. It is the locus of a cogniser's subjective point of view. Given RT's re-interpretation of Fodor's notion of a concept, the relation between RT's distinction and the wide-narrow distinction is as follows (interpretation 2).

INTERPRETATION 2

linguistic semantics ~ real semantics

non-truth-theoretic narrow content ~ individualistic/holistic psychology

[11] However, the notion of non-truth-theoretic linguistic semantics cannot be sustained for the reasons discussed in section 2.1.

[12] But it is not so apparent with schemas – see footnote 7.

Evident here is that the amendments to interpretation 1 leave RT without a notion of content which allows for truth-evaluation (linguistic semantics is non-truth-theoretic and real semantics is private). Given this problem, as well as arguments against non-truth-theoretic semantics discussed in 2.1.1, RT's more recent characterisation of linguistic semantics in truth-theoretic terms seems to be an inevitable move. Indeed, by defining linguistic semantics in terms of full-fledged concepts – by definition, Fodorian causal-externalist content – and, on top of that, by arguing that encoded word meaning/semantics is publicly available to a given speech community (Sperber & Wilson 1998, Carston 2002), RT seems now doubly armed against the problem arising for interpretation 2 (but, as discussed, RT's deviation from Fodor questions RT's causal-externalism). The relation between RT's distinction and the wide-narrow distinction now looks as follows (interpretation 3).

INTERPRETATION 3

linguistic semantics ~ real semantics

causal/social-externalist wide content ~ individualistic/holistic psychology

Construing linguistic semantics in terms of wide content may have solved the problem which arises for interpretation 2; however, it seems at odds with the philosophical legacy. On the one hand, RT has shared causal- and/or social-externalist linguistic semantic content, which can function just as RT theorists want it to – i.e. as a piece of 'evidence' (e.g. Carston 2002: 365) for the thought communicated by a given utterance in a given context. But what about the individualistic/holistic content of real semantics? Well, even though real semantics on interpretation 3 cannot be publicly available because it is private, RT still needs to maintain that its content is truth-theoretic – after all it is THE REAL semantics. To see the full scale of RT's problem here, we need to look at interpretation 3 from the perspective of the distinction between coarse-grained wide content and fine-grained wide content.

INTERPRETATION 3A

linguistic semantics ~ real semantics

causal/social-externalist wide content ~ individualistic/holistic psychology

coarse-grained wide content/ ~ fine-grained wide content
fine-grained wide content

Interpretation 3A represents the current state of affairs in RT. Word meaning/semantics is constituted by some coarse-grained wide content (of 'public language') and by fine-grained full-fledged lexical concepts, i.e. concepts whose content is sensitive to external co-variances (because they are pragmatically constructed). The fine-grained content is publicly available in virtue of its context-sensitivity, but at the same time individualistic because constructed through personal inference. As such, linguistic semantics is both public and private. Indeed, the same conflict is evident in RT's characterisation of real semantics (and was identified in Interpretation 3 as a conflict between privacy and truth-theoretic value of real semantics). Like linguistic semantics, real semantics is private, because individualistic, and wide in the fine-grained sense, because its content is derived in contexts – i.e. it is sensitive to mind-external co-variances relative to which its truth can be evaluated. The problem is that to argue that a semantics is both public and private seems inconsistent.

Now, the private-public inconsistency involved in the notion of fine-grained wide content is (in a way that will become clear in 3.3) unavoidable. I will later argue (and in a sense agree with RT) that there is nothing conceptually wrong in mental content being private/individualistic and publicly available at the same time. Indeed, it seems to follow from Perry's characterisation of wide content: if, as Perry argues, narrow content is a function from context to wide content, then wide content is derived on the basis of (private) narrow content and (publicly available) context.[13] However, the problem I have with Perry's and RT's accounts is that they assume two kinds of content. Notice that if Perry's notion of wide content (i.e. content which is publicly available in virtue of its context-sensitivity) can be applied to holistic content, no other notion of content seems necessary. I return to this point in 3.3.

The issue of RT's instability about the nature of linguistic semantics in particular and mental content is general has been discussed at some length in chapter 2 and in this chapter. All in all, I think that in allowing individualistic content (i.e. full-fledged lexical concepts) into linguistic semantics, RT (e.g. Carston 2010) has made the right choice insofar as it counts as a step towards the conclusion that shared/stable content is not conceptually necessary and that real semantics is the only kind of semantics. In this respect, my criticism of RT is aimed at their, unnecessary, in my opinion, adherence to shared/stable content.

[13] Indeed, Perry's wide content *seems* to match RT's real semantics (with the exception of interpretation 1).

In chapter 2, I argued that, from a linguistic perspective, the notion of shared/stable content – i.e. linguistic semantics – is problematic and unnecessary to account for meaning in language. In the next section, I argue that, also from a philosophical perspective, the notions of shared/stable content – i.e. non-truth-theoretic narrow content, causal- and social-externalist wide content – are problematic. I argue that holism is the only plausible account of mental content and show that the problems envisaged for holism by its opponents (e.g. Fodor & Lepore 1992) do not actually arise.

3.2 All roads lead to holism

In this section, I critically evaluate two-factor approaches to semantics, i.e. approaches which endorse the wide-narrow distinction. I reject all notions of shared/stable content and argue that holism is the most plausible account of mental content.

3.2.1 *Two-factor semantics*

As mentioned in 3.1.1, one way in which narrow content has been defined is in terms of Block's (e.g. 1986) notion of CONCEPTUAL or INFERENTIAL ROLE SEMANTICS[14] (henceforth C/IRS). This treats of inferential relations that a given concept or proposition bears to other concepts or propositions in a cogniser's belief system. The main purpose of positing such internalist content is to provide psychological explanation of a cogniser's behaviour – the assumption being that what is 'in the head' (i.e. a mental state) has functional (e.g. behavioural) consequences (Fodor & Lepore 1992). The C/IRS narrow content, however, is not the only kind of content that Block (*ibid*.) posits. Block's argument is that both narrow and wide kinds of content are needed if certain puzzles about meaning/semantics and reference are to be resolved. Following Putnam (1975), Block assumes that wide content is necessary if the Twin Earth puzzle is to be resolved. However, wide content cannot be all there is to meaning/semantics because it cannot on its own handle the so-called Frege's cases.

Frege (1960) famously drew attention to the fact that if mind-external reference is all there is to meaning/semantics, the substitution of co-referring expressions in a sentence should preserve its truth value. That, however, is not the case in opaque contexts. Consider (1) and (2) below, where expressions *the morning star*, *the evening star* and *Venus* all refer to the same planet in the mind-external world.

[14] Conceptual role semantics is concerned with the content of concepts (i.e. sub-propositional entities), whereas inferential role semantics is concerned with the role a proposition plays in a set of inference patterns (Pagin 2006).

(1) Mary knows that the morning star is Venus.

(2) Mary knows that the evening star is Venus.

In a scenario where Mary knows that *the morning star* and *Venus* are co-referential expressions, but is unaware of the fact that *the evening star* is co-referential with them too, (1) comes out as true whereas (2) comes out as false. Cases like this, where the substitution of co-referential expressions does not preserve the truth value, are abundant (e.g. Cicero=Tully, Superman=Clark Kent, Köln=Cologne, etc.). They all point to the conclusion that something mind-internal has a role to play in the explanation of meaning/semantics. Block's (1986) proposal incorporates both externalist and internalist observations in an approach that can be referred to as a TWO-FACTOR (i.e. internalist/narrow plus externalist/wide) SEMANTICS. This approach came in for severe criticism from Fodor & Lepore (1992).

One of the problems that Fodor & Lepore (1992: 170) have with Block's two-factor approach is that there is nothing in Block's framework that 'keeps the two factors stuck together' – the narrow and wide kinds of content are independent of each other. On the one hand, this independence allows that Oscar's and Toscar's *water* terms in Putnam's scenario have different wide contents despite having the same narrow content. Conversely, it also allows that the terms *the morning star* and *the evening star*, despite having a single wide content, may have two distinct narrow contents (as was the case with Mary in (1) and (2)). On the other hand, this independence worries Fodor & Lepore (1992: 170), who argue that, if wide content and narrow content are independent, nothing in the theory prevents there being an expression that has narrow content appropriate to *4 is a prime* but truth conditions (and thus independent content) appropriate to *water is wet* (Fodor & Lepore 1992: 172). The two-factor approach thus does not resolve what can be called the CONTENT-REFERENCE PROBLEM.[15]

Another problem that Fodor & Lepore (1992) have with two-factor semantics relates directly to holism. As discussed in 3.1.1, C/IRS theories have a choice to constrain narrow content to some set of content-constitutive inferences or not to constrain it and argue for holistic narrow content. According to Fodor & Lepore (1992), both solutions are problematic.

The most common way to constrain narrow content to content-constitutive inferences is to adopt the analytic-synthetic distinction. This immediately brings us back to Fodor's (1998, 2008) criticism of that distinction (discussed in 1.1). Put simply,

[15] As discussed in 1.1, Fodor's referential account faces a similar problem. I return to this point in 3.2.2.

Fodor thinks that C/IRS cannot be defined in terms of the analytic-synthetic distinction because there is no way of deciding which inferences are and which are not content-constitutive. In the light of this, Fodor (*ibid*.) argues that any non-atomistic account of conceptual content inevitably leads to holism. And, according to Fodor, holism is problematic because (a) it does not satisfy the publicity constraint (which states that all humans share the same concepts), and relatedly (b) it does not offer a plausible account of compositionality of thought. Fodor & Lepore's (1992: 176-177) argument is that (holistic) inferential roles are not compositional because they are constituted by whatever beliefs a cogniser happens to have, i.e. they are unconstrained. For example, if someone happens to believe that cows are dangerous – i.e. that an inference to DANGEROUS is part of the content of the concept COW – then an inference from BROWN COW to DANGEROUS should, by compositionality, be part of the content of BROWN COW. But that cannot be the case, so the argument goes, because 'it doesn't look as if the inference from "brown" or "cow" to "dangerous" is compositional' (Fodor & Lepore 1992: 177). This leads Fodor & Lepore (1992: 181) to conclude that C/IRS suffers from a circularity problem: the only way to make C/IRS compositional is to adopt the analytic-synthetic distinction, but since there is no such distinction, C/IRS is an implausible theory because it cannot account for compositionality.

It must be noted, however, that there is also a certain circularity/inconsistency in Fodor & Lepore's criticism of C/IRS; their argument that the inference from BROWN COW to DANGEROUS 'doesn't look' compositional relies on the acceptance of the analytic-synthetic distinction – they actually write that the inference from COW to DANGEROUS is 'clearly synthetic' (Fodor & Lepore 1992: 178).

More fundamentally, however, and regardless of Fodor & Lepore's stand on the analytic-synthetic distinction, their criticism of C/IRS is grounded in the publicity constraint. It is undeniable (and I'm sure Fodor & Lepore would agree) that if there is a COW → DANGEROUS inference in an individual mind, then, all things being equal, this inference may licence the BROWN COW → DANGEROUS inference. The latter inference arises from the fact that the concept BROWN COW has the concept COW as its component, i.e. it is compositionally constituted by it. For Fodor & Lepore (1992), the problem with the subjective/individualistic inference from BROWN COW to DANGEROUS is that it is subjective/individualistic. As Fodor & Lepore (1992) see it, the subjective/individualistic aspect of holism precludes explanation of *the* human mind and so makes the prospects for psychological generalisations/laws doomed to failure.

The crux of the problem is that, given holism, it is very unlikely that two cognisers will ever share a mental state with the same content. I dismiss this criticism in 3.3.

As signposted at the end of 3.1.2, I sympathise with (Block's adoption of) holism but reject the argument that wide content is also necessary. The content-reference problem identified for two-factor semantics by Fodor & Lepore (1992) is genuine. Given that two-factor semantics fails to resolve this problem (acknowledged by Block himself (1993, 1998b)), there are two directions which one may take in aid of resolving it. One may argue that there is only one kind of content and that this content is wide. This position is taken by Fodor (1998, 2008). Alternatively, one may argue that there is only one kind of content and that this content is holistic. This position, endorsed in this thesis, is taken by Bilgrami (1992). In the next section, I discuss problems with Fodor's 'solution' to the content-reference problem and argue that contrary to claims made by Fodor, Fodor's account is implicitly relying on two-factor assumptions. My arguments are supported by claims made in Bilgrami (1998b).

3.2.2 *Fodor's referentialism revisited*

Fodor's (1998, 2008) answer to the content-reference problem is the rejection of narrow content and with it of the wide-narrow distinction. For Fodor, referential (i.e. causal-externalist) content is the only content there is (inferential roles are, by assumption, non-content constitutive).

However, I have argued in 1.1.1 that without internalist semantics, Fodor has no way of explaining why a given concept has the content it does – for example, why the concept CAT locks onto/refers to cats and no other things in the world. Briefly, Fodor's problem was that the content of the concept (e.g. CAT) cannot determine the referential relation (between the concept CAT and cats in the world) because it is the referential relation that determines the content of the concept. This circularity problem, I argued, gives rise to Fodor's (1998, 2008) mind-dependence thesis, which, in turn, commits Fodor to internalist, compositional semantics.

There are two interesting things to notice about this situation. Firstly, Block's and Fodor's accounts are, in fact, not so far apart – given Fodor's (1998, 2008) mind dependence thesis and his re-introduction of internalist semantics, Fodor's account is in effect a two-factor account.[16] Given that, Fodor's (1998, 2008) proposal has to face the content-reference problem just as Block's account does. Secondly, if we combine

[16] Even though Fodor (2008: 86) *explicitly* denies that his account is a two-factor account.

Fodor's mind dependence thesis with his overt rejection of the analytic-synthetic distinction, we simply get a holistic account of what is 'in the mind'. Fodor's proposal (especially as presented in 1998, 2008) thus cannot escape the charge of holism. My argument receives support from Bilgrami's (1998b) observations.

Like any other theory of mental content, Fodor's (1998, 2008) theory is supposed to be equipped well enough to account for the Frege cases (section 3.2.1). We have seen that Block's solution to the problem of the substitution of co-referring expressions in opaque contexts was to posit the existence of narrow content on top of wide content. For Block, the substitution of co-referring terms fails in (1) and (2) because Mary's terms *the morning star* and *the evening star* have different narrow contents. As for Fodor (1998, 2008), despite being *covertly* committed to internalist semantics (as argued earlier), *overtly* Fodor rejects the existence of narrow content. This means that Fodor cannot, *at least overtly*, say what Block does. What Fodor does, however, is argue that this problem has a syntactic solution; the concepts THE MORNING STAR and THE EVENING STAR have different constituent structures and that is why they can behave differently in opaque contexts. For Fodor, 'syntax can do what senses were traditionally supposed to do' (Fodor 2008: 61).

When it comes to primitive concepts, the solution is the same. For Fodor, the concepts CICERO and TULLY have exactly the same (referential) content, but they are syntactically different. For Fodor, it is the assumed difference in the syntax of these two concepts that allows for and explains situations where people believe that Cicero was a great philosopher, but Tully was not. The question that Bilgrami (1998b) asks is what makes the two concepts syntactically different in the first place. Presumably, it cannot be orthography or phonology since concepts are not the sort of thing that can be spelt or pronounced (see also Fodor 2008: 79).[17] Bilgrami (*ibid*.) argues that even on the assumption that syntax solves these puzzles, syntax cannot solve them if it is independent of cognisers' conceptual/inferential roles. He writes:

> **Quote 3**: No amount of talk of the syntax of Mentalese or talk of brain-writing is going to avoid the route to syntax via interpretation of inferential behaviour. So if one insists that it is syntax that will solve the puzzles, then the proper way to proceed is to see functional role as the dog and syntax as the tail, with the former

[17] Furthermore, there exist puzzles where one and the same term has to correspond to two distinct Fodorian concepts, i.e. concepts with the same content but distinct syntaxes. An example is provided by the so-called Paderewski puzzle. Paderewski was a famous Polish pianist as well as Prime Minister. Thus, we can easily imagine a situation where a person who mistakenly believes that Paderewski the pianist and Paderewski the politician were two different people has a further belief that they have met Paderewski the politician, but not Paderewski the pianist in person. Fodor's solution to this puzzle is to argue that this person has two syntactically distinct concepts which share the same content.

> wagging the latter. There is no leaving functional role out, […] Of course, someone might protest that syntax, in general, is surely not to be thought of as exclusively posited on the basis of such things as inference and behaviour, even if causal and inferential role is. That is plausible. But if it is plausible, the lesson to be learnt from it is that one should not appeal to syntax to solve puzzles which are essentially puzzles that are defined upon the failures of rationality in inferences. (Bilgrami 1998b: 111)

Fodor (2008: 87) argues that 'concepts have both referents and a congeries of beliefs (etc.) in which they are embedded', where the latter correspond to a concept's role (including inferential role) in mental processes. He also argues that 'theories about how minds represent things ought to be sensitive to data about which inferences one does (or doesn't) accept' (2008: 77). That said, the relation between concepts (and their formal, i.e. syntactic properties) and inferential roles is for Fodor one-directional; his claim is that 'only *formal* differences among Mentalese expressions can affect mental processes' (2008: 77). Bilgrami's point is that Fodor cannot invoke the route from formal properties to mental processes without acknowledging the primacy of the route from mental processes to formal properties. But once the primacy of mental processes is acknowledged, there is no need to appeal to formal properties in solving the puzzles. If formal properties are nevertheless made use of, one ends up with the formal properties, i.e. the syntax, answering to the whole subjective/individualistic system of inferences. Fundamentally, Bilgrami's argument (1998b) shows that if Fodor insists that inferential roles are not content-constitutive and invokes syntax to solve the puzzles, his notion of syntax is seriously compromised. The paradox is that Fodor's goal of keeping the theory of mind free from individualism/holism has led him to an account where individualism/holism governs the formal properties of the mind.

Now, Fodor (2008: 86) *explicitly* denies that his account is a two-factor account. This is because for him all inferential roles are by definition non-content-constitutive, and referential content is the only kind of content. However, given my arguments in 1.1 and here, as well as Bilgrami's (1998b) criticism of Fodor, Fodor's denial is untenable. Furthermore, his denial is circular – inferential roles are non-content-constitutive because semantics is referential and semantics is referential because inferential roles are non-content-constitutive. My conclusion is that regardless of what Fodor asserts, Fodor's (1998, 2008) theory just *is* a two-factor theory with holistic narrow content (given his rejection of the analytic-synthetic distinction). I argue that Fodor has simply failed to escape the charge of holism.

In the next section, I argue that individualistic, and thus holistic, assumptions underlie the notion of social-externalist content.

3.2.3 *Problems with social-externalist content*

As discussed in 3.1.1, for theorists who posit social-externalist wide content, the misuse of a given word (e.g. *arthritis*) can only be properly explained when we assume the existence of content constituted by community norms – in this case by the way the experts use it. However, perhaps the most obvious objection to social externalism is that it gives rise to the problem of 'incomplete understanding' – according to a social-externalist, an individual is ascribed thoughts composed by concepts of which this individual may have an incomplete grasp (Wikforss 2004).

A related problem is that the alleged social explanations of language use (and misuse) actually have an internalist-individualistic justification. Chomsky (2000a) observes that if a speaker uses, for example, the term *arthritis*, the hearer initially assumes that the speaker's usage is like that of the hearer's. In other words, the hearer assumes that the speaker uses the word to mean the ailment of joints. If the hearer realises that the speaker uses the term to mean something else, the hearer makes modifications to the interpretation as is required by the circumstances. Reference to 'public language' or social-externalist content of the word 'sheds no further light on what is happening between [interlocutors], even if some clear sense can be given to the tacitly assumed notions' (Chomsky 2000a: 32).

Chomsky also discusses some problems with the social co-operation and the role of experts in determining the reference of words. Chomsky shares the intuition that under some circumstances we can agree that what we are referring to when using the word *elm*, for example, is what is meant by some expert. But the expert to whom the speaker is deferring the reference of *elm* may be an Italian gardener who does not speak a word of English but yet is able to correct the speaker's usage through his knowledge of Latin terms. The problem is, Chomsky (2000a: 71) observes, that it is impossible to derive a useful notion of 'language' or 'language community' from such considerations.

Chomsky's internalism-individualism about word meaning/semantics is not incompatible with there being speech communities (in some imprecise sense). As Pateman (1987: 80) comments on Chomskyan general approach to the study of language, 'nativism does not exclude the possibility of a sociolinguistics; what it does do is constrain its possible form'. The point is that because it is impossible to precisely define/delimit the notion of a speech community, the notion is of little theoretical

import in the explanation of the phenomenon we are interested in. Furthermore, an internalist-individualistic approach to word meaning/semantics offers insight into what is happening in cases of 'misuse' of a given word without making the implausible assumption that meaning/semantics is collectively shared by members of the same speech community. If anything happens to be shared, it is shared distributively (Pateman 1987: 91) and even then the distribution is non-uniform across individuals.[18]

Judging language misuse (or correct use) ultimately depends on our assumption that people do things the way we do, i.e. that they use words to communicate what we do in using them (Pateman 1987: 131). The judgement of correct use/misuse happens not in virtue of collective agreement, but distributive agreement, i.e. agreement grounded in internalist-individualistic terms. Burton-Roberts (2012) frames it in terms of 'I-linguistic assumptions, […] amounting to an I-assumption about others' words – effectively, the I-assumption that others implement the same conventions as 'I''.

Now, Chomsky's individualism about words' meaning/semantics is radical, as discussed in 1.2. Since radical individualism – together with the network aspect – is defining of holism (section 3.0), Chomsky's arguments against social-externalism and for an individualistic explanation are, in effect, arguments for a holistic explanation.

The problems with both types of externalism that I have discussed in this section strongly suggest that holism underlies all externalist explanations. For this reason, holism is not merely one of the many possible options – it is, I argue, the only option to be seriously taken into account.

3.3 'Problems' with holism dismissed

In this section, I defend holism against three criticisms it has come under (especially, Fodor & Lepore (1992)). First, I look at the 'problems' of mental generalisations and compositionality. Then, I dismiss the sense-reference 'problem' in the light of Bilgrami's thesis about unity and locality of content. Finally, I discuss how a holistic account is compatible with the claim that meaning is publicly available.

[18] In the unlikely event that it were possible to investigate people's mental contents, conceptual primitive by conceptual primitive, and if we found a single conceptual structure recurrent across individuals in association with a given semiotic label (see my discussion of RT's process of 'concept narrowing' in 2.1.3.2), we could not extrapolate from such an observation to the existence of collectively shared mental content. Relatedly, we could not extrapolate from such finding to the existence of the cognitive process of deterministic decoding of such content (see discussion in 2.2.3.3).

3.3.1 *Holism meets the requirement for mental generalisations and compositionality*

Fodor & Lepore (1992) object that holism, as a thesis, lacks explanatory potential. This problem is exemplified by the inability of holism to state psychological generalisations such as: 'The belief that one is in immediate danger causes release of adrenaline'. The problem identified here is that for holists there is nothing like *the* belief that one is in immediate danger, because each individual's beliefs are subjective/individualistic.

However, Block (1998a) argues that holistic accounts of mental content are able to provide universally quantified laws, i.e. laws that generalise about contents without specifying them. He gives the following example of a universally quantified psychological law: 'For any action *a* and any goal *g*, if one wants *g* and also believes that *a* is required for *g*, then one will try to do *a*'. This quantified psychological law has been stated without committing oneself to the existence of two agents having exactly the same goals.

Another argument in favour of such quantified laws is related to my criticism of Fodor's atomicity (in 1.1.2). Fodor's assumption that his theory allows for psychological generalisations like 'The belief that one is in immediate danger causes release of adrenaline' is based on his claim that the concept BELIEF is atomic and species-shared. The problem for Fodor – and for the explanatory potential of his psychological statements – is that (as I argued in 1.1.2), Fodor's atomism, on which such generalisations rely, is theoretically and methodologically implausible.

Contrary to Fodor's (1998, 2008; with Lepore 1992) claims, I argue that certain general psychological *laws* can be deduced by observing instances of subjective/individualistic inference patterns. For example, my earlier observation (3.2.1) that the fact that the existence of the subjective/individualistic COW → DANGEROUS inference licences the subjective/individualistic BROWN COW → DANGEROUS inference is subsumed under the following generalisation.

(3) $\forall x\,(Y(x)\ \&\ (z \subseteq x)) \supset Y(z)$

(3) states that for all things *x* such that *x* has the property *Y* and *z* is a subset of *x*, it follows that *z* has the property *Y*. Accordingly, if someone believes that all cows are dangerous and that brown cows are a subset of the set of all cows, it follows that this person believes that brown cows are dangerous. Given (3), we can predict an infinite number of diverse subjective/individualistic inferences: from BROWN COW to DANGEROUS or CUDDLY or APPETIZING, from LITTLE POODLE to SCARY or

FRIENDLY or INCAPABLE OF RECURSIVE THOUGHT, etc. Similarly, we can predict that, all things being equal, a subjective/individualistic inference from COW to APPETIZING will not licence the inference from NOT COW to NOT APPETIZING, just as the subjective/individualistic inference from LITTLE POODLE to FRIENDLY will not licence the inference from NOT LITTLE POODLE to NOT FRIENDLY. This fact can be subsumed under the following generalisation.

(4) $\sim[[\forall x(C(x)) \subseteq \forall y(A(y))] \supset [\exists z(\sim C(z)) \supset \sim A(z)]]$

(4) states that it is not the case that if all things *x* which have the property *C* (e.g. cow) form a subset of all things *y* which have the property *A* (e.g. looks appetising), then if there is a thing *z* such that *z* does not have the property *C*, then *z* does not have the property *A*. In simple terms, (4) predicts that if all cows look appetizing but not all things that look appetizing are cows, then, if something is not a cow, it does not necessarily mean that it does not look appetising.

Furthermore, we can also predict that (4) will not hold if there is a logical relation of equivalence between the two variables, *x* and *y*. For example, it will not hold for inferences from THE PRESENT QUEEN OF ENGLAND (*x*) to THE PRESENT SUPREME GOVERNOR OF THE CHURCH OF ENGLAND (*y*). In cases of logical equivalence like that, the negation of one variable follows from the negation of the other and vice versa. This rule can be subsumed under the following formula, which is licensed by the truth-functional properties of logical equivalence.

(5) $(x \equiv y) \supset (\sim x \equiv \sim y)$

Just like the previous generalisations, so (5) predicts and explains a variety of inferences, including ones that are peculiar to particular cognisers. For example, (5) predicts that if someone happens to believe that the natural phenomenon of winds blowing at a speed of more than 40 kilometres per hour is always concurrent with the pressure drop by 230 hectopascals, then this person is bound to infer from the occurrence of one to the occurrence of the other (and vice versa) and also from the non-occurrence of one to the non-occurrence of the other (and vice versa).

The point I am making here is that the probability or improbability of certain inference patterns with subjective/individualistic content can be predicted by reference to the laws of logic. This shows that holism about mental content is absolutely compatible with positing psychological generalisations, grounded in general rules of

logic, as laws underlying general human inferential patterns. I argue that given the availability and explanatory power of such laws, Fodor's (1998, 2008; with Lepore 1992) argument that holistic accounts hinder psychological generalisations and therefore thwart prospects for robust and explanatorily powerful psychological theories is unjustified.

A further, related, aspect of holism that Fodor & Lepore (1992) object to has already been mentioned in 3.0 as 'the network effect'; the conceptual role of a mental symbol is such as to relate it more or less directly to many other (and though indirectly perhaps all) mental symbols within the same system, making the symbols mutually dependent for content. As mentioned in 3.2.1, Fodor & Lepore object to the network aspect of holism by arguing that it does not offer a plausible account of compositionality of thought.

Fodor & Lepore (1992: 182-183) argue that compositionality is normally determined *locally*; for example, the idea that syntax is compositional is the idea that the structural description of an expression is determined by the '[…] lexical items that it contains'. The problem for them is that, even if holism is compatible with some sort of compositionality, such compositionality is not determined *locally* but *globally*, i.e. by features of one's whole belief system.

However, the 'problem' of global compositionality does not arise if one allows for immediate pragmatic constraints on inference.[19] As discussed in 2.2.3, Hintzman's (1986) context-dependent process of echo retrieval constrains pragmatic inference so that not all memory traces stored in secondary memory (i.e. not whole holistic system) resonate strongly enough to make contribution to the echo which is retrieved in the process of utterance interpretation (and which, for Hintzman, constitutes the interpretation of a given utterance). Thus, Hintzman's account clearly shows that *not all* subjective/individualistic content is relevant to the retrieval of context-sensitive conceptual structures (i.e. thoughts). Add to this the fact that compositionality is a phenomenon which operates on contextually sensitive concepts (as convincingly argued by Recanati (2005), section 2.2.3.3), Fodor & Lepore's (1992) objection to the *global* nature of compositionality involved in holistic accounts is, again, ill-founded.

The point is that a holistic account of mental content is compatible with and only with a *local* account of compositionality as long as it is combined with a radically contextualist account of utterance interpretation – like the one I argued for in 2.2.3. On a

[19] That is, the 'problem' arises if one – like Fodor and Lepore – fails to acknowledge the role of pragmatic constraints.

radically contextualist account – which presupposes holistic mental content – compositionality *does not* operate globally, i.e. it does not combine all imaginable inferential roles of a given concept, because all imaginable inferential roles of a given concept are constrained by contextual factors (those which are incoming and those which are stored in secondary memory). This, very roughly, means that if we are sitting in a restaurant eating steaks and I am telling you that I love beef so much that to me even the look of a cow grazing in a meadow is mouth-watering, I am not communicating to you, and you are not taking me to be communicating to you, all my (somewhat inconsistent) concerns about cruelty to animals in slaughterhouses. That is because, even though in my mind there are inferential routes from COW to APPETIZING, from COW to CRUELTY THAT ANIMALS ARE SUBJECTED TO IN SLAUGHTERHOUSES, and from COW to MILK, the second and third inferences are not constitutive of the thought I am communicating to you in that particular situation. Thus, contrary to what Fodor & Lepore (1992) argue, holism does not commit one to a global approach to compositionality. Indeed, holism implies that compositionality is a context-sensitive process (which is compatible with Recanati 2005).

In this section, I have dismissed two arguments that Fodor & Lepore (1992) have wielded against holism: (i) lack of prospects for psychological generalisations and (ii) commitment to global compositionality. There is one remaining issue that needs to be discussed. We have seen that both Block's and Fodor's accounts face the sense-reference problem. In the next section, I argue that the content-reference problem does not arise on a holistic account. My argument is grounded in Bilgrami's (1992) thesis about the unity and locality of content.

3.3.2 *Bilgrami's unity and locality of content*

In this section, I introduce Bilgrami's (1992) unity and locality of content thesis. I discuss its relevance to the sense-reference 'problem' and explain how Bilgrami's thesis allows for public availability of meaning. However, I identify certain tensions in Bilgrami's thinking about concepts and argue that they can be resolved by adopting the Representational Hypothesis' account of semantics and meaning.

3.3.2.1 *Unity and locality of content*

Bilgrami's (1992) unity of content thesis states that there is just one kind of content – holistic content. Crucially, however, Bilgrami argues that the holistic content as such

(i.e. content at the network level) is not involved in the explanation of behaviour (including linguistic behaviour). What it does is provide 'a pool of resources' for selective use in the interpretation of behaviour of others. Bilgrami (e.g. 1992: 10, 142-144) calls the holistic level the 'meaning-theoretic', 'aggregative' or 'idiolect' level.

Bilgrami's (*ibid*.) locality of content thesis states that holistic content is constrained at the 'local' – i.e. context-specific – level and that it is this locally constrained content which plays a role in the explanation of behaviour.

The distinction between aggregative and local levels can be illustrated with the word *water*. Let us assume that in the mind of a person A, the word *water* is associated with some phenomenological properties, complex beliefs about its chemical composition, the feeling of quenching thirst, the fear of drowning, etc. But in the mind of a chemically ignorant person B, the word *water* is associated only with phenomenological properties, the feeling of quenching thirst, the pleasure of swimming, etc. Bilgrami's locality of content thesis states that in the context where a chemically ignorant person (B) and a chemically aware person (A) are drinking water from the tap to quench thirst, or when A is asking B for a glass of water, only the beliefs about quenching thirst properties of water will be relevant to the explanation of their behaviour. In this locality, A's beliefs about the chemical composition of water or A's associations of the term *water* with the fear of drowning are not selected from the specifications of the aggregative level. Thus, even though A's and B's beliefs associated with the term *water* are necessarily different at the aggregative (holistic) level, they may (and often do) converge at the local level.

Bilgrami's locality of content thesis offers a perspective on the 'problem' raised for the network and individualist aspects of holism – i.e. the criticism that 'no two people are ever likely to have the same concepts since they will almost certainly associate different beliefs with their respective terms' (Bilgrami 1992: 10). Bilgrami (1992: 147) argues that this sort of criticism follows from the mistaken assumption that content at the holistic level plays a role in the explanation of behaviour.

However, a particular shortcoming of Bilgrami's account is that he (1992: 12, 143) underestimates the importance of the aggregative level and suggests that, because this level summarises all individualistic beliefs associated with a given term, it is 'of no specific interest or use at all', that it has 'no psychological reality' and that it is merely a 'trumped up theoretical posit'. Perhaps this is why Bilgrami (1992: 145) worries that his account may be criticised for not offering a 'clear principle' on how the local content is 'distilled' from the aggregative level.

I disagree with Bilgrami on this point and argue that the locality of content is not plausible unless it is acknowledged that the aggregative level is psychologically real. This is because the aggregative level is essentially involved in determining/constraining local contents. Crucially, I argue that Bilgrami's aggregative level finds parallels with Hintzman's (1986) trace aggregate level and, furthermore, that Hintzman's process of echo retrieval offers a 'clear principle' for how Bilgrami's local content is selected from the aggregative level.

From Hintzman's perspective, the locality of content thesis is satisfied by the context-sensitive process of echo retrieval. As discussed in chapter 2, on Hintzman's account, the interpretation of acoustic stimuli is a wholly pragmatic process of retrieving a conceptual structure (i.e. the echo) via the matching of the incoming contextual information (i.e. the probe) with multiple traces stored in secondary memory (i.e. the trace aggregate level). This means that for the echo – which can be understood as Bilgrami's local content – to be retrieved, there has got to be the trace aggregate level to match the incoming stimuli with. This suggests that Bilgrami's aggregative level is much more than a 'trumped up theoretical posit' – in fact, it is necessary for the retrieval of local content.

3.3.2.2 *Bilgrami's externalism and public availability of private thought*

Bilgrami (1992) argues that his locality of content thesis requires a special kind of externalism. This special kind of externalism, he goes on to argue, should allow for public availability of thought but, emphatically, should not amount to causal- or social-externalist kinds of wide content, which Bilgrami rejects as implausible.

Bilgrami (1992: 4, 200) argues that some form of externalism is necessary to explain how we are able to understand one another. In other words, it is necessary to allow for the public availability of belief and meaning. External environment is important, Bilgrami (1992: 6-7) argues, insofar as 'one fixes concepts' – that is, attributes concepts to other people (e.g. 1992: 7, 160) – by correlating the terms in the utterances, in which concepts are 'expressed', with saliencies in the environment in which these utterances are made. It must be pointed out, however, that for Bilgrami (1992: 134), 'there is no way to think of concepts being individuated by external objects, except as constrained by an agent's beliefs'. Accordingly, Bilgrami's externalism amounts to the claim that the agent's perceptions of (beliefs about) mind-external environment are essential in attributing concepts to other agents. For Bilgrami (e.g. 1992: 160), the mind-external world plays an *evidential* role in that concepts are

attributed to agents by observing correlations between the uses of a particular term by agents and perceptions of external items in the context of these utterances. Thus, the sense in which Bilgrami's (e.g. 1992: 160, 200) externalism allows for public availability of private thought – i.e. for 'one mind's access to another's mind and its contents' – is in virtue of it being *communicated in context*. Because this 'access to another's mind' relies on context-dependence, it must, I assume, consist in pragmatic inference. Therefore, (perceived) mind-external context correlated with the use of a given term constitutes the evidence for pragmatic attribution of concepts, i.e. for inferring the intended interpretation of one's interlocutor's communicative behaviour.

Bilgrami (e.g. 1992: 7) is careful to point out that finding the right mind-external saliencies to correlate with an agent's use of a given term has to 'fit in with' the other contents one has already attributed to a given agent. Notice that for this to work, Bilgrami has to acknowledge the importance, and the evidential role, of the aggregative level: if the current contextual information 'fits in with' other contents attributed to a given agent, the contents attributed in the past must be stored somewhere. This 'somewhere' is the aggregative level and the 'fitting in', I argue, happens through Hintzman's (1986) process of echo retrieval. When the importance of the aggregative level is recognised, Bilgrami's notion of public availability (or evidentiality) in terms of context-dependence corresponds to my argument (2.2.3.4) that contextual information (both incoming and stored in secondary memory) constitutes primary, indeed the only, evidence for intended interpretation.

Now, Bilgrami's characterisation of public availability is reminiscent of Perry's (3.1.1) characterisation of wide content, and thus the content's public availability, in terms of its context dependence (and regardless of the notion of shareability qua stability). Perry's wide content is dynamic – i.e. it co-varies/changes with context – and, at the same time, it is publicly available precisely because dependent on the mind-external world, which (unlike someone else's thought) is available to all of us for scrutiny. Of course, Bilgrami's account differs from Perry's in that Bilgrami rejects the problematic (as argued in 3.1 and 3.2) narrow-wide dichotomy. This rejection has serious consequences for the sense-reference problem which, as discussed in 3.2, arises for two-factor accounts (including Fodor's, despite his claims to the contrary).

Bilgrami (e.g. 1992: 134, 140) argues that what 'individuates concepts' is not the mind-external world but agents' conceptions of things (i.e. their beliefs). Indeed, Bilgrami (e.g. 1992: 32-33, 134) denies that the notion of reference has got 'anything to do with meaning and content'. His externalism is not the claim that the mind-external

world provides, i.e. is constitutive of the content for concepts. Bilgrami's externalism is the claim that the mind-external world provides evidence for inferring what the (non-referential) content is. Thus, Bilgrami does not abandon externalism, but in rejecting reference restrains its function (Bilgrami 1992: 139-140). Having rejected reference, Bilgrami's account effectively dissolves the sense-reference problem: there is no such problem, because reference is not content-constitutive (or, more precisely, there is no reference).

3.3.2.3 *Holism and concepts*

Overall, I am sympathetic to Bilgrami's argument that the unified/holistic content is dependent on external environment but not constituted by it. Much of my agreement, however, depends on what really is the locus of externalism and holism and what exactly is meant by the term 'concept'.

Given that for Bilgrami it is the local level which bears explanatory potential, his 'concepts' seem to be the context-dependent entities which are taken to be communicated by the speaker on a given occasion of use. Using RT's terminology, Bilgrami's concepts are ad hoc.[20] Thus, when Bilgrami talks of 'individuation of concepts', what he has in mind is an inference from contextual information to concepts taken as communicated in a particular communicative situation. For Bilgrami, thus, concepts are temporal mental objects, dynamically constructed on the basis of context and a cogniser's holistic belief system (i.e. the aggregate level). Bilgrami's externalism about 'individuation of concepts' is not a thesis about conceptual content but about the inferring (attributing) of communicative intention – it is a relation between an uttered term and communicated concepts which is externally determined.

The problem, for Bilgrami, is that he does not seem to be consistent in the distinction between concepts and their content (which *is not* the domain of externalism) and the inferring (attributing/individuating) of concepts (which *is* the domain of externalism). This is evident from Bilgrami's stance on innateness and internalism. Consider the following quotes:

> **Quote 4**: What, in the evolutionary story, explains this fact about innate conceptual structure that the species is endowed with. There can be no answer to

[20] This interpretation seems to be right especially in the light of Bilgrami's (1992: 12) argument that the locality thesis 'dissolves the very idea of content composed of context-invariant concepts'. He (*ibid*.) argues that the locality thesis offers 'a way out of the dogmas of a certain way of thinking which leave us on the one hand with a false distinction [wide-narrow] and, on the other, with an artificially tidy picture of the mind'.

> this other than that which invokes adaptability to an external *environment*. […] Externalism, in the sense I have been insisting on, is, therefore, a core, an essential component, of the explanation that even the *nativist* is committed to. It is this core which is a necessary condition for the possession of these concepts. So even if nativism is true, externalism is still in the picture. (Bilgrami 1992: 224-225)

> **Quote 5**: I must admit internalism still gives me bad nights […] internalism is a hidden commitment of the widespread belief in the necessity and viability of a notion of narrow content […] (Bilgrami 1992: 243)

Bilgrami's externalism – as presented in relation to the locality thesis – does not concern concepts and their content but the inferential ascription of communicated concepts (let me call it 'ascriptive externalism'). Indeed, Bilgrami (e.g. 1992: 134) emphasises that ascriptive externalism is restrained by an agent's beliefs about the mind-external world – the mind-external world determines 'concept individuation' insofar as it is the mind-external world as the agent perceives it (believes it to be). However, externalism as discussed in quote 4 concerns concepts and their content (let me call it 'content externalism'). Bilgrami argues that if there is innate conceptual structure, it is there as an evolutionary response to the mind-external environment. In this sense, ontogenetic innateness phylogenetically presupposes externalism. This shows that Bilgrami is also concerned with content externalism. In fact, even though content externalism *is* compatible with nativism (quote 4), Bilgrami never explicitly acknowledges that, overall, his externalism is indeed very internalist.[21] This is because, as Bilgrami explains in quote 5, internalism is too strongly associated with the narrow content and thus with the wide-narrow distinction, which he rejects. This worry of Bilgrami's is also noteworthy because the wide-narrow distinction – especially as presented by Putnam, Burge, Block and Fodor – is a thesis about content and not ascription. If Bilgrami's worries about internalism concern the wide-narrow distinction, then his worries about internalism concern conceptual content. The problem is that defining conceptual content as externalist but not referential and nativist but not internalist effectively leaves it undefined. Is there a way to resolve this externalist-nativist schism?

Indeed, notice that even though Bilgrami (e.g. 1992: 6) distinguishes between concepts, on the one hand, and 'terms in the utterances' which 'express' concepts, on the other, Bilgrami uses 'concept' to talk of the 'counterpart' to 'term'. In the light of

[21] However, Bilgrami (2002) overtly sympathises with Chomsky's (2000a) internalist-individualistic philosophy.

Bilgrami's distinction between concepts and terms, his use of 'counterpart' suggests that even though concepts are not terms (are not equivalent to terms), the notion of a concept is evoked when the notion of an uttered term is evoked. This peculiar form of isomorphism may be, at least partly, responsible for the externalist-nativist schism.

The nature of Bilgrami's problem becomes clear when we recall Burton-Roberts' (2012) signification precondition (section 2.2.2), which is a logically necessary requirement that, in any signification relation, a signified must exist prior to and independent of the fact of signification. In the linguistic context, this means that if an acoustic event ('uttered term') is to function as a sign for a concept, the concept must exist prior to and independent of the fact of signification. Thus, the Representational Hypothesis' unambiguous distinction between linguistic signs and concepts highlights the lack of any necessary dependence between terms/linguistic signs and concepts.

If concepts were merely 'counterparts' to terms, as Bilgrami seems to suggest, there would be no place for concepts, and thus thoughts they compose, outside the communicative context. I do not think this is what Bilgrami has in mind, but it emphasises the point that defining concepts with respect to terms is problematic.

When concepts are unambiguously distinguished from the way they might be said to function in language (i.e. from the way they are associated with linguistic signs), it becomes clear that the nativist stance about concepts and their content is compatible with an externalist stance (in the sense of Bilgrami's ascriptive externalism) about meaning as defined in the Representational Hypothesis (2.2.2). Having acknowledged the signification precondition and the Representational Hypothesis' notion of meaning-as-relation, it becomes possible to see that Bilgrami's externalism is not a thesis about concepts, but about meaning.

Along the lines argued in chapter 1 and 2, and in line with the Representational Hypothesis, I propose that primitive concepts (i.e. Jackendoff's quarks, Hintzman's primitive properties) as well as the (compositional) capacity to combine them into structured concepts are innately specified. Together, primitive concepts and the compositional capacity constitute LOT. In this sense, the range of structured concepts which *can be* entertained[22] (and assigned a semiotic label) is innately pre-determined. However, the range of structured concepts which *are* in an actual fact entertained (and

[22] I do not use the term 'entertain' in the sense of Perry (1977). Rather, an entertained concept is a concept that has a function in a mental life of an individual.

associated with a given semiotic label) is delimited by an individual's experience and psychology[23].

Given the distinction between concepts and the way they function in language, it also transpires that holism (especially the network effect, on which below) is not a thesis about concepts and their content but about the associations between concepts and semiotic labels. Indeed, such an unambiguous distinction between concepts and the way they function in language sheds light on the analytic-synthetic distinction. Clearly, the analytic-synthetic distinction is impossible to maintain in the domain of associations between concepts and semiotic labels as these are holistic. But this may not be true of concepts in and of themselves. If primitive concepts as well as the compositional capacity are innately pre-determined, any structured concept is analytic, whereby analyticity is defined in terms of the concept's compositional structure (i.e. its content).[24]

However, much remains to be said about such a construal of analyticity. Notice that if we assume, as is normally assumed, that analyticity is to be defined in terms of the concept's compositional content (its content-constitutive relations), the analyticity that we get is rather weak in the sense that it does not presuppose shareability qua stability. Any concept which composes any of our thoughts (communicated or not) comes out as analytic in the sense that its structure is constrained by a given locality. Therefore, it may be argued that what we are talking about here is not analyticity, but simply (local) compositionality. Notice also that analytic relations such construed are non-holistic in the sense that they are locally constrained and thus do not give rise to the network effect. However, it is clearly not the case that such analytic relations are not individualistic. It follows then that it is the network aspect of holism, but not radical individualism that distinguishes concepts from the way they are associated with semiotic labels. Indeed, this conclusion seems to follow naturally from the distinction I made earlier in this section between innately predetermined range of structured concepts which *can be* entertained (and assigned a semiotic label) and the experientially determined range of structured concepts which *are* in an actual fact entertained (and associated with a given semiotic label). Important though the issue of analyticity is, I do not expand on it here, but leave it as an area for further research.

[23] Though, of course, given that the range of concepts is innately pre-determined, individual experience will be constrained by the range of concepts with which the world can be experienced.
[24] The claim that analyticity pertains to concepts in and of themselves is compatible with my interpretation of Chomsky's stance on analyticity (1.2). But, of course, it is not compatible with Chomsky's double-interface assumptions.

However, a conclusion which can be drawn at this point is that whereas semantic (i.e. mental) content at the aggregative level is holistic, meaning (in the sense of Burton-Roberts) is local because it is contextually and thus externally (in the sense of Bilgrami) constrained to a degree which, all things being equal, allows for successful communication.

3.4 Conclusion

This chapter has engaged with two related issues. The first concerned philosophical legacy in Relevance Theory's notion of linguistic semantic content. The second, more general, issue concerned the question of whether any philosophical notion of shared content can withstand the challenge of holism.

As for the first issue, I argued that RT's introduction of individualistic content (i.e. full-fledged lexical concepts) into linguistic semantics is inconsistent with their own view of linguistic semantics as shared/stable. However, I also argued that RT's introduction of individualistic content into linguistic semantics is a right move insofar as it is a move towards a more radical conclusion that shared/stable content is in fact unnecessary to explain meaning in language.

As for the second issue, I argued that all, wide or narrow, notions of shared/stable content are problematic. Having dismissed three criticisms of holism – the 'problem' of mental generalisations, the 'problem' of compositionality, and the sense-reference 'problem' – I maintained, in line with the arguments presented in chapter 2, that holism is the most plausible account of mental content.

I further argued that Bilgrami's unity and locality of content thesis, with its specific form of externalism, offers a compelling account of public availability of a personal thought in virtue of its being communicated in context. Bilgrami's account of public availability of thought is important (especially in the context of RT's introduction of individualistic content into linguistic semantics) as it effectively makes the wide notion of content conceptually unnecessary.

Finally, I showed that the externalist-nativist tension present in Bilgrami's proposal is due to lack of consistency in distinguishing between concepts and the way concepts function in language. I argued that approaching the externalist-nativist issue from the perspective of the Representational Hypothesis, not only resolves the tension, but it also demarcates the separate (though related) domains of holism (i.e. aggregative level relations between semiotic labels and concepts) and analyticity (i.e. concepts, whether communicated or not).

The arguments presented in this chapter and in chapter 2 make a strong case *against* a traditional two-step (deterministic decoding and then pragmatic inferring) account of utterance interpretation and *for* a wholly pragmatic, wholly inferential account of utterance interpretation – interpretation which is primarily (i.e. immediately) constrained by a personal (i.e. holistic) inference about the communicative intention of the speaker in a given conversational context.

Importantly, an RH-based wholly inferential account which I propose endorses nativist assumptions about conceptual content (if only in the light of arguments presented in chapter 1 and here). Thus, it combines identity of LOT (i.e. primitive concepts and compositional capacity) among the human species with Bilgrami-type public availability of the conventional way in which linguistic signs are used to signify conceptual structures, which are allowed by shared LOT.[25] In this sense, it is both the shareability of LOT among our species and Bilgrami-type notion of public availability of communicated concepts, which allow for successful communication without linguistic semantics. This conclusion ends the theoretical part of the thesis.

In the second part, I investigate whether my theoretical arguments can and should be applied to linguistic analysis. In particular, I investigate whether the arguments put forward so far apply to the analysis of the relation between the linguistic sign *if* and material implication. In chapter 4, I argue that the claim that *if* semantically encodes material implication cannot be maintained even when supported by pragmatic explanation. I argue that material implication only ever applies in pragmatically inferred thought. In chapter 5, I offer a wholly inferential account of the interpretation of *if*. This account has certain implications for the explicit-implicit distinction.

[25] As discussed earlier, Fodor's rejection of holism is motivated by the publicity constraint which seeks to account for the identity of concepts. Since for Fodor concepts just are word meanings/semantics, his publicity constraint effectively seeks to account for the identity of mental entities which are theoretically and empirically (see discussion of cross-linguistic data in section 1.1.2) dubious. My point is that even though publicity constraint is important, Fodor's version of it applies to implausibly defined concepts.

PART II

Chapter 4. Does *if* encode Material Implication?

4.0 Introduction

In chapter 2, I argued that it is impossible and unnecessary to posit a specifically linguistic semantics and that the process of deterministic decoding of such content is redundant in accounting for meaning in language. In chapter 3, I argued that philosophical notions of shared content (i.e. non-truth-theoretic narrow content, causal-externalist wide content and social-externalist wide content), which have been posited as linguistic semantics, should be rejected. I further argued that holistic content is the only kind of mental content there is. In this chapter, I discuss and then critically evaluate a more specific claim (e.g. Grice 1989, Noh 2000): that material implication is the encoded – i.e. linguistic – semantics of the word *if*. I have chosen this particular item for its significance to the distinction between (linguistic) semantics and pragmatics. Even though the claim that *if* semantically encodes material implication has been controversial in philosophy, semantico-pragmatic (i.e. two-step) accounts of the interpretation of *if* (e.g. Grice 1989) have taken some heat off that controversy. For that reason, semantico-pragmatic accounts of the interpretation of *if* may be thought of as one of the central achievements of pragmatic theory (e.g. Mauri & van der Auwera 2012).

However, I argue that it is impossible to maintain the claim that *if* semantically encodes material implication for two reasons. The first is that certain conditional beliefs involve radical deviations from material implication. Because such deviations occur at the level of belief, they cannot be explained away in terms of the conversational principles which are appealed to in aid of defending the claim that *if* semantically encodes material implication (4.2.1). The second reason concerns the problem of pragmatic intrusion into encoded semantics (4.2.2). This problem, I argue, arises with Grice's (1989) and Relevance Theory's (e.g. Noh 2000) accounts of conditionals. Both problems point to the conclusion that conditional beliefs and (or, rather, thus) the interpretation of conditionals is a matter of a wholly pragmatic and thus holistic inferential process.

For brevity, I use the term 'encoded semantics' to mean linguistic semantics encoded by a given expression and thus deterministically decoded in the process of interpretation of an uttered expression.

4.1 Material Implication as the encoded semantics of *if*

In 4.1.1, I introduce the logical operator of material implication and then, in 4.1.2, explain why it has been argued that material implication constitutes the encoded semantics of *if*. The discussion focuses on indicative conditionals. The main reason for this limitation is that, as I intend to show, the variation found in the interpretation of this group of conditionals suffices to illustrate my argument that semantic content is holistic and accessible solely through pragmatic inference. I use '⊃' for material implication, '&' for conjunction, 'v' for disjunction, '~' for negation, '≡' for equivalence and '├' for entailment.

4.1.1 *Facts about Material Implication*

The logical operator of material implication '⊃' connects two propositions *p* (the antecedent) and *q* (the consequent) such that the truth value of the complex proposition (*p*⊃*q*) is a function of the truth values of the propositions it operates on (*p, q*). Material implication (henceforth MI) is a truth-functional operator, i.e. *nothing* other than the truth values of *p* and *q* are required to compute the truth value of (*p*⊃*q*). The truth table for MI is given in (1) below.

(1)

	p	**q**	**p ⊃ q**
1.	T	T	T
2.	**T**	**F**	**F**
3.	F	T	T
4.	F	F	T

This truth table states that (*p*⊃*q*) is false (F) in situation 2, where the antecedent is true and the consequent false, and true (T) in all other situations. Thus, it follows that MI can be defined in terms of negation and conjunction as well as negation and disjunction in the following way.

(2) (a) (*p*⊃*q*) ≡ ~ (*p*&~*q*)
(b) (*p*⊃*q*) ≡ (~*p*) v *q*

As illustrated by (2b), the falsity of *p* or the truth of *q* entails the truth of the implication ([~*p* ├ (*p*⊃*q*)]; [*q* ├ (*p*⊃*q*)]).

The truth table for MI licences two inferences: Modus Ponens and Modus Tollens, represented in (3) and (4) respectively.

(3) Modus Ponens

1. $p \supset q$
2. p
3. q

(4) Modus Tollens

1. $p \supset q$
2. $\sim q$
3. $\sim p$

According to Modus Ponens, if ($p \supset q$) and p is T, it follows that q is T. According to Modus Tollens, if ($p \supset q$) and q is F, it follows that p is F.

From the truth table for MI nothing follows about the truth value of p from the truth of q or about the truth value of q from the falsity of p. This means that if ($p \supset q$) and q is T, the inference to the truth of p is invalid and that if ($p \supset q$) and p is F, the inference to the falsity of q is invalid. These invalid inferences, which are called the Fallacy of Affirming the Consequent (FAC) and the Fallacy of Denying the Antecedent (FDA) respectively, are represented in (5) and (6).

(5) FAC

1. $p \supset q$
2. q
3. p

(6) FDA

1. $p \supset q$
2. $\sim p$
3. $\sim q$

4.1.2 *Truth-functional semantics of 'if'*

It has been argued (e.g. Grice 1989, Smith & Smith 1988, Noh 2000) that MI constitutes the encoded semantics of natural language connective *if*. In this section, I explain why such claims have been made.

The aim of truth-functional approaches to conditionals (e.g. Grice 1989) is to show that the truth-conditions of '*if... then*' sentences are of a truth-functional kind. If this can be shown, the semantics of conditional statements can be explained in terms of MI.

Crucially, the claim that MI constitutes the encoded semantics of *if* rests on the assumption that any deviation from MI in the interpretation of conditionals can be

explained in terms of pragmatic principles. In the Gricean tradition, any such pragmatic deviation should be amenable to explanation in terms of a conversational Cooperative Principle.

I first introduce Grice's distinction between semantics and pragmatics as well as his Cooperative Principle (4.1.2.1). In 4.1.2.2, I discuss examples whose interpretation seems to be predicted by MI in a straightforward way. In 4.1.2.3, I look at Grice's analysis of some non-truth-conditional aspects of meaning of utterances of *if* (his Indirectness Condition) in terms of conversational implicature. Section 4.1.2.4 deals with biconditional (*if and only if*) interpretations of conditionals and 4.1.2.5 with a relevance theoretic account of metarepresentational uses in conditionals. This is followed by an interim summary in 4.1.2.6.

4.1.2.1 *Grice's semantics and pragmatics*

Grice's (1989) distinction between semantics and pragmatics can be defined in terms of the following list of dichotomies (adapted from Burton-Roberts 2005).

Semantics	~	Pragmatics
What is said	~	What is implicated
Explicit	~	Implicit
Context-free	~	Context-sensitive
Truth-conditional	~	Non-truth-conditional

For Grice, semantically encoded content is the truth-conditional content of a word and any expression composed of words. In uttering an expression, the speaker commits herself to the expression's semantically encoded content (WHAT IS SAID); this means that such content cannot be cancelled without (creating a) contradiction, i.e. without contradicting what the speaker is committed to in uttering that expression. For Grice, semantic content (i.e. what is said) is context-invariant and it is EXPLICITLY communicated. In contrast, pragmatics deals with non-truth-conditional, context-dependent aspects of 'meaning'. Because of the non-truth-conditional relation of what is pragmatically inferred to what is linguistically encoded (i.e. explicit), pragmatically communicated aspects of 'meaning' can be cancelled without contradiction. Pragmatically communicated aspects of 'meaning' are not said/explicated but IMPLICATED by the saying of what is said/explicated. Consider the following example:

(7) a) Tom got up *and* read a paper.

b) Tom got up *and then* read a paper.

c) Tom got up and read a paper, *but not in that order*.

For Grice, the word *and* semantically encodes the logical functor of conjunction. Accordingly, (7a) semantically encodes and hence explicitly communicates only that both conjuncts are true. Any suggested temporal (or causal) link between the two conjuncts (7b) is non-truth-conditionally linked to what is explicitly communicated and so can be cancelled without contradicting what is said/explicated (as in (7c)). For this reason, the temporal link (7b) is, for Grice, implicitly communicated – it is an IMPLICATURE.

The implicature communicated by the use of *and* in (7) is classified by Grice as a Generalised Conversational Implicature (GCI). This kind of implicature arises generally, i.e. regardless of the particularities of a given context (Grice 1989: 37, Levinson 1983: 126). Nevertheless, it is context-dependent in the sense that it can be cancelled in a particular context, as shown in (7c).

Another type of conversational implicature proposed by Grice is a Particularised Conversational Implicature (PCI). PCIs are implicated by exploiting the contexts in which utterances that give rise to them occur – were it not for the particular context, these implicatures would not arise. Consider (8).

(8) a) A: Are you hungry?

B: I've had breakfast.

b) *B has had breakfast today*.

c) I've had breakfast – *in my life, but not today*.

d) *B is not hungry.*

e) I've had breakfast, *but actually I'm still hungry*.

B's utterance in (8a) gives rise to the implicature in (8b) – given that breakfast is a daily occurrence, when people say they have had breakfast they are generally taken to communicate that they have had breakfast on the day of the utterance. Because the implicature in (8b) would arise even in the absence of A's question in (8a) – i.e. particular context is not necessary for it to arise – it can be classified as a GCI. This GCI is cancellable, as shown in (8c). But B's answer in (8a) can also give rise to the PCI in (8d). It is a PCI because it can be derived only in the particular context of A's question – take away A's question and the PCI does not arise. Because PCIs are pragmatically

derived, they too can be cancelled without contradicting what was said/explicated, as shown in (8e).

The question that needs to be answered now is how conversational implicatures arise in the first place. To answer this question, Grice (1989) proposes that in conversation there is a tacit agreement between interlocutors to cooperate. This tacit agreement is explained by Grice in terms of a Cooperative Principle, a set of four Maxims – Quantity, Quality, Relation and Manner – which speakers exploit (by obeying or flouting them) in order to give rise to implicatures.

The maxim of Quantity concerns the quantity of information to be provided in conversation. It is divided into the following two sub-maxims:

i. Make your contribution as informative as required (for the current purposes of the exchange)
ii. Do not make your contribution more informative than is required[1]

The maxim of Quality states to make one's contribution true and its two sub-maxims are:

i. Do not say what you believe is false
ii. Do not say that for which you lack adequate evidence

The maxim of Relation states to make one's contribution relevant and the maxim of Manner to be perspicuous. This last maxim is further subdivided into:

i. Avoid obscurity of expression
ii. Avoid ambiguity
iii. Be brief (avoid unnecessary prolixity)
iv. Be orderly

Given Grice's Cooperative Principle, the derivation of the GCI in (7) (*Tom got up and read a paper*) can be explained in terms of the maxim of Manner. By saying 'S_1 *and* S_2' the speaker of (7) is implicating *and then* because the speaker is assuming that the hearer takes her to be observing the maxim of Manner and thus will be able to infer that the speaker is presenting the events in a orderly way.

As for the PCI in (8), it can be explained in the following way. Speaker B in (8) (*A: Are you hungry?; B: I've had breakfast*) is implicating that she is not hungry by

[1] Grice (1989: 26-27) suggests that the second sub-maxim is disputable because its transgression merely amounts to 'a waste of time' or because its effect can be explained by the maxim of Relation.

exploiting the maxim of Relation – B assumes that A takes her to be observing the maxim of Relation and will thus be able to infer that B's answer is relevant to A's question.

In the following sections, I discuss Grice's (1989) claim that *if* semantically encodes the logical functor MI and the related assumption that any deviation from MI in the interpretation of conditionals can be explained in terms of the Cooperative Principle.

4.1.2.2 *Simple cases*

Let us look at the extent to which natural language conditional sentences correspond to MI. Consider (9) below (taken from Edgington 2008).

(9) If it's a square, it has four sides.

Let us refer to the *if*-clause as the 'antecedent', the matrix clause as the 'consequent' and the complex statement they together constitute as a 'conditional'. Now, if the word *if* is truth-functional, we should be able to predict the truth value of the conditional from the truth values of the antecedent and the consequent. For (9), this prediction seems to be borne out. If both the antecedent and the consequent are true, the conditional is true. If the antecedent is true but the consequent is false, the conditional is false (it is impossible for a square not to have four sides). If the antecedent is false, the conditional is true regardless of the truth value of the consequent (if something is not a square, it may or may not have four sides). Consider also conditional (10), whose consequent is patently false, and (11), whose antecedent is necessarily true.

(10) If Tom can drive a truck, then I am Mickey Mouse.

(11) If two plus two equals four, then my client is innocent.

In using (10), the speaker is communicating that she does not believe in Tom's ability to drive a truck. This interpretation can be modelled by MI; on the assumption that the speaker of (10) is thereby saying something true (abiding by Grice's maxim of Quality) and given that the consequent is patently false, the antecedent has got to be false if (10) is to be true. As for (11), the speaker here is communicating that it is certain that her client is innocent. This interpretation too can be modelled by MI; on the assumption that the speaker of (11) is saying something true and given that the antecedent is necessarily true, it follows from the truth table for MI that the consequent has got to be true as well;

that it is as certain that the speaker's client is innocent as it is certain that two plus two equals four.

4.1.2.3 *Indirectness Condition*

Conditionals are generally used to communicate more than ($p \supset q$). For example, in (9) there seems to be a consequential relation between *x* being a square (*p*) and *x* having four sides (*q*) such that the truth of *q* can be inferred from the truth of *p*. Grice (1989: 58) refers to this extra, non-truth-functional, aspect of 'meaning' associated with conditionals as the INDIRECTNESS CONDITION and points out that this condition can be formulated in different ways: 'that *p* would, in the circumstances, be a good reason for *q*', 'that *q* is inferable from *p*', 'that there are non-truth-functional grounds for accepting $p \supset q$'.

In order to maintain the claim that semantics of *if* is truth-functional, Grice needs to show that the Indirectness Condition is not part of the semantics of *if*, but that it is pragmatically inferred. In particular, Grice (1989: 60) argues that the Indirectness Condition is a GCI – it arises generally, but it is cancellable. Consider (12).

(12) If you put that bit of sugar in water, it will dissolve.

Grice (*ibid*.) argues that the Indirectness Condition GCI can be cancelled by adding a clause such as '*though so far as I know there can be no way of knowing in advance that this will happen*'. The addition of the cancellation clause, he goes on to argue, marks (12) as a guess or prophecy.

Another reason for analysing the Indirectness Condition as an implicature is that it is not always present. Grice offers the conditional clause in (13) as an example.

(13) Perhaps if he comes, he will be in a good mood.

There are two possible interpretations of (13). On the first one, the Indirectness Condition is present and the speaker is, in Grice's terms, implicating that *p* will be a reason for believing *q*. On the second interpretation, there is no Indirectness Condition implicature and the speaker is simply communicating something to the effect of '*he will happen to be in a good mood*' – *p* merely materially implies *q* and there is no other relation between *p* and *q* beyond consistency.

But how does the Indirectness Condition arise (when it is present)? Grice (1989: 61-62) argues that to say/explicate the very general ($p \supset q$) is not normally taken to be

informative enough. Thus the saying/explicating of ($p \supset q$) gives rise to the assumption that the speaker is violating the maxim of Quantity[2]. That, however, is counterbalanced by the assumption that the speaker is doing so because she is obeying the maxim of Quality, which states that one should have adequate evidence for what one is saying. The evidence in question is the Indirectness Condition. Grice's (*ibid.*) proposal thus is that the saying/explicating of ($p \supset q$) implicates that there is some evidence for the saying of it, i.e. that there is some non-truth-functional relation between the antecedent and the consequent.

Based on such arguments, Grice argues that the Indirectness Condition is not part of the semantics of *if* and that the semantics of *if* is truth-functional. However, Grice (1989: 77) concedes that 'generally or at least in special contexts [it might be] impossible for a rational speaker to employ the conditional form unless, at least in his view, not merely the truth-table requirements are satisfied but also some strong connection [i.e. the Indirectness Condition] holds'. I return to this point in 4.2.

4.1.2.4 *Conditional perfection*

Another common interpretation of conditionals which is stronger than ($p \supset q$) is the biconditional (i.e. if and only if) interpretation. Consider the following examples:

(14) If you pass the exam, I'll take you to a restaurant.
(15) If you do something illegal again, we'll lock you up.

It may be argued that in uttering (14) the speaker is not actually (semantically) making any undertaking as to what she will do if the hearer fails the exam, only what she will do if the hearer passes the exam. This would make the interpretation of (14) compatible with the truth table for MI. According to MI, the promise is not broken if there is a meal following a passed exam (situation 1), if there is a meal despite a failed exam (situation 3 – the speaker did not say what she would do if the hearer failed the exam) and if there is no meal because of a failed exam (situation 4). The promise is only broken if the exam is passed but there is no meal (situation 2).

The same analysis is available for (15), where the conditional is used to make a threat. It can be argued that the speaker of (15) is not actually saying what she will do if the hearer refrains from doing something illegal, only what she will do if the hearer

[2] The Indirectness Condition implicature also seems to involve the assumption of the violation of the maxim of Relation. Grice (*ibid.*), however, does not mention the maxim of Relation in this context.

does something illegal again. This leaves open a possibility that the speaker will lock the hearer up even though he does not do anything illegal.

However, as pointed out by Geis & Zwicky (1971: 652), the human mind has a tendency to 'perfect conditionals to biconditionals'. In other words, 'a sentence of the form X ⊃ Y invites an inference of the form ~X ⊃ ~Y' – an inference modelled by equivalence. This tendency for CONDITIONAL PERFECTION is manifest in examples like (14) and (15).

It has been argued that the phenomenon of conditional perfection (also known as BICONDITIONAL STRENGTHENING, i.e. strengthening of *if* to *iff*) can be analysed in terms of Grice's conversational implicature (Comrie 1986, van der Auwera 1997). In particular, van der Auwera (1997) argues that it can be analysed in terms of Grice's scalar quantity implicature, a kind of a GCI.

Scalar implicatures involve semantic scales of entailment – this is why they are predictable generally, i.e. without reference to a particular conversational context. For example, '*p and q*' is on a semantic scale with '*p or q*' such that '*p and q*' entails '*p or q*'. This semantic scale allows one to predict that by saying '*p or q*' the speaker is conversationally implicating that, as far as the speaker knows, the stronger claim does not obtain, i.e. that 'it is not the case that *p and q*'.

Since the phenomenon of conditional perfection involves, it is assumed, a strengthening of material implication ($p \supset q$) to equivalence ($p \equiv q$), and since ($p \equiv q$) entails ($p \supset q$), the scale involving these two logical functors is the first one to investigate. However, the problem for the assumption that the [($p \equiv q$) $\vdash$ ($p \supset q$)] scale is involved in conditional perfection is that the use of *if* in examples like (14) and (15) does not communicate the negation of the stronger *if and only if* (henceforth iff) by which it is entailed, but the assertion of it. This has led Levinson (e.g. 1983: 146-147) to suggest that at work here is a principle of informativeness, which in some circumstances allows speakers to implicate more than what is explicitly communicated. This principle is linked to the second sub-maxim of Quantity, which states that one should not make one's contribution more informative than required.

However, van der Auwera (1997: 267) criticises the principle of informativeness for its triviality; in principle, it should be at work whenever what is implicated involves what is said/explicated plus more. Van der Auwera (1997: 262) argues that the strengthening of *if* to *iff* can be explained in terms of Grice's scalar GCI with reference to the following scale:

(16) …

if p, q and if r, q and if s, q

if p, q and if r, q

if p, q

In this scale, the higher assertions entail the lower ones. Thus, by saying *if p, q*, one is implicating that the conjunction of *if p, q* with another conditional which involves a different antecedent (*r*) but the same consequent (*q*) does not obtain. In other words, in saying *if p, q*, one is implicating that the truth of *q* does not follow from any other antecedent but *p*; *q* is T iff *p* is true. In van der Auwera's (1997: 262) words, 'when one supplies only the one sufficient condition *p*, one conversationally implicates that there is no second – and no third, etc. – sufficient condition'.[3, 4]

However, there is a problem with the assumption (van der Auwera's and Levinson's) that the iff interpretation can be explained with reference to a semantic scale. Scalar inference depends on the elements which form a scale being lexical items (e.g. Levinson 1983: 133), as is the case with *and* and *or*, *all* and *some* or *excellent* and *good*. But even though equivalence entails MI, equivalence is not lexicalised in English[5] (or any other language). Thus, the biconditional interpretation cannot be explained with reference to semantic scales.

Smith & Smith (1988) argue that the biconditional interpretation can be easily explained in relevance-theoretic terms. As discussed earlier, the challenge for truth-functional approaches to the encoded semantics of *if* is that MI predicts that the consequent will be true irrespective of the truth of the antecedent. Smith & Smith's (1988) solution is to use Relevance Theory's (e.g. Sperber & Wilson 1995) notion of relevance, which is calculated in terms of positive cognitive effects and processing effort (as discussed in chapter 2). Smith & Smith (1988: 333) claim that in saying '*if p, q*', the speaker is implying that ($\sim p \supset \sim q$) for the following reason: if the hearer believed that ($p \supset q$) and that ($\sim p \supset q$), she should have said "q" to spare the hearer unnecessary

[3] Even though van der Auwera (1997) explains conditional perfection in terms of Grice's GCI, van der Auwera (1986) argues that *if* semantically encodes more in a way of relation between *p* and *q* than offered by a truth-functional account. For van der Auwera (1986), *if* semantically encodes a non-truth-functional sufficiency condition (that '*p* is a sufficient condition for *q*'). However, because the effect of van der Auwera's (1997) conditional perfection GCI can be stated with reference to the sufficiency condition (one implicates that there is no other sufficient condition) or without reference to it (one implicates that the truth of *q* does not follow from any other antecedent but *p*), I assume that there is no relevant difference between the two approaches in this respect. See also footnote 12.

[4] In other words, this scale is supposed to account for the implicature that a given sufficient condition is in fact necessary (see e.g. Horn 2000).

[5] The closest English gets to equivalence is *if and only if*.

processing effort. Because the speaker did not simply say "q", the hearer assumes that *p* must be interpreted as relevant to *q*.[6] Smith & Smith (1988: 333) claim that this pragmatic explanation of biconditional strengthening makes it possible to defend the view that MI is the encoded semantics of *if*. I return to this point in 4.2. and in chapter 5 (where I discuss the relevance of *p* to *q* in more detail).

It is important to point out that the uses which involve the iff interpretation do not form a homogenous group. For example, Comrie (1986: 78) argues that the iff interpretation is natural with prohibitives, of which the threat in (15) is an example. However, as exemplified by (14), this interpretation is also found with promises[7]. Furthermore, some uses of conditionals to communicate general truths (so-called 'universal conditionals') may also involve the iff interpretation, as in (17), whereas others do not, as in (18)

(17) Water begins to boil if it reaches 100°C.
(18) If one is a man, one cannot get pregnant.

The difference between (17) and (18) is that whereas in (17) the falsity of *q* follows from the falsity of *p* (Denying the Antecedent is a licit inference here), in (18) nothing about the truth value of *q* follows from the falsity of *p* (Denying the Antecedent is fallacious here). To complicate things more, (19) (due to Horn 2000: 319) and (20) are examples of conditional injunctions involving the iff interpretation.[8]

(19) If thy right eye offend thee, pluck it out.
(20) If there is no other choice, amputate his leg.

But not all conditional injunctions involve the iff interpretation. Consider (21).

(21) If you want to make your parents happy, quit smoking.

In (21), unlike in (19) and (20), the speaker does not seem to be communicating that the truth of *q* does not follow from any other antecedent but *p* – only that ($p \supset q$).

[6] A similar proposal is advocated by Horn (2000: 309-310), who argues that, because the conditional assertion of *if p then q* contains more information than the assertion of *q*, *p* must be taken as relevant. According to Horn, what makes a condition relevant is its (communicated) necessity.
[7] In chapter 5, I will actually argue that not all conditional promises (and, in fact, not all conditional threats) involve the iff interpretation.
[8] Horn (2000) criticizes Dancygier & Sweetser's (1997) account on which biconditional is built into the encoded semantics of predictive conditionals. As Horn's (19) shows, non-predictive conditionals can also receive the biconditional interpretation.

In summary, the iff interpretation cannot be explained by reference to Levinson's informativeness principle (which invokes the [$(p \equiv q) \vdash (p \supset q)$] scale) because of its triviality (as argued by van der Auwera 1997). Neither can it be explained with reference to van der Auwera's (1997) scale, presented in (16). This is because equivalence is not lexicalised.[9] Smith & Smith (1988) argue that it can be explained with reference to Relevance Theory's notion of relevance. However, the iff interpretation is not restricted to any homogenous group of uses of conditionals. Thus, it remains to be explained under what conditions this interpretation arises.

4.1.2.5 *Relevance Theory on metarepresentational uses of conditionals*

The claim that MI constitutes the encoded semantics of *if* has been challenged by what has been referred to as 'speech act domain' (Sweetser 1990) conditionals or 'metarepresentational' (Noh 2000, subsuming Horn's (1989) 'metalinguistic') conditionals. Consider (22) (due to Noh 2000: 199).

(22) [The door bell is ringing.]
Mary to Jane: If that's John, I'm not here.

Given that Mary is in the location referred to by *here*, MI incorrectly predicts that the conditional in (22) should be true where the antecedent and the consequent are true. Relatedly, in (22) it is impossible to infer the truth of *q* from the truth of *p* or the falsity of *p* from the patent falsity of *q*. This means that the inference rules of MP and MT, which are licensed by MI, do not apply to this example.

As observed by Noh (2000: 178-179), conditionals with so-called 'given' antecedents, i.e. antecedents which are recoverable from context or known/accepted by the speaker and hearer, also often resist a truth-functional analysis. Consider (23).

(23) A: Two and eleven makes thirty.
B: If two and eleven makes thirty, you need more work on maths.

The problem posed by this example is that the antecedent of B's answer in (23) is false. The truth table for MI predicts that if the antecedent is false, it does not matter whether the consequent is true or false for the whole utterance to be true. However, Noh (1996, 2000: 189) observes that this simply is a wrong interpretation because B in (23) clearly

[9] A more fundamental problem with van der Auwera's account is discussed in footnote 12.

is communicating that the hearer needs more work on maths, i.e. that the consequent is true.

However, Noh (1996, 2000) argues that the relevance-theoretic notion of metarepresentation actually allows one to retain MI as the encoded semantics of *if* even in examples like (22) and (23).

In Relevance Theory, the notion of metarepresentation involves the use of one representation to represent another, e.g. an attributed utterance or thought, which it resembles (e.g. Noh 2000: 185, 186). Following Sperber and Wilson (1995), Noh (2000: 73) argues that resemblance is to be understood in terms of sharing analytic and contextual implications in a given context. Metarepresentation may resemble the metarepresented object in terms of content or form. Uses which exploit resemblance of content are called INTERPRETIVE USES, and those which exploit resemblance of form are called METALINGUISTIC USES. Consider examples below, taken from Noh (2000).

(24) [The material in brackets represents American pronunciation of *tomatoes*]
A: I like tom[eiDouz]
B: If you like tom[eiDouz], you must be from America.

(25) A: I loved her.
B: If you loved her, why didn't you come to the party?

(26) [A is talking to B about her meeting with her supervisor]
A: Then what did she say?
B: The argument is invalid.

Example (24) involves a metalinguistic antecedent; here B does not metarepresent the content of A's utterance but its pronunciation[10]. This means that the attribution of pronunciation is included in the proposition expressed in the antecedent: the proposition expressed is something like '*If you like things you call 'tom[eiDouz]*'' or '*If you say you eat 'tom[eiDouz]*'' (Noh 2000: 190). Noh (2000: 187-189) argues that (25) and (26) are interpretive uses. In (25), the antecedent of B's conditional metarepresents the proposition expressed by A's utterance. More specifically, Noh (2000: 187) argues that the antecedent metarepresents A's utterance including the higher-level explicature of '*if you say/believe that …*'. (26) is similar to (25) in that it is an interpretive use, but different in that it does not involve the interpretation of A's utterance but the previous

[10] Other metalinguistic uses involve metarepresenting morphology, word stress, etc.

utterance(s) of the speaker's supervisor. Noh (1996, 2000) argues that reanalysing certain conditionals in terms of attributive metarepresentational use provides a more explanatory and uniform account of their semantics and pragmatics. In fact, she argues that a metarepresentational approach allows one to maintain MI as the encoded semantics of all uses of *if*.

Importantly, Noh (2000: 189, 203, 211) argues that the truth value of a metarepresentational use is not established with respect to the truth value of what is metarepresented but with respect to the *faithfulness* of the metarepresentation to the metarepresented content or form. For example, B's utterance in (26) is true if and only if it is 'a faithful enough interpretation of what [B's] supervisor said' (Noh 2000: 189).

Noh's point is best illustrated by example (23) (*A: Two and eleven makes thirty; B: If two and eleven makes thirty, you need more work on maths*). As discussed, the truth-functional relation between the antecedent and the consequent of (23B) cannot hold at the descriptive level – i.e. between the content of '*two and eleven makes thirty*' and '*you need more work on maths*' – because it does not explain how B is communicating that the consequent is T. On Noh's (2000: 187) metarepresentational account, however, the antecedent in (23B) is used to metarepresent A's utterance including the higher-level explicature of '*if you say/believe that …*'. At that meta-level, the antecedent is in fact true: whereas '*two and eleven makes thirty*' is false '*you say/believe that two and eleven makes thirty*' comes out as true. According to the truth table for MI, if the antecedent is true, the consequent has got to be true for the conditional to be true. For Noh, this explains why B in (23) is communicating that the consequent is T and allows maintaining the truth-functional semantics of *if*.

Similarly, Noh (2000: 199) argues that a metarepresentational treatment in terms of MI is available for (22) (*If that's John, I'm not here*). In this example, it is the consequent, not the antecedent, which receives a metarepresentational analysis. Noh (2000: 199) argues that the consequent in (22) is metarepresenting 'an utterance which the speaker wants the hearer to make in the situation described by the antecedent'. This, according to Noh (2000: 204), allows one to treat the relation between the proposition '*that's John*' and the higher-level explicature '*you tell him I'm not here*' in truth-functional terms; the hearer of (22) is being requested to make the conditional true by making sure that *q* is T in a situation in which *p* is T.

However, there are problems with Noh's metarepresentational analysis. I discuss them in 4.2.2.5.

4.1.2.6 *Interim summary*

The goal of this section was to discuss reasons behind the claim that the word *if* semantically encodes MI. To this end, I first introduced Grice's distinction between semantics and pragmatics and his Cooperative Principle (4.1.2.1). In 4.1.2.2, I discussed certain conditionals whose acceptability, it seems, can be explained in terms of MI without any reference to the notion of implicature. I then looked at certain interpretations which deviate from MI (Indirectness Condition in 4.1.2.3 and conditional perfection in 4.1.2.4) and discussed how such deviations are explained in pragmatic terms by reference to Grice's Cooperative Principle. Finally, in 4.1.2.5, I discussed a relevance theoretic extension of a truth-functional analysis to metarepresentational uses.

In the next section, I argue that the claim that MI constitutes the encoded semantics of *if* is impossible to maintain.

4.2 Why *if* does not encode Material Implication

In this section, I argue that there are two general reasons why the claim that *if* encodes MI cannot be maintained. In 4.2.1, I discuss Edgington's (2008) observation that certain conditional beliefs deviate from MI. In 4.2.2, I discuss the problem of pragmatic intrusion into encoded semantics. I argue that contrary to claims made in Carston (2000: 99-100), Relevance Theory's re-analysis of Grice's explicit-implicit distinction does not resolve this problem.

4.2.1 *Deviations from Material Implication at the level of belief*

In 4.1.2.3 and 4.1.2.4, I discussed how, in the Gricean tradition, deviations from MI in the interpretation of conditionals have been explained in conversational terms. Some explanation has also been offered within the framework of Relevance Theory (Smith & Smith 1988). For example, one of the problems for truth-functional approaches is the logical fact that, since $\sim(p \supset q)$ is equivalent to $(p \& \sim q)$, $\sim(p \supset q)$ entails the truth of its antecedent. Inconsistently with this logical fact, when someone says that 'It's not the case that if the peace treaty is signed, war will be avoided', they are not communicating that the peace treaty will be signed. Thus, facts about what is communicated by the use of a conditional are at odds with the facts of logic.

Now, Smith & Smith (1988: 334) argue that there is a simple relevance-theoretic solution to this problem. For them, a speaker who wanted to use $\sim(p \supset q)$ with 'the conjunctive force' of $(p \& \sim q)$ would not be observing the principle of relevance: it

puts the hearer to unjustifiable effort. For this reason, Smith & Smith (*ibid*.) argue that the use of ~($p \supset q$) would not normally be interpreted as the assertion of what they call 'the troublesome conjunction'.

However, Edgington (2008) observes that certain deviations from MI in the interpretation of conditionals – including 'the troublesome conjunction' – cannot be explained in terms of conversational principles because they take place at the level of belief. The problem for an MI analysis arises due to the fact that, in thinking about the world, people often accept and reject propositions with different degrees of certainty.

Edgington (2008) considers a scenario in which a person A does not believe that 'If the Republicans win, income tax will double' and does not believe that 'the Republicans will win'. In other words, A rejects both ($p \supset q$) and p. In rejecting these, A accepts both ~($p \supset q$) and ~p, which are logically contradictory. Now, one of the logical facts about MI is that its truth is entailed by the falsity of its antecedent: $\sim p \vdash (p \supset q)$. Since $\sim p \vdash (p \supset q)$, MI predicts that a disbelief in Republicans' winning (i.e. ~p) entails a belief that if they win, income tax will double (i.e. ($p \supset q$)). But this is not the case with A's beliefs; inconsistently with MI but nevertheless rationally, A can be almost certain that Republicans won't win (i.e. certain that ~p), and find it unlikely that if they win, they will double income tax (i.e. certain that ~($p \supset q$)). The logical facts about MI are at odds with what a rational person can consistently believe or disbelieve – this cannot be explained in terms of conversational principles.

Similarly, Smith & Smith's (1988) relevance-theoretic proposal does not explain why a cogniser may, contrary to the logic of MI but nevertheless rationally, not believe that the peace treaty will be signed and at the same time not believe that if the peace treaty is signed, war will be avoided.[11]

The problems discussed above arise in virtue of one of the paradoxes of MI, namely that the falsity of the antecedent is sufficient for the truth of the implication (hence, $\sim p \vdash (p \supset q)$).[12] In fact, this paradox is an example of a more general paradox of

[11] Nor can this fact be explained away in terms of Gricean Conversational Implicature.

[12] This problem arises for truth-functional as well as non-truth-functional accounts of the encoded semantics of *if* (Edgington 2008). On non-truth-functional accounts, the conditional operator is modal (e.g. van der Auwera 1985: 184), which means that the truth value of a conditional cannot be predicted solely from the truth values of the simple propositions on which it operates. Non-truth-functionalists agree with truth-functionalists that when the antecedent is T but the consequent F, the conditional is F (situation 2 in the truth table for MI). However, they do not agree that the truth value of a conditional can be logically deduced in other situations. For example, van der Auwera (1985) requires some consequential relation (corresponding to Grice's Indirectness Condition) to hold between p and q for *if p then q* to be T. Also, non-truth-functional approaches, like truth-functional approaches, make no prediction as to the truth value of the consequent when the antecedent is F (i.e. [$\sim p \vdash (p \supset q)$] is valid on

MI, namely that a falsity or a contradiction materially implies any and hence every proposition.

Consider the following scenario (taken from Edgington 2008). A person B is sitting in front of a box. B knows that inside the box is a geometric figure. B can rationally assume that 'it's not the case that if it is a pentagon, it has six sides': i.e. that ~($P \supset S$). The problem arises when B is told that the figure inside the box is not a pentagon: i.e. that ~P. B believes that ~($P \supset S$) and that actually ~P. Logically, to believe ~P and ~($P \supset S$) is contradictory because ~($P \supset S$) is equivalent to (P&~S) which entails P. The problem for the applicability of MI to B's rational patterns of thought is that since (~($P \supset S$) & ~P) is contradictory, it can entail anything, including ($P \supset S$) – logically, (~($P \supset S$) & ~P) $\supset$ ($P \supset S$) is a valid formula. Thus, MI predicts that someone who believes of an unseen geometric figure that it is not the case that if it is a pentagon, it has six sides, but that actually it is not a pentagon may be wrong because it may be the case that if it is a pentagon, it has six sides.

Consider also the logical fact that ~(p&q) $\equiv$ ($p \supset$~q). Edgington (2008) imagines a scenario in which a person C is almost certain that it is not the case that C will be hit by a bomb and injured today. But at the same time C thinks it highly unlikely that if C is hit by a bomb, C will not be injured. In other words, C consistently and rationally believes that ~(p&q) and that ~($p \supset$~q). The problem is that ~(p&q) and ~($p \supset$~q) are logically contradictory. As a matter of logic, C should not be judged a rational thinker.

The next problem pointed out by Edgington (*ibid*.) concerns the logical paradox that all conditionals with T consequents are T. Logically, the truth of q is sufficient for the truth of ($p \supset q$): $q \vdash (p \supset q)$. The problem arises when a person D is almost certain, but not one hundred percent certain that q is T. For example, D is almost certain that Ed is in Rome (R) right now and rejects the conditional that if Ed's plane has crashed on its way to Rome, Ed is Rome now (D rejects $C \supset R$). The problem for a truth-functional approach is that D rejects (i.e. believes it is F) a conditional ($C \supset R$), which according to MI is T. MI predicts that as long as the consequent is T (and D is almost certain that it is), the whole conditional is T – including a situation in which it is T that Ed's plane has crashed on its way to Rome. Since logically $R \vdash (C \supset R)$, how can one be certain that R, but find it unlikely that ($C \supset R$)? This logical prediction, Edgington

---

both approaches). Consequently, this problem with conditional beliefs arises for non-truth-functional approaches just as it does for truth-functional approaches.

(*ibid*.) argues, is incompatible with rational patterns of thought and the deviation of such rational patterns of thought from MI cannot be explained in conversational terms.[13]

Edgington (2008) also discusses the application of MI to conditional commands. Consider the following example:

(27) If the patient is still alive in the morning, change the dressing.

Since $(p \supset q) \equiv (\sim p \vee q)$, the command to make (27) true is logically equivalent to 'Make it the case that either the patient is not alive in the morning, or you change the dressing'. Given that a disjunction is T if at least one of the disjuncts is T, a nurse may argue, on logical grounds, to have obeyed the command by making sure that the patient is not alive in the morning ($\sim p$ is T) and not changing the dressing (q is F).

Edgington (*ibid*.) observes that a truth-functional theorist may appeal to pragmatics in the resolution of the problem with conditional commands; it may be argued that the hearer of (27) 'tacitly understands' that killing the patient is not a reasonable or intended way to make the truth-functional conditional true. Edgington dismisses such a pragmatic explanation as unconvincing on the grounds that it is 'stretching pragmatics rather far'. In 4.2.2, I strengthen Edgington's opinion by arguing that pragmatic accounts stretch pragmatics too far to be consistent with the claim that MI constitutes the encoded semantics of *if*.

4.2.2 *Pragmatic intrusion into encoded semantics*

As discussed in 4.1.2.1, Grice's semantics-pragmatics distinction can be defined as follows:

- Semantics concerns the content of what is said/explicitly communicated; it is context-free and truth-conditional
- Pragmatics concerns the content which is not said but implicated (i.e. implicitly communicated); it is context-sensitive and non-truth-conditional with respect to what is said (the semantics)

[13] From another perspective, B accepts (R) and rejects ($C \supset R$). To reject ($C \supset R$) is to accept $\sim(C \supset R)$, which equals accepting ($C \& \sim R$). However, to accept both R and ($C \& \sim R$) is to have logically contradictory beliefs.

However, Grice's distinction between semantics and pragmatics and hence between explicit and implicit communication is problematic. In what follows, I discuss some problems with Grice's distinction and its relevance to the study of conditionals.

4.2.2.1 *Grice's circle*

Burton-Roberts (2005) argues that Grice's notion of what-is-said, and hence of what is explicitly communicated, is vague with respect to two types of saying: A-SAYING and B-SAYING. A-saying is the saying-of-"s", i.e. it concerns the words uttered. B-saying is the saying-that-*p*, i.e. it involves an assessment of the explicitly communicated thought. Consider the following examples:

(28) It's twelve.
(29) It's cold.
(30) Jest zimno. [Polish translation of *It's cold*]

In uttering (28), two people can A-say the same thing (i.e. utter the words "It's twelve"), but B-say different things; e.g. one can B-say that it is midday and the other that it is midnight. Conversely, different A-sayings represented in (29) and (30) can be used to B-say the same thing, i.e. to explicitly communicate the thought that it is cold.

The problem for Grice is that to determine what is B-said, one needs to do a lot of pragmatic work, including reference assignment and disambiguation (Grice 1989: 25). Consider (31) below.

(31) She took it.

In order to determine what the speaker of (31) B-said, the hearer needs to pragmatically derive (i.e. derive by reference to the context) the referents of *she* and *it* as well as the particular concept communicated by the word *take* (e.g. ACCEPT as in *take the money* or SWALLOW as in *take the pill*).

This Gricean problem of 'pragmatic intrusion' into what is semantically encoded – often referred to as 'Grice's circle' (e.g. Levinson 2000: 186) – has important consequences for Gricean accounts of the semantics and pragmatics of conditionals.

In 4.1.2.3, I discussed Grice's analysis of a causal-consequential or inferential relation between the antecedent and the consequent (Grice's Indirectness Condition) in terms of a conversational implicature. Importantly, I pointed out Grice's (1989: 77) acknowledgement that the use of conditionals by a rational speaker is *motivated by* the

speaker's assumption that there is some 'strong' connection (i.e. the Indirectness Condition) – and not merely the relation of MI – between the antecedent and the consequent. Acknowledging this, I argue, amounts to acknowledging the problem of pragmatic intrusion into semantics. Consider (32), where the conditional is used to make an instruction.

(32) If you press the red button, the conveyor belt will stop.

Let us assume that in uttering (32) the speaker is instructing the hearer (an apprentice) how to stop a conveyor belt. If (32) is to be used as an instruction on how the apprentice is supposed to stop the conveyor belt, the speaker must be communicating that there is a causal-consequential relation between the antecedent and the consequent such that the truth of the consequent follows from the truth of the antecedent. Notice that for (32) to count as an instruction, it cannot possibly be used to communicate anything else in terms of truth values; it cannot communicate that the consequent will be true regardless of whether the antecedent is T or F because even though consistent with MI, it would not be instructive at all.[14] Arguably thus, the thought communicated by the speaker – what is B-said – involves more than what is predicted by MI. Crucially, what is B-said involves the assignment of particular truth values and the assignment of particular truth values depends on the assumption that the stopping of the conveyor belt is a consequence of the pressing of the red button. In other words, what is B-said involves Grice's Indirectness Condition.

A similar problem arises on a Gricean analysis of the iff interpretation of conditionals in terms of biconditional strengthening. Consider (33).

(33) I'll tell you something interesting, if you promise to keep it a secret.

Arguably, the thought explicitly communicated by (33) – i.e. what is B-said – is not that the speaker will tell the hearer something interesting regardless of whether the hearer keeps it a secret or not ($p \supset q$), but if and only if the hearer makes the relevant promise ($p \equiv q$). The problem for Grice is that if pragmatic processes are involved in the derivation of what is B-said (as in (32) and (33)), then, for Grice, pragmatic processes are involved in the derivation of semantic content. This Gricean problem of pragmatic intrusion into semantics constitutes the driving force behind Relevance Theory's re-analysis of the explicit-implicit distinction. In the next section, I give an outline of

[14] I return to this example in chapter 5.

Relevance Theory's re-analysis of this distinction. Its implications for the theory of conditionals are discussed in 4.2.2.3, 4.2.2.4 and 4.2.2.5.

4.2.2.2 *Relevance Theory's explicit-implicit distinction*

Relevance Theory diverges significantly from Grice in claiming that what is linguistically encoded does not equal what is explicit. In RT (e.g. Carston 1988, 2002), explicitly communicated content (i.e. explicature) is derived on the basis of what is linguistically semantically encoded and the context – the derivation of explicatures requires substantive pragmatic work. In particular, in order to derive the explicature, the hearer needs to assign reference, disambiguate ambiguous terms, supply ellipsis and derive ad hoc concepts from the purportedly encoded lexical concepts. In acknowledging the importance of pragmatic processing in the derivation of explicatures, RT rejects Grice's distinction between context-free linguistically encoded semantics and context-sensitive pragmatics as defining of the explicit-implicit distinction.

An explicature, in RT, constitutes what-is-said, but without identifying what-is-said with what is encoded (i.e. RT's explicature equals B-said). Implicature, on the other hand, is an implicitly communicated proposition which is pragmatically derived on the basis of what is explicated and the context. In Sperber and Wilson's (1995: 182) words: 'Any assumption communicated, but not explicitly so, is implicitly communicated: it is an *implicature'*. Whereas the speaker commits herself to what is explicated (Carston 2002: 123-124), implicature is a risky inference in the sense that, being implicitly rather than explicitly communicated, the speaker does not necessarily endorse it. Given that, it is reasonable to expect that, since the speaker's endorsement is present in the case of explicatures, they should not be cancellable[15], and that implicatures, freed from such commitment, should all be cancellable. This, however, is not the case.

In RT, an explicature is said to be a DEVELOPMENT of the encoded logical form (LF) of the sentence uttered[16]. Even though the notion of 'development' lies at the heart of RT's explicature-implicature distinction, it is not a clearly defined term. The picture that emerges from Carston's (1988) discussion is an account of development in logical terms (Burton-Roberts 2005); a communicated proposition is a development/enrichment of LF of an utterance if and only if it entails that LF. Thus, although explicit does not

[15] 'Cancellable' here means 'cancellable without producing a contradiction of what is said.
[16] In RT, the logical form of a sentence corresponds to the sentence's encoded linguistic semantics.

equal encoded in RT, it still depends on what is encoded. As observed by Burton-Roberts (2005: 397), this works in many, but not all, cases.

For example, Grice's GCIs are in general re-analysed in RT's terms as explicatures. Thus, the temporal order GCI involved in the interpretation of (7a) (*Tom got up and read the paper*) is re-analysed as an explicature. This is because the pragmatically enriched '*and then*' entails what is encoded by *and*, i.e. the truth of both conjuncts[17]. Also, on the assumption that the iff interpretation of conditionals involves the pragmatic strengthening of encoded MI to equivalence (as discussed in 4.1.2.4), it has to be re-analysed in RT as an explicature because equivalence entails MI (see e.g. Smith & Smith 1988: 333). Similarly, since the Indirectness Condition is for Grice a GCI, it too needs to be re-analysed in RT as an explicature.

Conversely, Grice's PCIs are (consistently with Grice) treated in RT as implicatures. For example, B's answer in (8a) (*A: Are you hungry?; B: I've had breakfast*) communicates the PCI in (8d) – that *B is not hungry*. This PCI is in RT treated as an implicature. This is because (8d) does not entail what is linguistically encoded, i.e. that the speaker is not hungry does not entail that the speaker has had breakfast at some point prior to the utterance.

RT's reanalysis of Grice's GCI in terms of an explicature is indicative of their rejection of the criterion of cancellation as defining of the explicit-implicit distinction. If Grice's GCIs are explicatures then, since Grice's GCIs are cancellable, explicatures must be cancellable. Indeed, RT (e.g. Carston 2002: 138) allows that explicatures can be cancelled.

In this context, Burton-Roberts (2005) emphasises Carston's (2002: 123-124) characterisation of explicatures in terms of 'expressing… commitment' and 'overtly endorsing' what is explicitly communicated. Burton-Roberts argues that the suggestion that explicatures can be cancelled is inconsistent with this characterisation of explicature since the speaker cannot cancel what she is committed to without contradicting herself.

To complicate things further, the definition of explicature in terms of logical development of the encoded LF, combined with the exhaustivity of the explicature-implicature distinction, means that RT has to allow that some implicatures cannot be cancelled. Consider the following example.

[17] Carston (2002: 256) entertains the idea that perhaps *and* 'has no linguistic meaning at all' (conceptual or procedural) and that the truth values of co-ordinated propositions fall out from syntactic co-ordination. I return to this point in 4.2.2.3.

(34) a) The judge is my father.

b) The judge is a man.

(34b) cannot be regarded as explicated by an utterance of (34a) as it does not entail (34a). On the contrary, (34a) entails (34b). In other words, if one adheres to the logical (entailment) account of development, (34b) is not a development of (34a). In RT's terms thus, there is no other alternative but to treat (34b) as a case of implicature (e.g. Carston 2002: 140). A general problem with this analysis is that (34b) is defined in mutually exclusive terms – it is both entailed by and implicated by (34a). This analysis goes against Grice, who defined implicature in non-truth-conditional terms. A particular problem with respect to the criterion of cancelation is that RT has to allow that there are some implicatures which cannot be cancelled because they are actually entailed by what is linguistically encoded.[18] Consider now (35) below.

(35) a) A: Have you invited any men to the function?

B: I've invited my father.

b) I've invited a man.

Burton-Roberts (2005) argues that (34b) ought to be treated as an explicature since (34b) constitutes part of the truth-conditional content of the explicated proposition. Similarly, (35b) should be treated as an explicature of B's answer in (35a). Burton-Roberts' point is that, if the truth-conditional content of (34a) and (35a-B) is treated as implicated, then all the truth-conditional content of the explicature would have to be regarded as implicated.[19]

Burton-Roberts (2005) also points to the fact that the definition of development in terms of entailment is problematic for the analysis of negation. Compare (36) and (37).

(36) a) I've had breakfast.

b) I've had breakfast today.

(37) a) I've not had breakfast.

[18] As pointed out by Burton-Roberts (2005), Carston (2002: 140-141) argues that there is nothing wrong in the treatment of (34b) as an implicature because in RT concepts are atomic á la Fodor (1998). However, given the arguments presented in chapters 1, 2 and 3, this argument is not valid.

[19] Burton-Roberts acknowledges that (35a-B) is not a STRICT ANSWER to A's question in (35a). He argues that an answer is strict if and only if the communicated truth-conditional content is both sufficient and necessary to answer the question. Even though (35a-B) *communicates* a strict answer (because it is sufficient), it *is not* itself a strict answer (because it is not necessary). Burton-Roberts (*ibid.*) argues that his notion of strictness of an answer captures the indirectness of B's response in (35a) without the problem associated with treating (35b) as merely implicated.

b) I've not had breakfast today.

(36b) above is analysed as an explicature of (36a) because (36b) entails the less specific (36a). With negatives, however, the relation of entailment holds in the opposite direction. Thus, in (37) above, it is (a) that entails (b) because (b) is less specific. This means that, given the logical definition of development, (37b) cannot – parallel to (35b) (*I've invited a man*) – be treated as an explicature of (37a). RT must analyse it as an implicature. This contrast in RT's analysis of positives and negatives is counter-intuitive – there is no reason, other than RT's anyway problematic logical definition of development, for why the relevant temporal restriction should be explicated in (36) but implicated in (37).

Now, Recanati (1991) criticises RT's treatment of some truth-conditions as implicatures. He calls attention to Carston's (1991) Independence Principle, which states that if a communicated proposition entails or is entailed by the LF of the sentence uttered, it must be considered as part of what is explicated. However, given Carston's (2002: 140-141) insistence that (34b) is an implicature (see footnote 18), the Independence Principle does not actually seem to play any role in defining RT's distinction between explicatures and implicatures.

In summary, RT correctly rejects Grice's distinction between encoded semantics and pragmatics as defining of the distinction between explicit and implicit communication. However, acknowledging the substantive role of pragmatics in the derivation of explicatures forces RT to reject (the principle of) cancellation as defining of the explicit-implicit distinction.

As discussed, RT (Sperber & Wilson 1995, Carston 2002) defines explicature as a development/enrichment of the logical form of an utterance. But there are several problems with this definition. If development is defined in terms of entailment (such that *X* is a development of *Y* if and only if *X* entails *Y*), RT's account gives rise to many inconsistencies: a) explicit content can be cancelled even though the speaker endorses/commits herself to it; b) there are cases which are logical developments but which are acknowledged in RT as implicatures; c) negative and positive utterances do not receive a uniform treatment.

Carston (2002: 188-190) tentatively suggests that if the notion of development did not pertain to logical but rather cognitive strength, (34b) – and thus (37b) – could actually be treated as explicated. However, Burton-Roberts (2005: 398) points out that exchanging logical strength (i.e. specificity) for cognitive strength (i.e. specificity) does

not solve the problem since the cognitive is, for RT, the locus of logical (i.e. truth-theoretic) properties. Cognitive strength should thus parallel logical strength.

One solution to RT's problems might be not to define development logically in terms of entailment. However, if development is not defined in logical terms, there is no definition of development. If there is no definition of development, there is no definition of explicature since explicature is defined in terms of development. If there is no definition of explicature, there is no definition of implicature. This is because in RT implicature is defined negatively with respect to explicature.

In this section, I have introduced RT's re-analysis of Grice's semantics-pragmatics distinction and discussed problems with this re-analysis. In the next three sections, I look at its implications for the study of conditionals.

4.2.2.3 *Relevance Theory's circle: the nature of linguistic semantics revisited*

In chapter 2, I discussed RT's distinction between real and linguistic semantics in terms of truth-theoretic value or lack of it, respectively. In contrast to Grice, RT's (standard) claim is that the encoded (linguistic) semantics is non-truth-theoretic – for RT, truth conditionality lies only in thought. Carston (e.g. 2002: 99-100) argues that the problem of 'pragmatic intrusion' does not arise once the explicit-implicit distinction is re-analysed in Relevance Theory's terms. Carston's (*ibid*.) point is that the problem of 'pragmatic intrusion' arises only if what-is-said is identified with linguistic semantics. As discussed in the previous section, RT identifies what-is-said with explicitly communicated content (i.e. explicature), but the explicitly communicated content is not identified with the encoded linguistic semantics (even though dependent on it). In (standard) RT, pragmatics is involved in the derivation of what is explicated, but the encoded linguistic semantics is 'autonomous with respect to pragmatics' (Carston 2002: 99). Insofar as pragmatics is involved in the derivation of semantic content, it is in RT involved in the derivation of real, i.e. truth-conditional, semantic content of communicated thought. In what follows, I argue that Carston's (*ibid*.) claim that encoded semantics is autonomous with respect to pragmatics is invalid. Crucially, if this claim of Carston's cannot be maintained, the very rationale for positing linguistic semantics is lost.

As discussed in chapter 2, RT's account of the acquisition of linguistic semantics is *post hoc* in that it presupposes prior understanding of utterances. If, as RT (e.g. Carston 2002: 364-365) argues, linguistic semantics is abstracted from multiple memory traces of experiences of various uses of a given word, then some utterances

must be understood prior to the acquisition of linguistic semantics. It is clear that if some utterances are understood *prior* to the acquisition of linguistic semantics, they must be understood through and solely through pragmatic inference. In chapter 2, I argued that acknowledging that the understanding of utterances does not require the mediation of linguistic semantics equals acknowledging that linguistic semantics is unnecessary. However, this acknowledgement also gives rise to another problem for RT – the problem of pragmatic intrusion into encoded (i.e. linguistic) semantics. In principle, if linguistic semantics is abstracted from pragmatically inferred interpretations, linguistic semantics cannot be autonomous with respect to pragmatics. Thus, even RT cannot deny that it is pragmatic inference that gets an interpreter to the purported linguistic semantics.

Indeed, as discussed in chapters 2 and 3, there is an instability in RT regarding the nature of linguistic semantics. Carston (2002, 2010) herself acknowledges that at least some lexical concepts are full-fledged, i.e. they are concepts which, before undergoing lexicalisation, had been constructed through (personal) pragmatic inference and which thus have truth-theoretic properties (are true or false with respect to a given state of affairs in the world). This, I argued, amounts to admitting that real semantics is not the only locus of truth-theoretic properties and raises a problem for Carston's (e.g. 2002: 365) claim that linguistic semantics constitutes 'evidence' for the intended interpretation. Here, I argue that allowing that at least some linguistic concepts (i.e. full-fledged concepts) have been constructed through pragmatic inference is another respect in which Carston's (2002: 99) claim that linguistic semantics is 'autonomous with respect to pragmatics' is implausible.

Also relevant here is RT's definition of explicature in terms of logical development. For explicature to entail the LF (i.e. the linguistic semantics of a sentence) of which it is a development, the LF must be a propositional entity, i.e. it must be an entity with truth-theoretic properties. Since 'implicature' is in RT defined solely in contrast to 'explicature' (as discussed in 4.2.2.2), RT's explicit-implicit distinction must rest on the assumption that linguistic semantics has truth-theoretic properties.

There is an interesting paradox involved here. Carston (2002: 99-100) argues that Grice's circle (section 4.2.2.1) can be avoided if Grice's explicit-implicit distinction is re-defined along RT's lines. Crucially, the success of RT's re-definition of the explicit-implicit distinction in solving Grice's circle rests on the argument that linguistic semantics, unlike real semantics, *does not have* truth-theoretic properties. However, RT's definition of the explicit – and hence of the explicit-implicit distinction – rests on

the argument that linguistic semantics *does have* truth-theoretic properties. Thus, RT's re-definition of Grice's explicit-implicit distinction achieves the opposite of what it sought to achieve in the first place. I shall refer to this problem of RT's as 'Relevance Theory's circle'.

As discussed, Relevance Theory's circle is evident when we consider RT's theoretical claims regarding the acquisition of linguistic semantics, the nature of linguistic semantics and their re-definition of the explicit-implicit distinction. It is also evident when we consider more closely Noh's (2000) claim that *if* linguistically semantically encodes MI.

Before I discuss Noh's proposal in more detail, an important comment is in order. As argued earlier, Relevance Theory's circle points to the conclusion that there is no principled distinction in RT between linguistic semantics and real semantics – especially in the light of my argument (2.1.3) against the necessity of positing the process of decoding. In this connection, consider the following quote from Carston (2002: 257):

> **Quote 1**: The truth relation holds between thoughts and states of affairs, so between propositions expressed by utterances (semantic/pragmatic hybrids) and states of affairs. Then, it is system of thought, rather than linguistic systems, for which a truth calculus, that is, a logic, should be devised.

I agree entirely with what Carston is suggesting here – thought is the only locus of truth theoretic properties. However, the problem of Relevance Theory's circle, which I identified for various aspects of Relevance Theory (acquisition of linguistic semantics, the nature of linguistic semantics and RT's explicit-implicit distinction), shows that RT, including Carston, has not fully embraced the radical consequences of this claim: that there is no linguistic semantics or the process of deterministic decoding, for that matter; that a wholly pragmatic inference gets an interpreter to a (holistic) thought, which is the only locus of semantics.

Indeed, consider Carston's (2002: 256) suggestion (mentioned in footnote 17) that the word *and* encodes no linguistic meaning (conceptual or procedural) and that the truth values of co-ordinated propositions follow from syntactic co-ordination. What Carston does here is distinguish between the 'semantics of *and*', which is empty, and 'the logic of *and*', which is a syntactic matter. Clearly, if the logic of *and* is a syntactic matter, then the logic of *and* (i.e. conjunction) is still a linguistic, though not semantic, property. However, this claim of Carston's is inconsistent with what she is suggesting in quote 1 – that is, that truth is not a property of linguistic representation but of thought. It

is not clear whether for Carston truth-theoretic properties apply in language or in thought. This inconsistency clearly reflects RT's instability about the nature of semantics. Therefore, as much as I agree with what Carston is saying in quote 1, I argue that this claim can only be maintained once its radical consequences are acknowledged.

4.2.2.4 *Relevance Theory's circle: pragmatic intrusion into encoded semantics of 'if'*

In this section, I show that the problem of pragmatic intrusion into encoded semantics arises for Relevance Theoretic accounts of conditionals (Smith & Smith 1988, Noh 2000). My arguments will strengthen Carston's claim made in quote 1 and call for the resolution of the Relevance Theory's circle by means of acknowledging the radical pragmatic consequences of this claim.

Consider again the biconditional interpretation of *if*, which, it has been argued (e.g. Smith & Smith 1988), is easily amenable to a linguistic semantic analysis in terms of MI. Earlier, I discussed examples (14) and (15), repeated below.

(14) If you pass the exam, I'll take you to a restaurant.
(15) If you do something illegal again, we'll lock you up.

According to MI, (14) and (15) should be false only if the antecedent is T and the consequent F, and true otherwise. As discussed in 4.1.2.4, it might be argued that in uttering (14) and (15) the speakers are not actually making any undertaking as to what they will do if the antecedent is false, only what they will do if the antecedent is T – i.e. that MI correctly predicts that the speakers of (14) and (15) have spoken falsely only in a situation where the antecedent is T and the consequent F. Thus, MI allows that the hearer of (14) may be taken to a restaurant even if he fails the exam, and that the hearer of (15) may be locked up even if he refrains from doing something illegal again. I will argue that there is in fact a problem with this analysis.

Examples (14) and (15) exemplify two types of conditional inducements, i.e. statements made to influence hearers' behaviour by telling them about the consequences of their behaviour (Beller 2002). Accordingly, (14) is used to influence the hearer's behaviour by *promising* that *q* will be T if *p* is T, and (14) is used to influence the hearer's behaviour by *threatening* that *q* will be T if *p* is T. Promises and threats differ. With promises, *p* is a desired behaviour for which there is a reward (*q*). With threats, by contrast, *p* is an undesired behaviour for which there is a punishment (*q*). With promises

the hearer is motivated to make *p* true, whereas with threats the hearer is motivated to make *p* false.

Beller (2002) argues that crucial to the interpretation of conditional inducements is the question of what the speaker is *obliged* to do and what the speaker is *permitted* to do when *p* is false. Importantly, Beller (2002: 114) argues that with promises if *p* is false, there is no obligation on the speaker to reward the hearer with *q*, but she is permitted to do so – i.e. the speaker is permitted to make *q* true even if *p* is false. However, presumably because of the difference in the social perception of reward and punishment, with threats, the speaker is not permitted to make *q* true if *p* is false. This suggests that the biconditional interpretation (i.e. $\sim p \supset \sim q$) is more strongly associated with threats than it is with promises.[20]

Now, if there really is such a difference in speaker commitment in the case of conditional inducements, not keeping a promise should correspond to '*p* and not-*q*' (consistent with MI), whereas not keeping a threat should correspond to 'not-*p* and *q*' (not consistent with MI). Indeed, Beller's proposal seems to be right.[21] In (14), the promise made by the speaker clearly is broken if the hearer passes the exam (*p* is T) and the speaker does not take him to a restaurant (*q* is F). However, if the hearer is taken to the restaurant by the speaker (*q* is T) even if the hearer has failed the exam (*p* is F), one can argue that the speaker is over-indulgent, but not that she has not kept the promise to '*q* if *p*'. In (15), however, there is a strong intuition that the speaker has violated the terms of the threat if the hearer does not do anything illegal again (*p* is F) and yet the speaker locks the hearer up (*q* is T).

In sum, the difference between threats and promises is this: whereas with promises the speaker *is not taken to have spoken falsely* in a situation where *p* is F and *q* is T, with threats the speaker *is taken to have spoken falsely* in a situation where *p* is F and *q* is T. Crucially, this means that MI predicts when a conditional promise is violated (false) but not when a conditional threat is violated (false).

The difference between threats and promises raises serious problems for the assumption that *if* semantically encodes MI. If one wants to maintain MI as the encoded semantics of *if*, one has to acknowledge an inelegant asymmetry; the falsity of a conditional is sometimes (e.g. with promises) semantically encoded and thus deterministically decoded in utterance interpretation, and sometimes (e.g. with threats)

[20] I say more about the interpretation of threats and promises in chapter 5.
[21] In fact, this prediction is supported with Beller's (2002) experimental study. Beller (2002: 117) reports that 100% of the tested subjects equate not keeping a promise with '*p* and not-*q*', and 95% of subjects equate not keeping a threat with 'not-*p* and *q*'.

it is pragmatically inferred. But even this inelegant solution would be difficult to maintain, for the simple reason that what counts as a threat or a promise is a matter of individualistic, holistic pragmatic inference. This is well illustrated by (38).

(38) If you do it again, I'll take you to the funfair.

Example (38) can be interpreted as a promise or a threat depending on whether the hearer finds funfairs a pleasure or not, respectively. Thus, it needs to be established whether the intended interpretation is a threat or a promise in order to establish whether the falsity of the conditional is modelled by MI – and thus encoded – or not. The problem for truth-functional approaches is that the distinction between threats and promises can only be made at the level of a holistic belief system – whether something is a threat or a promise is not a matter of logic.

The point I am making here is that if we agree that (a) MI predicts when a promise is F, but not when a threat is F and that (b) the distinction between promises and threats is a pragmatic distinction, then (c) MI models and thus applies at the level of pragmatically (i.e. holistically) inferable thought, and not at the level of linguistic semantics (LF).

In fact, this problem arises not only for conditional inducements, but also for conditional injunctions. Consider (20) and (21) again.

(20) If there is no other choice, amputate his leg.
(21) If you want to make your parents happy, quit smoking.

The hearer of (20) will be taken not to have complied with the injunction – i.e. will be taken to have made (20) false – if he makes *q* T (the hearer amputates the patient's leg) in a situation where *p* is F (where there is some other choice of treatment). This is incompatible with MI, which would be true in that situation. The hearer of (21), however, will not be taken to have failed to comply with the injunction – i.e. will not be taken to have made (21) false – if he quits smoking (*q* is T) even if he does not want to make his parents happy (*p* is F). This is predicted by MI.

The difference between (20) and (21) may be due to the difference in the perception of leg-amputation and quitting smoking. Amputation is highly undesirable and requires a highly restricted set of circumstances/conditions in order to be performed. For this reason, the injunction in (20) would be normally taken to communicate a biconditional command ($\sim p \supset \sim q$). However, quitting smoking is usually

perceived as a desirable behaviour and thus its performance does not require a highly restricted set of circumstances/conditions in order to be performed. In fact, it is plausible to imagine that the context of (21) is such that the speaker and hearer are talking about how to make the hearer's parents happy. In this context, the antecedent (*If you want to make your parents happy*) is not a condition on the consequent (*quit smoking*) at all. Rather, the consequent suggests *means* to achieve the *goal* represented in the antecedent (and embedded in the propositional attitude represented by *want*). Unlike (20), (21) is not a command, but a suggestion. As with threats and promises, here the problem for truth-functional approaches is that the determination of whether MI applies or not depends on a distinction which can only be made at the level of a holistic belief system – whether something is a desirable action or not, a command or a suggestion, is not a matter of logic.[22]

Consider also the universal conditionals (17) and (18).

(17) Water begins to boil if it reaches 100°C.
(18) If one is a man, one cannot get pregnant.

MI guarantees that the conditional in (17) is false when *p* is T and *q* is F. This is a correct prediction. But MI also guarantees that (17) should be true if *p* is F and *q* T. This is an incorrect prediction – it is impossible for water not to reach 100°C and boil. In contrast, the conditional in (18) is F when *p* is T and *q* F, and T otherwise. The truth value of (18), but not of (17), can be predicted by MI. Again, the difference between (17) and (18) and thus the determination of whether MI applies or not is established at the holistic level of belief – it depends on a cogniser's (so-called 'encyclopaedic') knowledge that there is no other boiling point for water than 100°C and that there are other reasons for not being able to get pregnant than being a man.

It is clear that the cases which I discussed above and which are not modelled by MI are cases of biconditional interpretation. Thus, one may object that the requirement for the pragmatic inference of the truth value of such conditionals follows from the assumption that what we are dealing with is a pragmatic process of biconditional strengthening of the encoded MI to equivalence. But this argument is invalid. To repeat, just as it is pragmatically established that MI does not apply in (15), (20), (17) or on the threat interpretation of (38), so it is pragmatically established that MI applies in (14), (21), (18) and on the promise interpretation of (38). This is because

[22] In chapter 5, I will actually question the assumption that the interpretation of (21) can be modelled by MI.

recognising the lack of need for biconditional strengthening is itself a pragmatic inference.

I made a similar argument in chapter 2, where I argued that the process of deterministic decoding of linguistic semantics is redundant. One reason for arguing so was that in cases where the communicated concept is the same as the purportedly encoded concept – as in Carston's (2010: 242-243) '*Children in most cultures dance spontaneously*' – one and the same general concept DANCE has got to be accessed by two distinct cognitive processes (it has to be pragmatically inferred by the pragmatics module after it has been deterministically decoded by the language module). This is because perceiving that no further concept adjustment (i.e. the pragmatic adjustment of the purportedly encoded DANCE to a more specific ad hoc concept) is itself a pragmatic inference from context. I argued that because such a concept has got to be (and can be) pragmatically inferred anyway, and because there are so many problems with the notion of linguistic semantics, the notion of linguistic semantics and the process of deterministic decoding of such content should be rejected.

The point I am making with respect to MI goes even further. Not only is it the case that (a) the application of MI has to be pragmatically determined but, because of (a), it is also the case that (b) if MI applies, it does so at the level of (holistic) thought – not at the level of the LF encoded by a generated linguistic expression.

It also needs to be pointed out that the assumption that MI constitutes the encoded semantics of *if* gives rise to unsolvable paradoxes. Consider again (27).

(27) If the patient is still alive in the morning, change the dressing.

As discussed in 4.2.1, Edgington's (2008) point with respect to this example was that since ($p \supset q$) is logically equivalent to ($\sim p$ v q), the nurse may argue to have obeyed the command in (27) by making sure that the patient is not alive in the morning ($\sim p$ is T) and not changing the dressing (q is F).

The problem is that MI predicts that the nurse's way of obeying the command is logically correct, but it is nevertheless an unreasonable way of obeying that command. But notice that the clash between what is logically correct and what is reasonable only arises on the assumption that (27) can be modelled by MI! Take away this assumption, and the clash/paradox – and the problem – disappears.

As noted, Edgington suggests that pragmatic explanations of examples like (27) are 'stretching pragmatics rather far'. The argument I have put forward in this

section – that if MI applies, it never does so prior to pragmatic inference – substantiates Edgington's claim. If, as shown in this section, pragmatic inference determines whether MI applies or not, MI cannot be the encoded semantic content of *if*. This is because encoded semantic content is supposed to be autonomous with respect to pragmatic inference.

In the next section, I give further arguments for why appeal to pragmatics in aid of maintaining MI as the encoded semantics of *if* actually shows that it is impossible to maintain MI as the encoded semantics of *if*.

4.2.2.5 *Problems with Noh's (2000) metarepresentational account*

In this section, I concentrate on Noh's (1996, 2000) claim that it is possible to retain MI as the encoded semantics of *if* even on metarepresentational uses. I argue that Noh's appeal to pragmatics in order to maintain MI as the encoded semantics of *if* stretches pragmatics too far to be consistent with the claim that *if* semantically encodes MI.

As discussed in 4.1.2.5, Noh (2000) argues that a truth-functional analysis of examples like (23) (*A: Two and eleven makes thirty; B: If two and eleven makes thirty, you need more work on maths*) is available at the higher-explicature level. Noh argues that a truth-functional relation holds between the meta-level explicature '*you say/believe that two and eleven makes thirty*' and '*you need more work on maths*'. Consider also (39).

(39) If you're thirsty, there's beer in the fridge.

Example (39) defies analysis in terms of MI at the descriptive level in the following way. The truth table for MI predicts that the conditional is T in a situation where *p* is F and *q* is F. But this is a wrong prediction for (39). This is because the truth of *q* ('*there's beer in the fridge*') is independent of the truth value of *p* ('*you're thirsty*'). Relatedly, (39) defies an MI analysis in that that in uttering (39), the speaker is asserting that *q* is T. This means that *q* is not a consequent, hence *p* is not an antecedent and hence (39) is not modelled by MI.

Noh (2000: 201) argues that when approached from a metarepresentational perspective, (39), and examples similar to it, do not defy an analysis in terms of MI. However, even though Noh criticises a descriptive level analysis of (39), she does not

actually offer a metarepresentational analysis of it. Instead, she gives an analysis of (40).[23]

(40) [Son to mother who is going out]
Mum, don't worry. If I'm hungry, there's a sandwich in the fridge.

Noh (2000: 201) argues that the consequent in (40) is an interpretive use – it metarepresents 'a desired thought', i.e. a thought which would become relevant to the speaker in a situation where the antecedent is T. Thus, for Noh, MI models the relation between '*I am hungry*' communicated by the antecedent and '*it will be relevant for me to remember that there's a sandwich in the fridge*' communicated by the consequent. But Noh's analysis is problematic for three reasons:

- it suggests (contrary to Noh's own arguments) that MI applies at the pragmatic rather than semantic level
- it undermines the theoretical significance of the notion of metarepresentation
- it does not offer an adequate explanation

Fundamentally, the meta-level where the application of appropriate (i.e. consistent with MI) truth values is supposed to take place for metarepresentational uses is a pragmatic, not semantic, level. If MI applies here, then it does not apply at the level of decoded linguistic semantics, but at the level of pragmatically inferred thought. This gives rise to the question of whether a metarepresentational analysis actually shows, as Noh (2000: 205) argues, that 'natural-language *if* is **semantically** equivalent to material implication' [my emphasis]. Consider also the following quote from Noh.

> **Quote 2:** […] metarepresentational antecedents can express propositions different from those expressed by descriptively used (i.e. ordinary) antecedents. This is possible as long as **pragmatic enrichment processes can flesh out the linguistically encoded material in such a way as to distinguish between descriptive and metarepresentational use**. As a result, the truth table for material implication applies to these 'non-basic' conditionals just as it does to more basic ones, and a standard argument against the truth table account dissolves. (Noh 2000: 190-191, my emphasis)

In this quote, Noh makes it clear that descriptive and metarepresentational uses are pragmatically distinguished. But if that is the case, then the application of appropriate truth values – not only to meta-representational but also to descriptive uses – must occur

[23] I shortly return to a metarepresentational re-analysis of (39).

after it has been pragmatically determined whether the use in question is descriptive or meta-representational. Significantly, this suggests that, contrary to what Noh argues, Noh's arguments are not about the encoded semantics of *if*, but about its pragmatics. Noh, however, does not acknowledge this point.[24]

As for the second problem, let me illustrate it with a metarepresentational re-analysis of (39) (*If you're thirsty, there's beer in the fridge*). This example is significant because Noh (2000: 191) explicitly states that a metarepresentational re-analysis of examples like (39) is crucial to maintaining the argument that MI is the encoded semantics of *if*. However, even if we assume, along with Noh, that a metarepresentational analysis can be applied to (40), its application to (39) is problematic.

Let us imagine a context in which (39) is uttered by Paul to his brother John who has just entered the house after laying bricks in direct sun for the past eight hours, and that John has not verbally or in any other way communicated to Paul that he is thirsty. In such a context, there is a choice of two metarepresentational analyses: (i) the consequent is an interpretive use or (ii) the consequent is a metalinguistic use. Both options are problematic.

On the first option (i), the consequent in (39), like the consequent in (40), metarepresents a desired thought which will become relevant to John if the antecedent is T. The problem with an interpretive analysis of this consequent is that if '*there's beer in the fridge*' is a metarepresentation of a thought, it is a metarepresentation of the speaker's own thought. If the consequent in (39) is interpretive, it is used to communicate that it will be relevant for the hearer to entertain the thought that the speaker himself entertains as T with respect to the world. In fact, the assumption of relevance of *q* to *p* depends on the speaker's entertaining this particular thought as T with respect to the world. The problem is that if the consequent in (39) is metarepresentational of the speaker's own thought – i.e. if one can metarepresent one's own thoughts – then every single utterance, including the antecedent in (39), is metarepresentational and the notion of metarepersentation loses its import.[25]

[24] The descriptive-metalinguistic distinction was originally put forward by Horn (1989) as a pragmatic distinction. Horn argues that when *if* is used descriptively, it equals MI, but when it is used metalinguistically, it does not. Interestingly, Noh (1996: 26) argues that Horn (1989) posits a pragmatic ambiguity which 'seem[s] to imply a semantic ambiguity'. If Noh is right, both Noh (for the reasons discussed above) and Horn fail to see that when MI applies, it does so in pragmatically inferred thought.
[25] In fact, the same problem arises for the consequent in (40).

On the second option (ii), the consequent in (39) would be a metalinguistic use; it would communicate something like '*I am saying that …*'.[26] But the problem here is that if the consequent in (39) could be analysed as a metalinguistic use, then every utterance that every single speaker makes could be so analysed and, again, the notion of metalinguistic use would lose its theoretical significance.[27]

The third problem with Noh's analysis concerns the notion of metarepresentational faithfulness. As discussed earlier, Noh's metarepresentational re-analysis of examples which defy an MI analysis at the descriptive level *relies on* the relevance of the notion of metarepresentational faithfulness (of metarepresentation to the metarepresented object). As discussed in 4.1.2.5, Noh (2000: 187) argues that 'what is important in metarepresentation is faithfulness to the original, rather than truthfulness'. Consider (26) again.

(26) [A is talking to B about her meeting with her supervisor]
A: Then what did she say?
B: The argument is invalid.

Noh argues that B's utterance in (26) is T if and only if it is a faithful metarepresentation of B's supervisor's utterance. Noh (2000: 189) writes: 'what is important is not whether the argument is valid or not, but whether the supervisor said something resembling that or not'. Notice that if it is irrelevant whether the argument is valid or not, it is irrelevant whether what B's supervisor communicated was T or F. For Noh, (26B) is T as long as it is faithful to what the supervisor said, but regardless of whether what the supervisor communicated was T or F, i.e. regardless of whether the supervisor held a T or F belief about the argument.

For Noh (e.g. 2000: 202), the truth value of what is being metarepresented is irrelevant also in the case of conditionals. Consider (23), repeated below.

(23) A: Two and eleven makes thirty.
B: If two and eleven makes thirty, you need more work on maths.

As noted earlier, Noh (2000: 187-189) maintains that even though a descriptive level MI analysis of (23B) yields wrong results, (23B) is modelled by MI at the meta-level. At the meta-level the antecedent ('*you say/believe that two and eleven makes thirty*')

[26] In fact, Noh (2000: 207) offers a metalinguistic analysis of the speaker's own utterance for conditionals with patently false consequents. I criticise this proposal of Noh's later in this section.
[27] I will discuss a simpler and less problematic solution to the interpretation of examples like (39) in chapter 5.

comes out as T on the assumption that it faithfully metarepresents A's utterance. Given that the antecedent is T, the consequent has got to be T for the whole conditional to be T. The metarepresentational analysis of the antecedent, thus, guarantees the assignment of T and only T to the consequent. Thus, on Noh's analysis the consequent in (23B) is T because the antecedent is a faithful metarepresentation of A's utterance and regardless of the fact that the antecedent is F at the descriptive level. It is precisely the importance of faithfulness rather than truthfulness which allows Noh to maintain MI as the encoded semantics of all conditionals.

However, Noh's metarepresentational account misses a very important point about (23B) – that the faithfulness of the metarepresentational antecedent simply *is not* the reason why B in (23) communicates that A needs more work on maths (that the consequent is T). In fact, contrary to what follows from Noh's proposal, crucial to the truth of the consequent is not *the fact that B has truly metarepresented A's belief,* but *the fact that A falsely believes that a patent falsity is true*. It is thus the falsity of A's belief which allows B to communicate that the consequent in (23) is T.

The following modification of (23) emphasises my point. If Noh were right in claiming that the 'truthfulness' of what is being metarepresented is irrelevant in metarepresentational uses, B would be able to communicate '*you need more work on maths*' by faithfully metarepresenting a true mathematical belief of A – e.g. a belief that '*two plus two makes four'*. However, because A's belief that '*two plus two makes four*' is T, B cannot communicate (or conclude) that A needs more work on maths even if B has faithfully metarepresented this belief of A's.

Now, MI does not predict the fact that the truth of the consequent in (23B) follows from the falsity of A's belief. Nevertheless, the use of *if* in (23B) does communicate that there is some (consequential) relation between *p* and *q*. This needs to be explained. Crucial to the explanation is, I argue, the falsity of A's mathematical belief. A believes that what B knows is a patent mathematical falsity is true. Learning about this false belief of A, B realises that A does not have necessary mathematical knowledge to be aware that this belief is actually false. As a solution to this problem, B suggests that A needs more work on maths. There clearly is a relation of relevance between *p* and *q* in (23B), but that relation is not modelled by MI.[28]

Also important and relevant here is Noh's analysis of (41).

[28] I return to this example in chapter 5.

(41) *Pope to a telephone operator in a small Swiss village*: I'm the Pope.
Operator: If you're the Pope, I'm the Empress of China.

Noh (2000: 207) argues that the antecedent in the operator's utterance in (41) is used to metarepresent the hearer's (the Pope's) utterance and the consequent is used to metarepresent 'an utterance that the speaker wants to make in the case where the proposition expressed by the antecedent is uttered/entertained'.[29] According to Noh (*ibid*.), the speaker 'is asserting that if the hearer is saying that he is the Pope, she will say that she is the Empress of China'. At the meta-level, T will be assigned to '*you say you're the Pope*' and to '*I say I'm the Empress of China*'.

Whereas Noh's meta-level analysis is indeed consistent with MI, it misses the crucial point that the operator in (41) is communicating her disbelief that the hearer is the Pope by inviting an inference from the patent falsity of the consequent to the falsity of the antecedent (Modus Tollens). This inference is licensed by MI, but it is licensed at the descriptive, not meta-representational level. Noh (*ibid*.) does mention that both utterances involved in the meta-level analysis are 'presented as blatantly false' and that they 'function as abstract echoes' of 'the property of being patently false'. However, the problem is that it is not clear how the meta-level can 'echo' the falsity involved at the descriptive level if, as argued by Noh (2000: 202), in metarepresentational uses 'the proposition literally expressed by the consequent is given guarantee not of truthfulness, but of faithfulness'.

I have been arguing that Noh's metarepresentational analysis of (41) is inadequate. This example can be more simply explained as inviting MT (an inference licensed by MI) in virtue of the patent falsity of the consequent. However, even though the interpretation of (41) is modelled by MI, MI applies at the level of holistic thought. The operator in (41) communicates her disbelief that the hearer is the Pope by inviting him to perform MT. The operator can assume that the hearer (i.e. the Pope) will be able to perform MT only because, given that the hearer (i.e. the Pope) assumes that at the other end of the line is an operator in Switzerland, the hearer (i.e. the Pope) will be able to pragmatically infer the blatant falsity of the consequent. Crucially, the hearer (i.e. the Pope) will be able to infer from the falsity of 'I'm the Empress of China' to the communicated falsity of 'you're the Pope' even though the hearer (i.e. the Pope) knows that 'you're the Pope' is in fact true. The Pope will be able to infer that MI models the

[29] See my earlier point that if consequents like the one in (41) can be analysed in metalinguistic terms, then all utterances can be so analysed.

relation between 'I'm the Empress of China' and 'you're the Pope' in the operator's mind even though it does not model the relation between these two propositions/thoughts in his mind. This shows that if MI models the relation between 'I'm the Empress of China' and 'you're the Pope', it does so not as a matter of logical necessity, but as a matter of an individual's holistic belief system.

Now, Noh's metarepresentational analysis seems to make correct predictions in some cases. For example, it seems plausible to agree with Noh that MI predicts the injunction interpretation of (22) (*If that's John, I'm not here*), which involves a metarepresentational consequent. However, for the reasons discussed in 4.2.2.3, 4.2.2.4 and in this section, one cannot extrapolate from the existence of such examples to the claim that MI is the encoded semantics of *if*.

The aim of this chapter was to answer the (title) question of whether *if* encodes MI. The answer is that it does not, because even where MI applies, it does so in pragmatically (i.e. holistically) inferred thought. It is thus not language, but thought which is the locus of MI.

4.3 Conclusion

In this chapter, I have argued that there are two reasons why it is impossible to maintain the claim that *if* semantically encodes MI. Firstly, there are conditional belief deviations from MI. Because such deviations take place at the level of belief, they cannot be explained in terms of conversational principles. Secondly, the problem of pragmatic intrusion into encoded semantics arises for Relevance Theory just as it did for Grice.

The problem of pragmatic intrusion has important consequences. If pragmatic inference determines whether MI applies or not, MI cannot be the encoded semantic content of *if*. This is because encoded semantic content is supposed to be autonomous with respect to pragmatic inference (e.g. Carston 2002: 99-100). If, as I have argued, the purportedly encoded semantic content is not autonomous with respect to pragmatics, the rationale for positing encoded semantic content is lost.

The conclusion that the rationale for positing encoded semantic content is lost is further strengthened by RT's account of the acquisition of linguistic semantics, the nature of linguistic semantics and their re-definition of the explicit-implicit distinction.

My argument that if MI applies, it does so at the level of holistic content suggests that conditional beliefs – and thus the interpretation of conditionals – can only be explained with reference to, because they arise at, the holistic/individualistic cognitive level. This argument also strengthens the claim of the Representational

Hypothesis (e.g. Burton-Roberts 2012) that thought is the only locus of semantic content, my claim (developed in chapters 1-3) that such content is holistic and thus accessible solely through pragmatic inference.

One final comment that needs to be made at this point concerns the relation between MI and LOT. In chapter 3, I argued that an RH-based wholly inferential account which I propose allows for successful communication by combining the identity of LOT – i.e. primitive concepts and compositional capacity – among the human species with Bilgrami-type public availability of conceptual structures (which are allowed by the shared LOT). Now, one of the main arguments of this chapter was that holistic thought, and not linguistic expressions generated by the language module, is the locus of MI. By this I meant that the application of MI has to be pragmatically determined. However, the pragmatically determined (i.e. holistic and individualistic) *application* of MI does not at all indicate that MI itself is not subject to a realist-naturalistic inquiry. Indeed, assuming that:

(a) MI is part of the logical and thus realist-naturalistic system
(b) The language module (on the assumption that it exists), or any expression generated by it, is not the locus of MI, as argued in this chapter
(c) LOT is subject to a realist-naturalistic inquiry

It follows that (d) MI is licensed/allowed by (most plausibly the compositional component of) LOT. Thus, even though the question of when MI applies is pragmatically determined, the fact that it does (and can) apply is naturalistically determined.

In the next chapter, I show how the interpretation of conditionals can be explained in terms of a wholly pragmatic inferential process cashed out in terms of Hintzman's (1986) multiple-trace theory of memory (introduced in chapter 2) and its extension to the study of conversational common ground by Horton & Gerrig (2005). I also discuss the implications that a wholly pragmatic account of utterance interpretation has for the distinction between explicit and implicit communication.

Chapter 5. Holistic and individualistic conditions on interpretation

5.0 Introduction

The aim of this chapter is to show how the interpretation of conditionals can be handled by a wholly pragmatic inferential process of utterance interpretation. I first introduce Horton & Gerrig's (2005) extension of a multiple-trace theory of memory (e.g. Hintzman 1986) into the study of common ground. I argue that once Horton & Gerrig's argument – that interlocutor-specific information places immediate constraints on utterance interpretation – is acknowledged, it is possible to explain the distinction between the weak and the strong interpretations of conditionals.

I discuss basic uses of conditionals (where *if* signals a relation between two propositional objects *p* and *q*) and look at various relations which hold between *p* and *q* in such uses. Then, I move on to extended uses of conditionals (where *if* signals a relation between a proposition and an utterance (or speech act)) and discuss why the analysis I defend is more adequate than Sweetser's speech-act analysis. Finally, I discuss implications of a wholly pragmatic approach to utterance interpretation for the explicit-implicit distinction.

5.1 Interpretation of conditionals and common ground

In 4.2.2.4, I discussed, among other things, the interpretation of conditional inducements. In line with Beller (2002), I argued that there is an important distinction between the interpretation of conditional promises and threats: whereas a conditional promise is generally taken to have been broken (i.e. the speaker is taken to have spoken falsely) when *p* is T and *q* is F, a conditional threat is generally taken to have been violated when *p* is F and *q* is T. This shows that MI predicts when a promise is false, but not when a threat is false. It also suggests that the biconditional interpretation is more strongly associated with threats than it is with promises.

I further argued that this difference in the interpretation of conditional threats and promises raises problems for the claim that *if* semantically encodes MI. Because the distinction between threats and promises is made at the level of a holistic belief system, when MI applies, it does so at the level of a pragmatically (i.e. holistically) inferred thought. I argued that the interpretation of conditionals – like all utterance interpretation – is a matter of a wholly pragmatic (holistic) inferential process.

Now, if all utterance interpretation is a wholly pragmatic process, it necessarily depends on and is thus constrained by (cognitive) contextual factors. For example, in 4.2.2.4, I discussed how the interpretation of (38), repeated here as (1), depends on the hearer's desires and dreads.

(1) If you do it again, I'll take you to the funfair.

Depending on whether the hearer likes or dislikes the funfair, (1) is intended as a promise or a threat. In turn, depending on whether (1) is intended as a promise or a threat, its interpretation is or is not modelled by MI.

One particular point I want to develop in the first section of this chapter is that part of contextual assumptions involved in the interpretation of conditionals is the interlocutors' knowledge of one another's holistic state of mind (i.e. cognitive context). My argument is compatible with, and strengthened by, recent (Horton & Gerrig 2005; Horton 2008) psycholinguistic work on the notion of common ground.

5.1.1 *Horton and Gerrig's notion of common ground*

Horton & Gerrig (2005) are interested in the notion of COMMON GROUND, which concerns the set of assumptions (knowledge, beliefs, etc.) which interlocutors in a conversation consider as shared for the purposes of communication (Clark 1994). Consider the following exchange (taken from Horton & Gerrig (2005)).

(2) A: Oh first of all I have Shana's shower coming up that I have to do.
B: Ah, that's right.
A: That's going to be like a huge like three day effort with all the cooking and cleaning and like actually party [sic] that I have to do.
B: Is there anyone you can get to help you?
A: Um Jessica's going to help and Beth might because you see, Diane is here now.
B: Oh okay.

Horton & Gerrig (2005) are concerned with what lies behind the successful use of referring expressions exemplified by *Shana's shower*, *Beth*, *Jessica* and *Diane* in (2). They are interested in what it is that makes B successfully interpret the use of these expressions; how it is that A knows what expressions to use to ascertain a successful interpretation of them by B; and how it is that A and B know that what A intended to

communicate by the use of a given expression is what B understood as intended by A. In other words, Horton & Gerrig are interested in psycholinguistic underpinnings of the notion of common ground.

As observed by Clark (1994: 989), successful communication, like any joint activity, depends on participants' co-ordination of their actions through making assumptions about each other. Clark (*ibid*.) distinguishes between COMMUNAL COMMON GROUND, which concerns the assumptions taken to be shared by members of the same general community, and PERSONAL COMMON GROUND, which refers to more specific mutual knowledge, etc. that interlocutors have inferred about each other from past experiences with each other. As Clark (1994: 990) puts it, common ground is 'the background, the context, for everything the participants jointly do and say in [discourse]'. The goal of Horton & Gerrig (2005) is to propose a set of mechanisms that give rise to the effect described as common ground.

Horton & Gerrig (2005: 2) argue that common ground is not a category of specialised mental representations, but 'an emergent property of ordinary memory processes acting on ordinary memory representations'. Horton & Gerrig (2005) assume that speakers produce utterances which are suited to particular addressees. Similarly, listeners assume that speakers' utterances were designed with the listeners' needs in mind (Horton 2008: 193). Thus, when B in (2) indicates an understanding of who Shana is, it means, for Horton & Gerrig (2005: 3), that A has 'correctly formulated her utterance against the belief that Shana is part of their common ground'. A general idea illustrated by this example is that speakers adjust their utterances to particular audiences by incorporating their assumptions about their interlocutors' knowledge, etc. – a phenomenon referred to as AUDIENCE DESIGN.

According to Horton & Gerrig (2005), two processes are relevant in audience design: COMMONALITY ASSESSMENT and MESSAGE FORMATION. Commonality assessment concerns cue-dependent retrieval of episodic memory traces[1] which store information associated with a particular interlocutor. The retrieval of interlocutor-specific information allows interlocutors to assume communal or personal common ground. Message formation concerns the use of such retrieved information in constructing an utterance. Thus, for Horton & Gerrig (2005: 4), utterances reflect the speaker's beliefs about common ground.

[1] Horton & Gerrig (2005: 11) argue that commonality assessment arises via the mechanism of resonance (echo retrieval) proposed by e.g. Hintzman (1986). Resonance provides a parallel search of memory and thus allows a wide range of associated information to be accessible.

Let us look at (2) again. The fact that A chose the proper name Shana (rather than a more descriptive phrase like *my sister Shana*, etc.) reflects A's belief that the concept of Shana is co-present for A and B (i.e. that it is in the common ground). This assumption of common ground (and the subsequent use of the proper name) relies on a strong pattern of associations in A's memory between (the concepts of) A, B and Shana. But how does the assumption of common ground arise? In other words, how is a particular set of memory traces (associations between A, B and Shana) activated? Horton & Gerrig (2005: 9-10) argue that conversational interlocutors serve as highly salient cues for the retrieval of traces which are involved in the assumption of common ground. Such cue-based retrieval of interlocutor-associated information accounts for the readiness in memory of records of interlocutor-related experiences and their availability to processes of speech production and comprehension.

Horton & Gerrig (2005: 10) argue that from a pool of stored information only those traces will be retrieved which are most consistently associated with the cue (here, with the interlocutor in a given context). Furthermore, only those traces will become relevant to the processes of language production and comprehension which are retrieved within an adequate time frame. Given such constraints, interlocutor-specific information which is too weak or too slow will not have an immediate impact on the language production and comprehension processes.

Horton & Gerrig (2005: 14-15) point out that the final products of commonality assessment do not always occur prior to message formation. The processes of commonality assessment and message formation interact so that partial commonality assessment may influence message formation and message formation may influence the (re)assessment of commonality. The interaction is dynamic because it is subject to the demands of fluent conversation. Relatedly, commonality assessment is not an infallible process. Hence, commonality mis-assessment and subsequent hearer-inappropriate message formation may take place, which results in a communicative failure.[2]

Horton (2008: 196) suggests that a process analogous to message formation in language production is the process of MESSAGE INTERPRETATION in language comprehension. Horton does not explain how this process works, but it seems plausible to assume that the process of message interpretation is the reverse of the process of message formation: the form of the utterance will influence commonality assessment

[2] Horton & Gerrig (2005: 28) argue that the process of message formation is not exclusively constrained by the commonality assessment but also by factors like lexical frequency, recency, etc.

and thus constrain the search for the intended interpretation. For example, let us assume that B in (2) knows two people called Diane – $Diane_1$ and $Diane_2$. Furthermore, B knows *of* another person called Diane, $Diane_3$, whom only A knows in person. Now, $Diane_1$ is known only to B, $Diane_2$ is known mutually to A and B, and $Diane_3$ is known only to A (B just knows *of* $Diane_3$). Accordingly, two factors related to (2) – (a) the utterance of the form *Diane* and (b) the utterance of it by A – are likely to constrain the interpretation to mutually known $Diane_2$. This is because (a) and (b), acting as cues, are likely to activate memory records with associations between (the concepts of) A, B and $Diane_2$. Such commonality assessment seems personal (because of the associations between A, B and $Diane_2$). However, the commonality assessment relevant to the interpretation of *Diane* in (2) may also be more communal in that the discourse preceding the utterance of *Diane* already constrains the interpretation to individuals who are co-present for A and B, but also to those who are co-present for Shana, Jessica and Beth.

In summary, Horton & Gerrig's (2005) and Horton's (2008) claim is that interlocutor-specific information, i.e. information about common ground, is one of many cues which are simultaneously integrated during language processing and that it serves as an immediate constraint on language processing. In the next section, I use Horton & Gerrig's account of common ground to explain the weak and the strong interpretations of conditionals.

5.1.2 *Weak and strong interpretations of conditionals*

So far, I have assumed, in line with Beller's (2002) findings, that the interpretation of conditional promises is modelled by MI, whereas the interpretation of conditional threats is not. Here, I argue that it is actually difficult to maintain such a neat distinction. Consider (14) from chapter 4, repeated here as (3).

(3) If you pass the exam, I'll take you to a restaurant.

Let us imagine a scenario – scenario (i) – on which the speaker of (3) is an over-indulgent grandmother of the hearer. In scenario (i), the hearer stores memory records of grandma promising *q* if *p*, where *p* corresponds to a successfully completed task (by the hearer or other member of his family) and *q* is a promised treat for *p*. Crucially, the hearer stores memory records of non-completion of a given task followed by the occurrence of treat. Put simply, due to grandma's tendency for over-indulgence, the

hearer remembers occasions on which *q* was T even though *p* was F. Thus, in scenario (i), the product of personal commonality assessment is likely to dictate to the hearer that he can make no assumptions as to what will happen if he fails the exam. In other words, the hearer is likely to interpret the indulgent grandma as communicating that from the truth of *p* the truth of *q* will follow, and no more than that. On this interpretation of (3), the grandmother is not taken to have communicated anything about what she will do if the hearer fails the exam and thus it would not be surprising for the hearer to ask '*What if I don't pass the exam?*'. This interpretation is modelled by MI; in particular, it involves the inference of Modus Ponens (MP) which is licensed by MI. I will refer to such an interpretation as the WEAK interpretation of conditionals.

Indeed, if the indulgent grandma wanted to communicate and be taken to communicate that *q* iff *p*, the product of commonality assessment would most probably dictate to her that the form in (3) with an unmarked intonation pattern would not be sufficient to communicate that *q* iff *p* to this particular interlocutor. Thus, if the indulgent grandma wanted to communicate to her grandson that *q* iff *p*, she would be likely to put stress on *if*. Putting the stress on *if* would have the (intended) effect of making the shared background assumptions – that is, the assumptions about the grandma's tendency for over-indulgence – irrelevant to this particular conversation. Alternatively, the indulgent grandma could use the more specific (4) to successfully communicate that *q* iff *p*.

(4) If you pass the exam, I'll take you to a restaurant but if you don't, I won't.

Thus, it is the utterance of (3) (with an unmarked intonation pattern), combined with the fact that it is uttered by a particular interlocutor (the hearer's grandma who is known to the hearer to be over-indulgent) which give rise to the weak interpretation. Similarly, the grandma's assumption that the form in (3) uttered with unmarked intonation is not sufficient to communicate that *q* iff *p* arises due to her assumption of common ground with this particular hearer.

However, it is also plausible to imagine a scenario – scenario (ii) – in which the speaker of (3) is the very strict father of the hearer. As in (i), so here, the utterance of (3) combined with the fact that it is uttered by the strict father of the hearer function as cues to activate records of relevant past experiences. In (ii), the hearer stores memory records of the father promising *q* (a treat) if *p* (an undertaking is successful) and also of consistent cause-effect associations between unsuccessful undertakings and ensuing

lack of treat. Given such commonality assessment, in scenario (ii), the hearer of (3) is likely to interpret his father as communicating that *q* iff *p*.

It is plausible to argue that given the personal commonality assessment relevant to scenario (ii), choosing a more precise form such as (4) above to communicate that *q* iff *p* would simply be redundant. Indeed, given the shared assumptions about the father's strictness, it would be unlikely for the hearer to ask '*What if I don't pass the exam?*'.

The strict father scenario interpretation is modelled by equivalence; in particular, it involves the inference of Denying the Antecedent (DA) which is licensed by equivalence, but blocked by MI. I will refer to such an interpretation as the STRONG interpretation of conditionals.

So far, I have argued that conditional promises do not invariably involve the weak interpretation. But do conditional threats invariably involve the strong interpretation, as suggested by Beller's (2002) study? Consider (15) from chapter 4, repeated here as (5).

(5) If you do something illegal again, we'll lock you up.

Let us imagine a scenario in which the speaker of (5) is a police officer in a lawful state. Horton & Gerrig's (2005) account allows us to explain why it is possible to successfully communicate that *q* iff *p* by uttering (5). In the scenario where the speaker is a police officer in a lawful state and the hearer is a member of that state, the communal commonality assessment is likely to yield the assumption in the speaker's mind that the hearer shares with the speaker similar assumptions about legal aspects of social order. Presumably, the shared assumptions about the social order include the assumptions about what the speaker is permitted to do in a situation when *p* of a conditional threat is F, i.e. the assumption that with threats there is no social/legal consent for the speaker to make *q* T when *p* is F. Indeed, it would seem rather odd for the hearer of (5) to respond to the utterance of (5) with the question '*What if I don't steal again?'*; the contextual assumptions about the legal aspects of social order in a lawful state are such that the succinct form of (5) to communicate that *q* iff *p* is informative enough.

However, it is not implausible to assume that conditional threats may receive a weak interpretation. Consider a scenario in which Anna and Mary are at the party. Anna is very irritated because she has just been verbally offended by Mark. She says to Mary:

(6) If he talks to me like that again, I'll tell him what I really think about him.

Anna has just threatened that if *p* is T then *q* will be T, but has made no undertaking that if not *p* then not *q*. Indeed, even though it is plausible to interpret (6) as a threat, it is not clear that Anna could be judged to have spoken falsely if *p* is F and *q* T – if, for example, Mark offends someone else at the party or Anna gets sufficiently drunk. In fact, given that Anna has been verbally offended by Mark, she is permitted to make *q* T in the absence of *p*, i.e. in the absence of further offence by Mark.[3] The difference between (5) and (6) is brought about by the difference in the nature and social context of the threat. Whereas the social context of the threat in (5) is legally limited, the one in (6) is not. Thus, whereas the product of communal commonality assessment will constrain the interpretation of (5) to iff, there is nothing in the (cognitive) context of (6) to place such a constraint.

As discussed in 4.1.2.4 and in this section, the strong (i.e. biconditional) interpretation of conditionals is not limited to any homogenous group of examples: conditional promises, threats, injunctions and universal conditionals can all receive a weak or strong interpretation.[4] We have seen that Smith & Smith (1988) argue for a relevance theoretic explanation of the strong interpretation; because the uttering of "p", like any utterance, carries with it a presumption of its relevance, the speaker assumes that the hearer will be in position to assume that *p* is relevant in the context of *q*.

However, I argue that this account of Smith & Smith's (1988) is actually too general to explain the difference between the weak and the strong interpretations. The problem for Smith & Smith is that *p* is relevant to *q* even on the weak interpretation. It must be relevant if we assume that this interpretation involves the MP inference; if *p* were not relevant to *q* on the weak interpretation, how could we ever infer that the speaker has communicated that from the truth of *p* the truth of *q* will follow? Consider also (10) from chapter 4, repeated here as (7).

(7) If Tom can drive a truck, then I am Mickey Mouse.

[3] Beller's (2002) study is important as it shows a significant difference between the interpretation of conditional threats and promises. However, it has a shortcoming – due to the artificial setting of Beller's experiment, his study could not possibly show the fine-graininess of conditional inducement interpretation. Because Beller's subjects were interpreting written conditional statements, no relevant assumptions about the interlocutors could be made and thus no interlocutor specific assumptions could place immediate constraints on the interpretation of inducements.

[4] Arguably, the product of personal commonality assessment influences the interpretation of conditional inducements more than it does the interpretation of universal conditionals. As discussed in 4.2.2.4, whether a universal conditional receives, and is indented to receive, a weak or strong interpretation depends on the cogniser's (so-called 'encyclopaedic') knowledge. It seems that because conditional inducements are used to influence the hearer's behaviour, the information about interlocutors is highly salient in this context and thus the personal commonality assessment is crucially involved in message formation and interpretation. However, what universal conditionals, conditional inducements and injunctions have in common is that their interpretation goes through a holistic belief system.

Example (7) involves the weak interpretation – the speaker is communicating that from the patent falsity of *q*, the falsity of *p* follows. Again, in contrast to what follows from Smith & Smith's argument, it is not plausible to assume that because (7) does not involve the strong interpretation, *q* is irrelevant to *p*. Clearly, (7) involves an invitation to perform MT; the speaker is communicating that the falsity of *q* is relevant to what she thinks the truth value of *p* is. But how could the hearer perform MT without assuming that because the speaker has uttered "q" in the context of *p*, the falsity of *q* must be relevant to what the speaker is communicating about *p*? The difference between the strong and weak interpretations, I conclude, cannot consist in the difference in the relevance of *p* to *q* or lack of it (or in the relevance of *q* to *p* or lack of it). If, as Smith & Smith argue, the strong interpretation is due to the relevance of the fact that the speaker has uttered "p", all conditionals would receive a strong interpretation. But they do not.

At issue here is not the question of whether *p* is relevant to *q* (any utterance will convey the assumption of its own relevance, as argued in RT), but whether *p* is relevant to *q* in a way which allows us to perform inferences licensed by MI or those licensed by equivalence.

As argued in this section, what sort of inferences we make, and are assumed to be able to make, will depend on the (cognitive) context, including the assumption of common ground (as in the over-indulgent grandma and strict father scenarios). Unlike earlier approaches to the interpretation of conditionals, a radically contextualist, wholly pragmatic approach defended in this thesis is able to explain the holistic (cognitive) contextual conditions under which the strong and the weak interpretations of conditional statements arise.

5.2 Basic and extended uses

In this section, I discus a distinction between basic and extended uses of conditionals. On basic uses of conditionals, *if* signals that there is some relation between two propositions in a given cognitive context, whereas on extended uses, *if* signals that there is a relation between a proposition and an utterance (or speech act). However, before I say more about this distinction, I explain why van der Auwera's (1985) and Sweetser's (1990) sufficient conditionality approach, as well as Relevance Theory's notion of procedural semantics, are inadequate to handle the variation found in the interpretation of conditionals.

5.2.1 *A digression on sufficient conditionality and procedural 'meaning'*

In 5.1.2, I argued, in contrast to Smith & Smith (1988), that the notion of relevance of *p* to *q* is not distinctive of the strong interpretation of conditionals. Indeed, any use of *if p, q* signifies that *p* and *q* are relevant to each other in some way or another. One way of thinking about the relevance of *p* to *q* is in terms of the notion of sufficient conditionality, which was first proposed by van der Auwera (1985) and later developed by Sweetser (1990).

5.2.1.1 *Sufficient conditionality*

Van der Auwera (1985) puts forward the Sufficient Conditionality Thesis, where it is argued that *if* is a modal operator, requiring some consequential relation – the sufficiency condition – to hold between *p* and *q* for *if p, q* to be T. For van der Auwera, the sufficiency condition constitutes the encoded non-truth-functional semantics of *if*.

Sweetser (1990) argues that van der Auwera's sufficient conditionality approach can be maintained as long as it is recognised that conditionality operates in three distinct cognitive domains: the content domain, the epistemic domain and the speech act domain. In the content domain, one state of affairs (*p*) is a sufficient condition for the occurrence of another (*q*). In the epistemic domain, a given premise (*p*) is a sufficient condition to derive a conclusion (*q*). In the speech act domain, *p* is a sufficient condition for the speech act made in uttering "q". However, there are problems with the sufficient conditionality account. Consider (23) from chapter 4, repeated below as (8).

(8) A: Two and eleven makes thirty.
B: If two and eleven makes thirty, you need more work on maths.

Noh (2000) argues – rightly, I think – that the conditional in (8B) cannot be explained in terms of sufficient conditionality. In the content domain, it cannot be argued that the truth of the antecedent is sufficient for the truth of the consequent because the antecedent is obviously false. The problem also arises in the epistemic and speech act domains – the truth of the antecedent cannot be sufficient for concluding or asserting (respectively) that the referent of *you* needs more work on maths because the antecedent is false. This shows that sufficient conditionality is not the unitary non-truth-functional aspect of the encoded semantics of *if* (contrary to what van der Auwera and Sweetser argue).

Indeed, in some instances of 'conditional' suggestions/advice, the antecedent clearly is not a condition on *q*. For example, in chapter 4, I gave example (21) (*If you want to make your parents happy, quit smoking*) and argued that the antecedent in (21) does not represent a condition on the consequent. Rather, the antecedent represents a (given) *goal*, which is embedded in the opaque context represented by *want* (suggesting that the truth of *p* is desired by the speaker). This goal, it is suggested by the speaker of (21), can be accomplished by *means* of *q*. In fact, if there is conditionality here, it runs in the opposite direction – in (21), it is *q* which represents a condition for the achievement of the goal in *p*! I return to this point shortly.

The second problem pertains to the fact that van der Auwera's (1985) sufficient conditionality account is a non-truth-functional account of the encoded semantics of *if*. As noted in 4.2.1 (footnote 12), non-truth-functional approaches, like truth-functional approaches, make no prediction as to the truth value of the consequent when the antecedent is F ([$\sim p \nvdash (p \supset q)$]). Consequently, the problem of conditional belief deviations from MI arises for van der Auwera's account just as it does for truth-functional approaches. For this reason, it cannot be maintained.

Furthermore, as shown in 4.2.2.5, Noh's (2000) metarepresentational MI analysis of (8B) is also problematic: it gives rise to the problem of pragmatic intrusion into encoded semantics, it undermines the theoretical significance of the notion of metarepresentation and, because of Noh's focus on metarepresentational faithfulness, it fails to acknowledge the relevance of the falsity of *p* to the truth of *q*. Clearly, another account in needed.

5.2.1.2 *Procedural 'meaning'*

In the preceding chapters, I provided many arguments against the notion of encoded semantics, in general, and encoded semantics of *if*, in particular. However, I need to return to this issue once more.

In Relevance Theory (e.g. Carston 2002, 2010), words encode concept schemas, full-fledged lexical concepts (as discussed in chapter 2) or procedural 'meanings'. These procedural 'meanings' are not concepts. Rather, they are constraints on pragmatic inference, i.e. they are processing instructions (Blakemore 1992). Consider (9), taken from Blakemore.

(9) (a) John can open Bill's safe. (b) He knows the combination.

Faced with (9), the hearer may not know how to process the relation between the propositions communicated by (9a) and (9b); there is no way of telling whether (a) is a premise and (b) a conclusion, or the other way round. If the speaker of (9) wanted to communicate that (a) is a premise and (b) a conclusion, she would use *so* in order to save the hearer the processing effort. Conversely, if the speaker of (9) wanted to communicate that (b) is a premise and (a) a conclusion, she would use *after all*.

Relevance Theory argues that, among other expressions, discourse markers (of which *so* and *after all* are examples) encode procedural information. Accordingly, *so* and *after all* encode instructions to the hearer on how to process the relation between the propositions communicated by (a) and (b). In other words, they specify the way in which (a) and (b) are relevant to each other. The question that needs to be explored is whether it can be argued that *if* encodes a procedure of that sort.

Hussein (2008: 77-78), working in Relevance Theory, makes a distinction between REAL conditionals and METAREPRESENTATIONAL conditionals. Hussein's (2008) metarepresentational category parallels Sweetser's speech act domain. His real conditional category subsumes uses in the content as well as the cognitive domain.[5] On real conditional uses, the word *if* operates at a representational (i.e. propositional) level where it relates two propositions. According to Hussein (2008, 2009), real conditional *if* semantically encodes MI plus some non-truth-functional causal-consequential relation.

On metarepresentational uses, however, *if* does not relate two propositions but is used to introduce a reason for uttering "q". The interpretation of metarepresentational use is, according to Hussein (2008), independent of MI; here, Hussein (2008: 78) argues, '*if* does not contribute to the semantic representation of the conditional but plays a role in the inferential part of the conditional interpretation by constraining the relevance of the second clause'.

Now, Hussein's claim that real conditional *if* semantically encodes MI cannot be maintained for the reasons discussed in chapter 4. But could Hussein's claim that metarepresentational *if* encodes a procedural constraint on the 'relevance of the second clause' be maintained and perhaps extended to real conditionals? Consider (10).

(10) John can open Bill's safe, so he knows the combination.

As discussed earlier, on RT's account, *so* encodes an instruction to process what precedes it as a premise and what follows as a conclusion. Thus, in (10), knowing that

[5] Noh (2000) also argues that Sweetser's content and epistemic domains fall under one category of descriptive uses (as opposed to metarepresentational uses).

John can open Bill's safe allows the speaker to conclude that John knows the combination. Consider now (11).

(11) If John can open Bill's safe, he knows the combination.

When interpreted as operating in the epistemic domain, it could be argued that, parallel to the use of *so* in (10), the use of *if* in (11) instructs the hearer to process *p* a premise on the basis of which the conclusion *q* can be drawn. However, the problem with this assumption is that it makes the encoded semantics of *if* and *so* identical. Notice that invoking some encoded non-truth-functional causal-consequential relation (e.g. sufficient conditionality) would not save such an argument either, as causal-consequential (more specifically, inferential) relation is present in both (10) and (11).[6]

Furthermore, the nature of the purportedly encoded instruction would vary depending on the context. Thus, in (11) above, *if* would have to instruct the hearer to treat *p* as a premise and *q* as a conclusion. But in (7) (*If Tom can drive a truck, then I am Mickey Mouse*), *if* would have to instruct the hearer to treat *q* as a premise and *p* as a conclusion. In (5) (*If you do something illegal again, we'll lock you up*), the instruction would be to treat *p* as a cause and *q* as effect (on the content domain interpretation). However, as argued earlier, the speaker of (21) (*If you want to make your parents happy, quit smoking*) suggests that the goal represented in *p* can be caused by *q*. Consider also (12), taken from Comrie (1986).

(12) If it will amuse you, I'll tell you a joke.

Comrie (1986: 81) observes that in (12), the causal relation holds in both directions – my telling a joke (*q*) is the cause of you being amused (*p*), but also the hearer's future amusement (*p*) is the cause for the speaker telling the joke (*q*).

Now, if the inferential (i.e. premise-conclusion) relation can hold in any direction, as illustrated by (11) and (7), and if the causal relation can hold in both directions, as in (12), *if* would have to encode a very general procedure. To complicate things yet further, there are also speech-act domain uses. Consider (13), taken from Hussein (2008).

[6] As observed by Strawson (1986), *p* and *q* are not asserted in '*if p … q'*, but they are asserted in '*p so q'*. However, as we have seen (section 4.2.2.5), some uses of '*if p …q*' involve assertion. For example, in (39) (*If you're thirsty, there's beer in the fridge*), *q* is asserted. Arguably (and depending on the definition of assertion, see e.g. MacFarlane 2010), the truth of *p* in (21) (*If you want to make your parents happy, quit smoking*) is as non-debatable (i.e. taken for granted, in common ground) as is the truth of '*John can open Bill's safe*' in (10). Of course, the truth of *p* in (21) is non-debatable on the interpretation where *p* is interpreted as a goal to which *q* is a means (i.e. when *p* is the topic). I return to this point in 5.2.2.2.

(13) If you are thirsty, there is a lemon juice in the fridge.

If in (13) is treated by Hussein (2008: 78) as encoding a procedural instruction to constrain the relevance of the second clause. According to Hussein, in (13), the fact that there is lemon juice in the fridge is relevant to the person referred to in the *if*-clause, but no inferential or causal-consequential relation holds between the propositions expressed by *p* and *q* in any direction.[7]

Because of the variation in the ways in which *p* and *q* can be relevant to each other, if *if* encoded some procedural 'meaning', it could only be an instruction to the hearer that *p* and *q* are relevant to each other in some way or another. However, given that relevance is a pragmatic (cognitive) notion, no such instruction can be semantically encoded without undermining the claim that encoded semantics is autonomous with respect to pragmatics. Indeed, since in RT, every utterance carries with it a presumption of its own relevance and since relevance is established in context, encoding such an instruction would simply be redundant.

To conclude this section, I have argued that neither the sufficient conditionality approach nor the notion of procedural semantics is able to handle the variation concerning the relevance of *p* to *q*.[8]

5.2.2 *Basic uses*

Modulo differences (discussed above), it seems to me that Hussein's (2008) two-fold distinction – between *if* which relates two propositions and *if* which relates a proposition with an utterance – is on the right track. Accordingly, I will employ the term BASIC USE for cases where *if* is used to communicate that there is some relation (on which below) between two propositions (*p, q*). I will employ the term EXTENDED USE for cases where *if* is used to communicate that there is some relation (on which below) between a proposition *p*, and the utterance of "q". In the remainder of this section, I discuss basic uses.

[7] From the hearer's thirst it cannot be concluded that there is lemon juice in the fridge (by MP), nor would it follow (by MT) that if there is no juice in the fridge, then the hearer is not thirsty. The hearer's thirst is not a cause or reason for the presence of the juice in the fridge, nor is the absence of the juice in the fridge a cause or reason for the absence of hearer's thirst.

[8] A procedural account can be saved as long as it does not involve commitment to encoding. That is, it can be saved as long as procedures are understood as emerging from the interaction between primary and secondary memory systems (in the sense of Hintzman 1986) and defined in terms of acquiring a relation between a morpho-phonetic label and an aggregate of conceptual structures. On this view, a procedure equals a (representational) convention, as defined in 2.2.3.4. However, I prefer to use the term 'convention' to stress my rejection of encoding.

5.2.2.1 *Basic uses and holistic inference*

I argue that in its basic use, *if* serves to indicate that *p* and *q* are relevant to each other so that it will be possible to infer the truth value of one with respect to the other. In line with the arguments presented in chapter 4, I will show that the establishing of the truth value of *q* with respect to *p* or *p* with respect to *q* always goes through in a holistic system of thought. I will show that the establishing of the truth value of one proposition with respect to the other is sometimes modelled by MI, sometimes by equivalence, and sometimes determined by other factors. Consider (9) from chapter 4, repeated here as (14).

(14) If it's a square, it has four sides.

The interpretation of (14) is modelled by MI – it allows the hearer to perform MP and MT, but not DA or AC. However, the fact that (14) is modelled by MI (and that we reject the validity of DA and AC) relies on a holistic inference – in order to know what inferences can and cannot be performed in the case of (14), one needs to know that rectangles too have four sides. It is in virtue of this premise, which is held in the holistic belief system, that the hearer of (14) can establish that the truth of *p* is relevant to the truth of *q* so that the truth of *q* follows from the truth of *p*, but not vice versa, and that the falsity of *q* is relevant to the falsity of *p* so that the falsity of *p* follows from the falsity of *q*, but not vice versa. Consider also example (15) (due to Noel Burton-Roberts).

(15) I can't exactly remember the shape of the figure that John drew, but if it wasn't a square, it was certainly a triangle.

MI falsely predicts that if it actually was a square (antecedent is F), then we can make no inference – in the light of (15) – as to whether the figure was a triangle or not. But what the speaker of (15) is committing herself to is that from the falsity of *p* in (15), the falsity of *q* follows. That is, what the speaker is communicating in (15) is modelled by equivalence (DA). As above, in order to know what inferences can and cannot be performed in the case of (15), one needs to know that nothing can be both a square and triangle. It is in virtue of this premise, which is held at the holistic level – and in the light of (15), which restricts the options to either a triangle or a square – that the hearer can establish that in the case of (15): the truth of *p* is relevant to the truth of *q* so that the truth of *q* follows from the truth of *p*, and vice versa, and that the falsity of *q* is relevant

to the falsity of *p* so that the falsity of *p* follows from the falsity of *q*, and vice versa. Thus, as was the case in (14), the inferential process triggered by the utterance of (15), takes place in a holistic domain of thought. Indeed, the correctness of the inferential process depends on its taking place in a holistic domain. Let us now re-consider (8).

(8) A: Two and eleven makes thirty.
B: If two and eleven makes thirty, you need more work on maths.

Arguably, (8B) can be analysed in three ways – (a) as representing the speaker's train of reasoning (reaching a conclusion *q*), (b) as communicating that *p* is F, or (c) as suggesting that the hearer makes *q* T. On interpretation (a), from A's utterance and two premises: (i) that A believes *x* (a mathematical belief that two eleven makes thirty) is T and (ii) that *x* is F, B draws a conclusion (iii) that A needs more work on maths.[9] But the second premise (ii) holds in holistic thought (in fact, both premises do). Thus, the inference from A's utterance to the conclusion (iii) must go through the holistic system; indeed, the plausibility of conclusion (iii) crucially depends on this holistic premise (ii).[10]

On interpretation (b), (8B) is used to communicate that two and eleven does not make thirty, i.e. that *p* is F. Presumably, A needs to somehow infer that *p* is F from the utterance of (8B) and the assumption that B's conclusion (*q*) is T. This inference clearly is not licensed by MI or equivalence. However, it can be explained once it is acknowledged that it goes through in the holistic system of thought; from three assumptions – (iv) that *q* is T, (v) that one needs more work on maths if one does not have mathematical knowledge which is relevant in the current context, and (vi) that the relevant mathematical knowledge concerns A's belief that two and eleven makes thirty, A can infer that B communicates that A's belief that two and eleven makes thirty is actually F.

Interpretation (c) seems to depend on (b). Notice that if (8B) is to be interpreted as a suggestion that A needs more work on maths, the falsity of A's belief that two and eleven makes thirty needs to be assumed (acknowledged). If (8B) suggests anything, it suggests a means to resolve what B perceives as a contradiction – i.e. A's F

[9] More specifically, from the premises (i) and (ii), B concludes that (i+ii) A holds a false belief. It is the conclusion (i+ii), which serves as a premise in the inference to (iii).

[10] As discussed in chapter 4, Noh's metarepresentational MI analysis of (8) fails to acknowledge that the faithfulness of the metarepresentational antecedent to what it metarepresents is NOT a reason for concluding that A needs more work on maths. Indeed, it is the falsity of what A represented in (8A), which is the reason. Thus, neither a metarepresentational MI analysis nor a descriptive level MI analysis can be maintained. The relation between *p* and *q* in (8B) is much more fine-grained.

belief that what is F is T. Indeed, I suggest that it is the logically driven need for the resolution of (perceived) contradiction – which is perceived in virtue of (iv), (v) and (vi) – which explains why the falsity of *p* in (8B) can be inferred and thus communicated.

Relatedly, the use of *if* in the context of A's utterance in (8) indicates that the truth value of *p* is at stake here – it communicates that, contrary to what A suggests/argues, the truth of the proposition communicated in (8A) should not be taken for granted. Notice that, *if* in (8B) cannot be exchanged with *given that*, which would indicate that that the truth of the proposition communicated in (8A) was taken for granted. Incidentally, *given that* could only be used if followed by *you believe*. This is because the addition of *you believe* would indicate that B is not committed to the truth of the proposition communicated in (8A) – the use of *if* seems a more succinct way to achieve the same communicative goal in this example. Consider (7) again, repeated below.

(7) If Tom can drive a truck, then I am Mickey Mouse.

Earlier, I suggested that (7) involves an invitation to perform MT (licensed by MI), i.e. to infer from the patent falsity of *q* that the speaker believes *p* is F. Now, the most probable occurrence of (7) is in the context of a response to a previous utterance, as in (16).

(16) A: Tom says he can drive a truck.
B: If Tom can drive a truck, then I am Mickey Mouse.

As was the case with (8B), so in (16B), the use of *if* signifies that B does not take for granted the truth value of the proposition communicated in (16A). By representing a patently F proposition in *q* – i.e. a proposition whose truth value is manifest – B is unambiguously inviting A to infer that B thinks *p* is F. Indeed, if (16B) is preceded by B's laughter in response to A's utterance, or if B utters the *if*-clause with the falling intonation (which indicates A's disbelief in Tom's ability), there is nothing in the context that prevents A from assuming that B thinks *p* is F. This assumption of A is then *confirmed* by the utterance of the second clause. Now, MP can also be performed in the case of (16B) – if you believe that Tom can drive a truck (*p* is T), you must believe that I am Mickey Mouse (*q* is T). However, because the falsity of *q* is patent, the assumption that B is inviting the hearer to establish the truth value of *q* with respect to *p* gives rise to a contradiction. This contradiction is resolved when the hearer realises that the

speaker's communicative intention is for A to infer the truth value of *p* with respect to the falsity of *q*.

Crucially, in order to know that the hearer is being invited to perform MT, the hearer must consult his ('encyclopaedic') knowledge to establish that *q* is F. Thus, the inference to the falsity of *p* must go through in the holistic system of thought. Indeed, the inferring of the falsity of *p* via MT (rather than by way of contradiction resolution as in (8B) *If two and eleven makes thirty, you need more work on maths*) is a result of a holistic interpretive process. Consider also (32) and (21) from chapter 4, repeated here as (17) and (18), respectively.

(17) If you press the red button, the conveyor belt will stop.

(18) If you want to make your parents happy, quit smoking.

As argued in chapter 4, (17) is a conditional instruction on what the hearer should do (i.e. he should make *p* T) to make *q* T. I suggested that this interpretation is not modelled by MI because according to MI, it does not matter whether *p* is T or F as long as *q* is T (because $q \vdash (p \supset q)$). The problem is that if (17) is to be understood as an instruction on what the hearer should do to make *q* T, the context of (17) must exclude a situation where there is a power cut (etc.) and the belt has stopped, as in such a situation there would be no point instructing the hearer on what to do to stop it. The (cognitive) context of (17) must be limited to the action that is supposed to be performed by the hearer. Thus, the assumption that the relation between the truth values of *p* and *q* in (17) is modelled by equivalence follows from the contextual limitation to the action performed by the hearer, which, in turn, depends on (17) being an instruction.

In chapter 4, I also argued that the antecedent of (18) does not represent a condition on the consequent but a goal which is embedded in the propositional attitude represented by *want* (suggesting that the truth of the embedded proposition is desired by the speaker). Accordingly, (18) is used to suggest that the truth of the proposition embedded in *p* (the truth which is desired by the hearer), can be accomplished by means of making *q* T. Notice that the truth of *p* cannot be debatable on the assumption that the speaker utters (18) because she knows that *p* is T (*p* is the topic). But how is this intended interpretation – that the truth of the proposition embedded in *p* will follow from the truth of *q* – arrived at? The fact that it can be communicated (and inferred) that the truth of the proposition embedded in *p* will follow from the truth of *q* suggests that this inference is modelled by equivalence (the inference is AC). Indeed, to perform AC, in the speaker's (hearer's) mind, the proposition (about the state of affairs) that the

hearer's parents are unhappy (~H) and the proposition (about the state of affairs) that the hearer smokes (S) must be logically equivalent. Because these propositions are equivalent in the speaker's (hearer's) mind, the only way to make (~H) false (i.e. to make (H) true) – which the hearer wants – is to make (S) false. Thus, (i) the equivalence relation between (~H) and (S) – which can only be held in holistic thought – and (ii) the assumption that the hearer wants to make (~H) false (also held in holistic thought), allow to communicate (and infer) that the truth of (~S) will result in the truth of (H).

So far, I have argued that in its basic use – i.e. where it signifies a relation between two propositions – *if* serves to indicate that *p* and *q* are relevant to each other so that it will be possible to infer the truth value of one with respect to the other. I have sought to show that the establishing of the truth value of one proposition with respect to the other always goes through in the holistic system of thought. Depending on a holistic interpretive process, the establishing of the truth value of one proposition with respect to the other is sometimes modelled by MI – as in (14) and (16B); sometimes by equivalence – as in (15), (17) and (18); and sometimes by the logically motivated need for the resolution of a perceived contradiction – as in (8B).

5.2.2.2 *A note on directionality of inference*

There is one more comment that needs to be made with regards to the variation found in the basic uses of *if*. Now, I have been arguing that *if* serves to indicate that *p* and *q* are relevant to each other so that it will be possible to infer the truth value of one, i.e. ANY one, with respect to the other. If I am right in arguing that, we should be able to find cases where the truth values of both propositions are open and thus mutually relative. However, we should also be able to find examples where the truth value of *q* is taken for granted (and serves as a premise), and the truth value of *p* is inferred on that basis. Conversely, we should also be able to find examples where the truth value of *p* is taken for granted (and serves as a premise), and the truth value of *q* is inferred on that basis. If true, this prediction would be in contrast to Strawson (1986), who argues that, in '*if p … q*', *p* and *q* are not asserted (see fn 6), and to Grice (1989), for whom it would not be rational to use a conditional where, for example, there is no doubt about the truth of *p* (see Chapman 2005: 108).

My prediction, I argue, is borne out. Accordingly, in (16B) (*If Tom can drive a truck, then I am Mickey Mouse*), the truth value of *p* is established with respect to the patent falsity of *q*; and in (8) (*If two and eleven makes thirty, you need more work on*

maths), the truth value of *p* is established with respect to the assumed truth of *q* (on interpretation b). Conversely, in a context where (14) (*If it's a square, it has four sides*) is uttered as a judgemental response to the question '*How many sides does this square have?*', the truth value of *q* is established with respect to the assumed truth of *p*. The same obtains for (19), which does not involve a judgemental reaction.

(19) A: A ten pence coin fits through that slot.
B: If a ten pence coin fits through that slot, a five pence coin will fit through that slot.

Let us assume that B knows that A has empirically confirmed that a ten pence coin fits through the relevant slot. In this context, we can assume that A has truly communicated that a ten pence coin fits through that slot and that *p* in (19B) is non-debatably taken as T. Thus, the truth value of *q* can be established with respect to the empirically confirmed (and contextually given) truth value of *p* via MP. Similarly, in the case of example (11) from chapter 4 (*If two plus two equals four, then my client is innocent*), the truth value of *q* is established with respect to the patent truth of *p*.

However, in the context of (15) (*I can't exactly remember the shape of the figure that John drew, but if it wasn't a square, it was certainly a triangle*), it is manifest that the truth value of neither proposition is to be taken for granted, but because *p* and *q* represent the only two options, their truth values, which are relative with respect to one another, can be calculated as predicted by equivalence. The truth values are also open in (17) (*If you press the red button, the conveyor belt will stop*) and the calculation of relative truth values proceeds as predicted by equivalence. Incidentally, both truth values are open in (14) (*If it's a square, it has four sides*) if it is said of an unseen geometric figure (i.e. a figure hidden in a box). However, because of the hearer's knowledge that other figures can have four sides, the calculation of relative truth values proceeds as predicted by MI.

5.2.3 *Extended uses*

In extended use, *if* is used to communicate that there is some relation between a proposition *p*, and the utterance of "q". Consider (39) from chapter 4, repeated here as (20).

(20) If you're thirsty, there's beer in the fridge.

An extended use analysis of (20) is simpler and less problematic than Noh's metarepresentational analysis (discussed in chapter 4). On an extended use analysis, the relation of relevance holds between the proposition/thought communicated by *p* and the uttering of "q" (or the speech act of informing the hearer that there is beer in the fridge). This correctly predicts that *p* is used to communicate a reason for the uttering of "q" – *p* explains why the speaker is bothering to utter "q", without undermining the theoretical significance of the notion of metarepresentation (as was the case with Noh's account). Notice that the extended use analysis explains why there is no truth-functional relation involved here, i.e. why the truth of *p* is independent of the truth of *q* and vice versa. Even though a conditional form is used, there can be no truth-functional relation here because utterances (and speech acts) are not propositional objects. It is thus the lack of truth-functional relation between *p* and *q* – which is due to the relation holding between a proposition and an utterance (speech act) – which makes extended uses different from basic uses.[11] Consider also (21).

(21) If you're thinking of ordering drinks, I've already ordered them.

Similarly to (20), in (21), *p* is used to communicate the reason for uttering "q" (or the speech act of informing the hearer that the speaker has already ordered drinks). The reason why the speaker is informing the hearer that the speaker has already ordered drinks is that the speaker thinks the hearer may be thinking of ordering drinks, and thus the uttering of "q" will be relevant, as it will save the hearer an unnecessary trip to the bar. As in (20), so here, the truth value of *p* is independent of the truth value of *q*, and vice versa, because the relation of relevance holds between a proposition and an utterance (speech act). In the remainder of this section, I defend the extended use analysis presented here against Sweetser's sufficient conditionality speech-act analysis.

5.2.3.1 *Sweetser's speech-act analysis versus extended use analysis*

Consider (22), taken from Sweetser (1990).

(22) There are biscuits on the sideboard if you want them.

[11] If there is conditionality here, it is the relevance of the speaker's utterance of "*q*" which seems to be conditional on *p*. But this conditionality is not modelled by MI as it concerns the relation between a propositional object and a non-propositional object (i.e. an utterance). This view is better than Noh's (2000) metarepresentation proposal, which undermines the theoretical significance of the notion of metarepresentation (as argued in 4.2.2.5).

According to Sweetser (1990: 119-122), (22) lends itself to the sufficient conditionality analysis at the level of a speech act. Sweetser (*ibid*.) argues that in the speech act domain, the speech act is conditional upon what is represented in the antecedent. Thus, in (22) the offer of biscuits is conditional upon its potential acceptability to the hearer; in Sweetser's terms, a conditional speech-act structure for utterances like (22) is 'I hereby offer X, if X is a felicitous offer'. The problem with this analysis is discernible in the following quote from Sweetser:

> **Quote 1**: The major felicity condition for an offer is (as mentioned above) that the speaker assumes the hearer to want the thing offered. An offer is therefore infelicitous (or even irrelevant, in Gricean terms) if it fails in fact to respond to a desire or need on the part of the addressee. (Sweetser 1990: 122)

Whereas I agree with the first part of the above quote – that the speaker's assumption of the hearer's needs is a felicity condition (the weaker characterisation), I disagree with the second one – that the assumption's compatibility with the actual need of the hearer is a felicity condition (the stronger characterisation).

Given the stronger characterisation of the felicity condition in the second part of the quote, the offer in (22) must be infelicitous if the hearer does not actually want the biscuits, i.e. if the speaker of (22) has misinterpreted the hearer's needs. However, the problem with the predictions that Sweetser's analysis of (22) makes is that the offer in (22) simply has been made – and the relevance of "q" to *p* has been communicated – regardless of whether the hearer does or does not actually want the biscuits.[12] In other words, the performance of the speech act simply is not conditional upon the truth of *p* (its faithfulness to the actual needs of the hearer), but on the speaker's assumption that the hearer may want the biscuits (the weaker characterisation).

The extended use analysis offers a simpler solution to (22). Like in (20), so in (22), *p* communicates the speaker's reason for uttering "q" (or for the speech act of informing the hearer that there are biscuits on the sideboard). There is no truth-functional relation between there being biscuits on the sideboard and the hearer wanting them (the biscuits are there regardless of whether the hearer wants them or not), but the speaker finds the presence of the biscuits potentially relevant to the hearer. Indeed, the

[12] Austin (1975) makes a distinction between locutionary, illocutionary and perlocutionary forces of a speech act. A locutionary act is the act of uttering something. An illocutionary act is the act of signalling the particular speech act (e.g. a question, order, offer, etc.) intended by the speaker. A perlocutionary act has to do with an effect the performance of the speech act has on the hearer. In talking about felicity conditions, Sweetser (e.g. 1990: 118) is interested in the illocutionary force of the speech act.

lack of correlation between the speaker's assumptions about the hearer's needs and the hearer's actual needs does not invalidate the relevance of "q" to *p*. Consider now (23).

(23) If (as we both know) you were at the party, how's Harry these days?

Sweetser (1990: 131) argues that because the speech-act world is built up by references to it, it is in constant flux. Since it is constantly changing, it is, in Sweetser's words, 'intangible'. So just like the epistemic world, the speech-act world is not subject to direct examination. Given that, Sweetser (*ibid*.) suggests that a conditional presentation of a speech act allows the speaker to indicate how the speech act 'fits into the structure of the jointly constructed conversational world'. Thus, so the argument goes, in (23) the speaker's enquiry about Harry is relevant because it fits to the constructed discourse world. Sweetser argues that such display of conditional relevance gives the question a context. I entirely agree with that. However, in contrast to the weaker characterisation of the felicity condition, and consistently with the stronger characterisation, Sweetser (*ibid*.) argues that the enquiry is relevant on condition that the hearer has been to the party.

The problem with Sweetser's analysis is clearly discernible if we omit the clause in brackets (*as we both know*). Sweetser's analysis misses the crucial point that in (23) the relevance of the speaker asking the question ("q") is not conditional on the truth of *p* but on the speaker's assumption that the hearer may have been to the party. The use of conditional indicates that the question about Harry's well-being is not ad hoc but motivated by the speaker's assumption that the hearer may know how Harry is because the speaker may have been to the party. The extended use analysis correctly predicts that even if it turns out that the speaker has not actually been to the party and is not in position to answer the question, the question has been asked (the speech act has been performed) and the motivation for (relevance of) asking the question (performing the speech act) has been represented. Consider also the following examples (taken from Sweetser 1990: 118).

(24) If I may say so, that's a crazy idea.
(25) If it's not rude to ask, what made you decide to leave IBM?

As discussed earlier, for Sweetser (1990: 118) the performance of the speech acts ("q") in (24) and (25) is conditional upon the fulfilment of *p*. Sweetser rightly argues that even though (24) and (25) 'purport' to state an opinion or ask a question (respectively)

conditionally on the hearer's permission, such politeness conditions do not actually prevent the performance of the speech act. Nevertheless, in consistence with the stronger characterisation of the felicity condition, Sweetser (*ibid*.) argues that in these examples 'the statement is made and the question is asked – **but not quite fully**' (my emphasis) presumably because such politeness conditionals 'somehow do allow the hearer a little more room to maneuver'. But if, as Sweetser argues, the speech acts in (24) and (25) are 'not quite fully' performed, then it is not clear whether they are actually performed or not.

I suggest that the performance of speech acts in (24) and (25) is not conditional upon the hearer's permission – in fact, the speech acts have been performed regardless of whether the hearer likes it or not. These uses, however, are indicative of the relation between the performed speech acts and the speaker's contextual assumptions about social inappropriateness of stating the opinion in (24) and asking the question in (25). In (24) and (25), like in (22) and (23), the speaker's contextual assumptions are represented in the *if*-clause and therefore the speaker is communicating that there is a relevance relation between the represented contextual assumptions and the speech act in "q". In (24) and (25), the relevance is such that the speaker is performing the speech act not *because of* the contextual assumptions (as in (22) and (23)) but *despite* the contextual assumptions that it may be inappropriate. By social convention, the 'display' of the speaker's awareness of such potential inappropriateness makes the speech act less inappropriate.

5.2.4 *Interim conclusion: objectively messy, subjectively predictable*

In chapter 4 and earlier in this chapter, we saw that thus far the difference between the strong (biconditional) and the weak (MI) interpretations of conditionals has not been adequately explained. The iff interpretation cannot be explained by reference to scales because equivalence is not lexicalised. A further problem for Levinson's [$(p \equiv q) \vdash (p \supset q)$] scale is that the informativeness principle it invokes is trivial. A problem with Smith & Smith's (1988) relevance theoretic account is that *p* is relevant not only on the strong interpretation, but also on the weak interpretation.

Another problem with the strong interpretation is that it is not limited to any homogenous group of uses – conditional threats, promises, injunctions and universal conditionals can all be weakly or strongly interpreted. It has proven impossible to explain and predict the conditions under which these two interpretations arise in terms

of some necessary (objective) principles. I have argued that the distinction between the weak and strong interpretations of conditionals can be explained and predicted from a holistic/individualistic perspective. As shown in 5.1.2, it is only when we take into account the cognitive contexts of particular interlocutors in particular conversational situations, that we are able to pin down the *cognitive contextual (individualistic) conditions* under which the weak or the strong interpretation arises.

My point is not that meaning cannot be explained in objective terms. It can – for example, the difference between basic and extended uses of conditionals can be reliably distinguished by whether the relation signified by *if* holds between two propositional objects or not. Of course, establishing whether the relation holds between two propositional objects or not requires pragmatic (holistic) inference, but an inference about particular interlocutors (personal common ground) does not play a crucial role in establishing that. Rather, my point is that sometimes – as with weak and strong interpretations of conditionals – even though objectively messy, i.e. not governed by any necessary principle, the use IS subjectively predictable.

5.3 Implications for the explicit-implicit distinction

In 4.2.2, I argued that contrary to claims made in Carston (2000: 99-100), Relevance Theory's re-analysis of Grice's explicit-implicit distinction does not resolve the problem of pragmatic intrusion into encoded semantics. RT's problem is that their 'solution' to Grice's circle is to argue that linguistic semantics does not have truth-theoretic properties, but RT's definition of the explicit – and hence of the explicit-implicit distinction – depends on the argument that linguistic semantics *has* truth-theoretic properties. This problem, which I referred to as Relevance Theory's circle, also emerges in RT's claims about the acquisition of linguistic semantics, the nature of linguistic semantics, and, more specifically, in RT's (Noh 2000, Smith & Smith 1988) account of the encoded semantics of *if*.

RT's (and Grice's) problems with formalising the explicit-implicit distinction raise the question of whether this distinction can actually be defined in formal terms. In the next section, I discuss Gibbs' (2002) proposal that there do not seem to be 'hard and fast' rules which distinguish explicit communication from implicit communication.

5.3.1 *Gibbs on the explicit-implicit distinction*

Gibbs (2002) argues that there is no psycholinguistic evidence for the existence of 'some canonical, non-pragmatic meaning that is automatically analysed at both the word and sentence level'. In other words, Gibbs maintains that there is no evidence for the existence of linguistic semantics or lexical concepts, respectively. Nevertheless, Gibbs (*ibid*.) argues that empirical evidence strongly suggests that speakers do differentiate between what is explicitly and implicitly communicated.

Like RT, Gibbs (*ibid*.) rejects Grice's explicit-implicit distinction in terms of a distinction between semantics and pragmatics. However, because Gibbs rejects the notion of linguistic semantics (i.e. 'non-pragmatic content which is automatically analysed at both sentence and word level'), Gibbs cannot appeal to RT's explicit-implicit distinction. Put simply, explicature cannot be defined in terms of a logical development of what is linguistically encoded because, according to Gibbs, there is no linguistic encoding. So how does the explicit-implicit distinction arise?

Gibbs (2002: 476-477) discusses two kinds of pragmatic information/knowledge which are activated during language comprehension. PRIMARY PRAGMATIC KNOWLEDGE refers to deeply held background knowledge/beliefs, whereas SECONDARY PRAGMATIC KNOWLEDGE refers to local information which is available in the communicative context. Gibbs (*ibid*.) tentatively suggests that primary pragmatic knowledge is more involved in the derivation of what is explicitly communicated, whereas secondary, i.e. local, pragmatic knowledge is more involved in the derivation of what is implicated. Consider the following example.

(26) The cat is on the mat.

Gibbs (2002: 477, drawing on Searle 1978) suggests that the primary pragmatic knowledge involved in the derivation of what is explicitly communicated by (26) consists in assumptions such as that 'the cat chose for some reason to sit on the mat, and that the cat and mat are on the ground operating under the constraints of physical laws like gravity and are not floating in space in such a way that the cat is on the mat by virtue of touching the underneath part of the mat as in *The fly is on the ceiling*'. Secondary pragmatic knowledge involved in the derivation of what is implicated (i.e. that the addressee should let the cat outside) consists in recognising features of the local context such as the manifestly shared (between the speaker and the hearer) assumption that when the cat sits on the mat, it typically wants to go outside.

The first problem with this distinction is, as observed by Gibbs (*ibid*.) himself, that it is 'difficult, if not impossible' to decide which pragmatic knowledge is primary and which secondary. Gibbs' suggestion (drawn from Recanati 1993) is that perhaps primary pragmatic knowledge is more salient and therefore accessed more quickly than the local secondary pragmatic knowledge. However, this description seems circular.

First of all, if we consider Gibbs' proposal in terms of the interaction between Hintzman's (1986) primary memory (which records current experiences) and secondary memory (which stores largely dormant memory traces), it would seem that Gibbs' distinction does not take into account the dynamic and interactional nature of the relation between the two memory systems. Gibbs' proposal presupposes that parts of the primary pragmatic knowledge (i.e. the memory traces stored in long term memory) are/become salient irrespective of the cogniser's current (i.e. local) experience. That, however, cannot be true for the following reason.

Let us assume that the salient part of primary pragmatic knowledge relevant to the interpretation of (26) is, as proposed by Gibbs, that 'the cat chose for some reason to sit on the mat', etc. The question that needs to be asked is what is it that makes this and no other part of primary pragmatic knowledge (e.g. that fish are animals) salient. The answer is that particular parts of primary pragmatic knowledge are salient in the context of particular utterances. Thus, the salient parts of primary pragmatic knowledge predicted by Gibbs for (26) are salient in response to hearing the utterance of (26). Even if one wished to argue that it is only morpho-phonetic properties of the utterance of (26) that are relevant to the salience of parts of primary pragmatic knowledge, morpho-phonetic properties are nevertheless part of the current/local experience. This means that the salience of relevant parts of primary pragmatic knowledge *does* depend on features of the current/local experience. I suggest that Hintzman's dynamic model of the interaction between records of current experience and old memory traces is more appropriate.

Another, related problem is that if one assumes that local pragmatic knowledge is involved in the derivation of what is implicated, then one has to say that, for example, the information that the speaker of (26) is communicating a thought about the cat and the mat which are in the perceptual (visual) presence of the interlocutors is an implicature. This is because such information is derivable from the current/local context. The problem is that the analysis of so-called reference assignment in terms of implicature is intuitively incorrect (as captured by RT).

As mentioned, Gibbs' suggestion regarding the relevance of the primary versus secondary knowledge in the derivation of explicatures and implicatures is only tentative, but I suggest it needs to be rejected. Nevertheless, Gibbs' experimental work clearly indicates that there is *some* intuitively felt distinction between explicatures and implicatures and that hearers use explicatures in the derivation of implicatures.

The first reported experiment (Gibbs 2002; Hamblin & Gibbs 2003) measured the comprehension time of the same utterance which was used in two different contexts to communicate different kinds of information. The utterances in contexts are given in (27) and (28) below.

(27) Ted and Michele ran into each other at the mall.
Ted asked Michele what she had been doing lately.
Michele said that she had been busy car shopping.
Looking for ideas, Michele decided to consult Ted.
Michele asked Ted about his own car.
Ted mentioned: ''I drive a sports utility vehicle''.

(28) Ted and Michele are planning a trip to Lake Tahoe.
Michele had heard that there was a terrible storm there.
She wondered if it was going to be safe for them to go.
Michele was concerned about the vehicle they would drive.
She asked Ted if he thought they would be okay.
Ted replied: ''I drive a sports utility vehicle''.

In (27) there is no further information that Ted wants Michelle to access beyond that (a) he drives a particular kind of car. In (28), however, Ted informs Michelle that (a) he drives a particular kind of car, but he also communicates that (b) his car is safer to drive in a storm. Let us assume, after Gibbs (2002) and Hamblin & Gibbs (2003), that (a) in (27) and (28) is explicitly communicated by Ted's utterances and that (b) in (28) is implicitly communicated. The results of Hamblin & Gibbs' (2003) reaction time experiment show that deriving implicatures (measured by the comprehension time of the last line in (28)) increases the processing effort beyond that which is necessary to comprehend what speakers explicitly communicate (measured by the comprehension time of the last line in (27)). Whereas subjects on average took 1604 ms to derive the explicature, they took 1751 ms to derive the implicature. Hamblin & Gibbs (2003)

argue that these results are consistent with the assumption that people analyse what is explicated as part of determining what is implicated.[13]

The second experiment (Gibbs 2002; Hamblin & Gibbs 2003) measured the comprehension time of two different utterances which communicate the same information – one does so explicitly and the other implicitly. Consider (29) below.

(29) Bill is a new tenant in an apartment building.
His neighbor Jack has lived there for four years.
Bill was concerned that the building might be too loud.
Bill decided to ask a neighbor about it.
Bill asked Jack since he was the only neighbor Bill had met.
Jack replied,
a) ‘‘This is a very noisy building.’’
b) ‘‘I usually sleep with earplugs.’’

The assumption is that whereas (a) communicates the information that the building is very noisy explicitly, (b) does so implicitly. If that is the case, people should take more time to comprehend (b) than they do to comprehend (a). The findings were compatible with the findings of the pervious experiment: on average, subjects took 1511 ms to comprehend explicatures and 1661 ms to comprehend implicatures.

According to Gibbs (2002), the challenge for research into the explicit-implicit distinction is to specify the conditions under which hearers’ analyses of what speakers communicate demand more cognitive effort. Gibbs’ (*ibid*.) suggestion is that what facilitates comprehension is ‘conventionality’[14]. On the assumption that a given usage is conventional if it is frequent and thus widely accepted in a given speech community, it follows from Gibbs’ suggestion that if a non-literal use (e.g. metaphorical use, irony or indirect speech act use) is the conventional use of an expression which can also be used

[13] It must be pointed out that Gibbs (2002) and Hamblin & Gibbs (2003) suggest that these findings are also compatible with the assumption that people access primary pragmatic knowledge sooner than they do secondary pragmatic knowledge. Notice, however, that this suggestion is made without having defined the primary-secondary knowledge distinction properly (as discussed earlier). Because of that, the suggestion is merely about accessing primary pragmatic knowledge – whatever that is – sooner than accessing secondary pragmatic knowledge – whatever that is. Given this and other problems with the primary-secondary knowledge distinction discussed above, I reject this particular suggestion. I agree, however, that the results clearly indicate that reaction times are shorter for comprehending information (a) in (27) and longer for comprehending information (b) in (28); that for this reason we can take information (a) in (27) as explicitly communicated and information (b) in (28) an implicitly communicated; and that the findings are compatible with the suggestion that deriving an implicature is preceded by deriving an explicature.

[14] ‘Conventional’ here means ‘licensed by general use’.

in a literal way, the non-literal interpretation will be accessed faster than the literal one.[15]

This assumption receives support from experimental work. Thus, Gibbs (2002: 459-460) reports that people can comprehend an ironic use of '*You're a fine friend*' to mean that someone is a bad friend 'as quickly as, sometimes even more quickly, than literal uses of the same expression in different contexts, or equivalent non-figurative expressions'. Similarly, Gibbs (2002: 472) reports that people take less time to comprehend an indirect request communicated by *Can't you be friendly?* (i.e. the request to be friendly) than they do to comprehend its literal counterpart (i.e. the question of whether the hearer is unable to be friendly).

The question that I am interested in is whether the faster derivation of non-literal interpretations is wholly down to conventionality or whether conventionality is only partially responsible for this phenomenon. An answer to this question is suggested by Gibbs' (2002: 463, 480) observation that it is not only conventionalised non-literal interpretations which are accessed more quickly than their literal counterparts, but also novel forms (e.g. novel metaphors, ironies or sarcastic indirect requests). If that is the case, there must be something in addition to pure conventionality which facilitates their faster comprehension. Indeed, Gibbs (2002: 462) argues that people take less time to analyse ironies (e.g. *You're a fine friend*) *if the context is right*, i.e. if 'the context itself sets up an ironic situation'. Gibbs (2002: 480) goes on to suggest that 'there may not be a hard-and-fast rule that determines which kind of irony is understood as an explicature, and which as an implicature'. I suggest that this hypothesis should be extended beyond the interpretation of ironies – perhaps there are no hard-and-fast (i.e. formal) rules about what is understood as an explicature and what as an implicature; perhaps it is all context-dependent.

In the next section, I explore this hypothesis and suggest that what counts as explicitly communicated and what counts as implicitly communicated is not formally, but contextually determined. I suggest that it is plausible to approach the explicit-implicit distinction from a 'local' perspective, i.e. from the perspective of what happens between particular interlocutors in particular conversational contexts. On the 'local' account, the explicit-implicit distinction is an *intuitive* epiphenomenon of the amount of processing effort. I argue that this account is compatible with the Representational

[15] Gibbs (2002) argues that the distinction between literal (non-figurative) and non-literal (figurative) uses is questionable, if only in the light of evidence that much of what may be termed 'literal' has roots in figurative thought and language. I am concerned with the literal versus non-literal distinction insofar as it illustrates the distinction between conventional and non-conventional uses.

Hypothesis (especially, Burton-Roberts 2012) and supported by Horton & Gerrig's (2005) account of conversational common ground.

5.3.2 *Towards a local account of the explicit-implicit distinction*

As discussed in 5.1.1, Horton & Gerrig (2005) argue that particular conversational contexts, which include particular interlocutors, place immediate constraints on language production and comprehension. I argue that once it is acknowledged that interlocutor-specific information constitutes part of immediate (i.e. not merely 'post-decoding) contextual constraints on language processing, it seems plausible to approach the intuited explicit-implicit distinction from the LOCAL perspective, i.e. from the perspective of particular conversations in particular contexts between particular interlocutors.

On the local approach to the explicit-implicit distinction, the role of Horton & Gerrig's (2005) commonality assessment is essential. Consider (3) again.

(3) If you pass the exam, I'll take you to a restaurant.

In 5.1.2, I argued that, depending on the result of personal commonality assessment, (3) may give rise to (and be intended to give rise to) the weak interpretation (the over-indulgent grandma scenario) or the strong interpretation (the strict father scenario). Thus, the use of (3) to communicate that *q* iff *p* depends on the strict father's assumption that the form of (3) is sufficient to communicate that *q* iff *p* to this particular hearer – the strict father must have assumed, as a result of the personal commonality assessment, that the hearer is likely to effortlessly infer what the speaker intends to communicate. Similarly, the common ground assumptions in the over-indulgent grandma scenario suggest to the grandma that the form in (3) is not likely to give rise to the iff interpretation in the mind of this particular interlocutor (her grandson). Likewise in the grandson's mind, the common ground assumptions will suggest that the grandma is merely communicating that from the truth of *p* the truth of *q* will follow.[16] In other words, the weak and strong interpretations of (3) rely on *the assumption of tacit agreement* – which arises due to the assumption of common ground – between the speaker and the hearer as to what interpretation the use of *if* may give rise to.

Presumably, if the assumption of tacit agreement is the product of the personal commonality assessment, we are dealing with communicative (representational)

[16] As discussed in 5.1.2, the same common ground assumptions will dictate to the grandma that in order to communicate that *q* iff *p* to this particular interlocutor, she needs to put the stress on *if*.

conventions used by a small community of speakers (e.g. members of a family). If the assumption of tacit agreement is (at least partially) the product of the communal commonality assessment we are dealing with conventions used by a larger community of speakers. Let me explain this with (26), repeated below.

(26) The cat is on the mat.

I suggest that (26) can, in the right circumstances, be used and be understood as used to explicitly communicate that the speaker wants the hearer to let the cat out. The speaker A of (26) may choose this utterance to explicitly communicate that she wants the hearer B to let the cat out because the product of the personal commonality assessment dictates to her that this form is sufficient for this particular interlocutor, i.e. for B, to effortlessly infer that A wants B to let the cat out. Such personal commonality assessment must be derived on the basis of memory records of the uses of the form in (26) to communicate this particular request. It seems plausible to assume that in order for the form in (26) to explicitly communicate the relevant request, it must be initially used to implicitly communicate it. With time and frequent usage the route from the utterance of (26) to the request is short-cut – it becomes a representational convention for these particular interlocutors, A and B. The successful use of (26) to explicitly communicate the request relies on the tacit assumption that A and B share a particular convention.

Let us assume now that there is another interlocutor C for whom the commonality assessment does not yield the assumption of tacit agreement to use this particular convention – A has no memories of using (26) to communicate the relevant request to C and no memories which would make A infer that C knows that A and B share such a convention. Presumably, if A wanted to communicate to C that A wants C to let the cat out, the use of (26) would not communicate it explicitly. This is not to say that C would be unable to infer the intention to (implicitly) communicate the request. If A wanted to explicitly communicate the request to C, the communal commonality assessment would probably dictate to A to use something like '*Could you please let the cat out?*'.

The local approach to the explicit-implicit distinction seems consistent with work in the Representational Hypothesis. As discussed in chapter 2, one of the central arguments of the Representational Hypothesis (e.g. Burton-Roberts 2007, 2012) is that meaning is not a property of anything, but a cognitive relation, for someone, between X (anything) and what has semantic content Y (a thought and only a thought). Burton-Roberts (*ibid*.) argues that the RH's distinction between meaning and semantics is

strengthened by the observation that thoughts themselves can be meaningful but not simply in virtue of having semantic content. Thoughts – simply in virtue of being thoughts – are not signs (have no significance), but a thought (T1) can be *significant* ("have" meaning)[17] to a cogniser if it leads the cogniser to have another thought (T2). Burton-Roberts (*ibid*.) argues that, in a person's mental world, there are *semiotic, inferential* relations between thoughts and that this *semiotic* fact, about relations between thoughts in an individual's mental world, is separate from facts about *semantic content* of thoughts. Burton-Roberts illustrates this point with the following example. Let us imagine that Anna sees a big pile of clean washing in the kitchen. That phenomenon (P) leads Anna to have thought T1:

T1: [the washing has been done]

The pile of clean washing (P) is *a sign* (to Anna). It has no semantics but its significance lies in its relation (for Anna) to T1 (it *communicates* T1 to Anna). Now, T1 has semantic content (conventionally represented in the square brackets). But, separate from its semantic content, Anna's having T1 might lead her to have another thought, T2.

T2: [I don't need to do any washing right now] (and T2 ~~> T3, and so on)

The *relation* between T1 and T2 is the *meaning* "of" thought T1 (what T1 *communicates* to Anna). For Anna, then, [T1] *means that* [T2]. But the relation [T1→T2], and T2 itself, are clearly distinct from T1's *semantic content*.

This argument of Burton-Roberts also applies to ostensive communication. Consider (30).

(30) There's a snail in the conservatory.

In line with what I have argued for (26) (*The cat is on the mat*), I argue that the utterance of (30) may be taken as explicitly or implicitly communicating the thought that the speaker wants the hearer to get the snail out. Whether this thought is taken as being explicitly or implicitly communicated crucially depends on the result of

[17] As in chapter 2, I am using inverted commas here to emphasise that in the RH meaning is not equivalent to the encoding of semantic properties. In the RH, a sign can "have" meaning, or meaning can be said to be "of" a sign, insofar as it leads a cogniser to have a thought.

commonality assessment. I will first look at the situation in which this thought may be taken as implicitly communicated.

Let us assume that (30) is uttered by Alice to Chris and that Chris is not aware that Alice considers snails disgusting to the extent that she cannot get them out herself. As a result of the process of echo retrieval (including commonality assessment) the utterance of (30) by Alice gives rise to T1 (conventionally represented in the square brackets) in Chris's mind:

T1: [there's a snail in the conservatory]

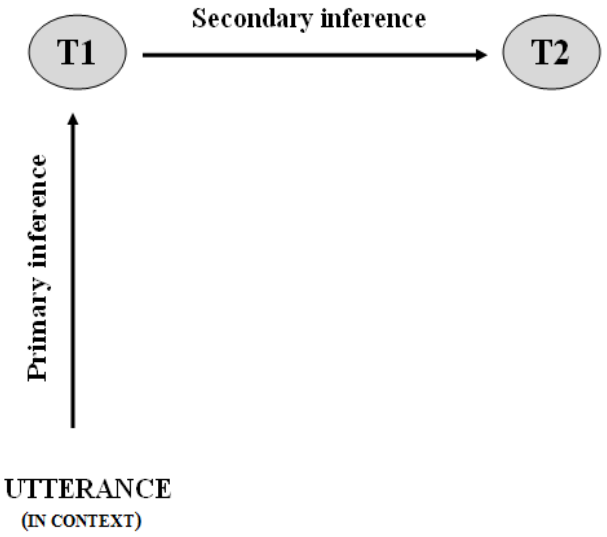
As the pile of clean washing was a sign to Anna in the earlier example, so the utterance of (30) is a sign to Chris. This sign has no semantic content (it is an acoustic phenomenon), but its significance lies in its relation (for Chris) to T1 – the utterance of (30) communicates T1 to Chris. Separate from its semantic content, Chris' having T1 might lead him (via the echo retireval process) to have another thought, T2.

T2: [Alice wants me to get the snail out] (and T2 ~~> T3, and so on)

As before, the relation between T1 and T2 is the significance of thought T1 (what T1 communicates to Chris). For Chris, [T1] means that [T2], but the relation [T1→T2], and T2 itself, are distinct from T1's semantic content.

I argue that, in ostensive communication, Burton-Roberts' T1 constitutes the explicitly communicated content, and that Burton-Roberts' T2 constitutes the implicitly communicated content. I shall refer to the pragmatic inferential process which gives rise to T1 as the PRIMARY inferential process (or inference), and to the pragmatic inferential process which gives rise to T2 as the SECONDARY inferential process (or inference). The following figure is a rough sketch of the primary and secondary inferential processes.

Figure 5.1: Primary and secondary inferential processes in ostensive communication



On the account I am proposing, T1 is the conclusion of the primary inferential process. The mechanics of this inferential process is to be cashed out in terms of Hintzman's (1986) information retrieval process (chapter 2). Accordingly, Hintzman's probe – i.e. a mental representation of a stimulus, e.g. an utterance, in a particular context – serves as a premise in an online derivation of the conclusion T1 (Hintzman's echo). T1, which is retrieved in the primary inferential process, is taken by the hearer to be explicitly communicated. Thus, I argue that the locus of explicit meaning is the (primary) semiotic relation between an utterance (in context) and a thought this utterance gives rise to (T1). Because (cognitive) context places immediate constraints on utterance interpretation, what is explicitly communicated by an uttered word is LOCAL, i.e. it varies across pairs of interlocutors and communicative situations. In the secondary inferential process, T1 serves as a premise on the basis of which the conclusion – i.e. T2 – is derived. T2, like T1, is derived via Hintzman's information retrieval process.[18] T2, which is retrieved in the secondary inferential process, is taken by the hearer to be implicitly communicated (if assumed to have been intended). Thus, I argue that the locus of implicit meaning is the (secondary) semiotic relation between (T1) and another thought (T2).

Coming back to the utterance of (26) by Alice to Chris, the thought that Alice wants Chris to get the snail out arises via the secondary inferential process as T2. However, we can imagine a second scenario in which (30) is uttered by Alice to Ben. Unlike Chris, Ben is well aware of the fact that Alice considers snails disgusting to the

[18] My proposal is that both primary and secondary inferential processes can be cashed out in terms of Hintzman's (1986) multiple-trace theory of memory – the two processes differ in terms of what constitutes the premise in an inference. This seems compatible with Gibbs (2002: 477).

extent that she cannot get them out herself. As a result of the process of echo retrieval, including commonality assessment, the utterance of (30) by Alice gives rise to T1 in Ben's mind:

T1: [Alice wants me to get a snail out of the conservatory]

In Ben's mind, T1 arises via the primary inferential process. The difference between Chris' and Ben's interpretations of (30) amounts to a difference in the processing effort required to infer a thought upon which Chris and Ben are likely to act in a way wished-for by Alice (i.e. get the snail out); Chris is likely to act upon his T2 in the same way as Ben is likely to act upon his T1.

Similarly, I argue that the utterance of (3) (*If you pass the exam, I'll take you to a restaurant*) can give rise to the iff interpretation via the primary inferential process – and thus be taken as explicitly communicated – in the strict father scenario, but not in the lenient grandma scenario. Now, my analysis of the iff interpretation as explicitly communicated (when it is explicitly communicated) differs from RT's analysis of this interpretation as an explicature in the following way. In RT, the iff interpretation arises as an explicature because it entails the purportedly encoded MI – for RT, the iff interpretation is secondary with respect to the deterministic decoding of MI (which is the necessary step in the interpretation of (3)). On my account, when the iff interpretation is explicit, it is explicit because the context allows an effortless derivation of it. For example, in the strict father scenario, the iff interpretation does not arise as a result of the strengthening of MI to equivalence, but is derived directly (i.e. without the mediation of MI) from the utterance in an adequate (cognitive) context.

Thus, the biggest contrast between RT's proposal and mine concerns the role of context in the determination of the explicit-implicit distinction. In RT, even though explicit does not equal encoded, explicit is dependent on what is encoded. As discussed in 4.2.2.2, in RT, explicature entails the purportedly encoded linguistic semantics, and any communicated assumption which is not an explicature, i.e. which does not entail the encoded semantics, is an implicature. In RT, thus, one does not need to look to context to determine what is explicitly and what is implicitly communicated; whether a proposition is explicitly – or therefore implicitly – communicated depends on whether that proposition entails what is encoded. On my account, however, whether something is explicitly or implicitly communicated is contextually determined and depends, in each particular case, on particular holistic states of mind.

I argue that the explicit-implicit distinction cannot be formalised because it is a psycho-processing distinction which is an intuitive epiphenomenon of the amount of cognitive effort undertaken (and intended to be undertaken) by the hearer in order to arrive at a given interpretation. Thus, less cognitive effort correlates with the intuition of explicitness and more cognitive effort correlates with the intuition of implicitness (as suggested by Gibbs 2002).

Now, it has already been suggested by Elsheikh (2010, 2011), who works within the Representational Hypothesis, that semiotic relations among thoughts (as proposed by Burton-Roberts) may be responsible for generating the inferences which are traditionally referred to as implicatures. Elsheikh illustrates this point with the following exchange.

(31) A: Are you bringing any drinks to the party.
B: Well, I have invited my father.

Let us assume that the cognitive context of the interlocutors taking part in (31) is such that they know that B's father is a conservative type. Elsheikh (2011) argues that, given such context, B's utterance communicates the thought that B has invited her father to the party (T1), but also that B is not bringing any drinks to the party (T2). For Elsheikh, T1 and T2 stand in a semiotic relation such that T2 is inferred from the semantic content of T1 and the context. Because T2 is derived from another thought (i.e. it is the meaning of T1 for A) and because B intended to communicate T2, Elsheikh (2011) is prepared to agree that T2 roughly corresponds to what is traditionally referred to as an implicature. She writes:

> **Quote 2:** It is this semiotic relation among thoughts which is responsible for the kind of inferences we label 'implicatures'. So, on the RH, what is standardly called 'implicature' might be construed as the meaning of a thought in a hearer's mind such [that] this meaning is intended and is intended to be recognised as intended. (Elsheikh 2011: 132-133)

The title of Elsheikh (2011) (i.e. *The Myth of Explicit Communication*) clearly indicates that, even though she is prepared to say that T2 in (31) is implicitly communicated, she is not prepared to say that T1 is explicitly communicated. Elsheikh's rejection of the notion of explicit communication is closely tied up with her rejection of this notion as defined in both Gricean and relevance theoretic traditions. I agree with Elsheikh's rejection of the traditional explicit-implicit distinction (as discussed in chapter 4); however, I also share Gibbs' view that despite the lack of prospects for its

formalisation, the intuitive distinction is still there. Hence, I do not object to the use of the terms *explicit* and *implicit* as long as it is made clear that both terms refer to interpretations that are wholly pragmatically – and locally – derived (and communicated). To repeat, Elsheikh's rejection of explicit communication is a rejection of the idea that there are any non-pragmatic aspects of communication. We agree on that. What we disagree on is that, unlike Elsheikh, I argue, in line with Gibbs (2002), that some intuitive notion of explicit communication must be acknowledged.

As discussed in 5.3.1, Gibbs (2002) suggests that conventionality facilitates comprehension. Now, were conventionality the only facilitative condition, then T2 in (31) could be taken as the result of the secondary inference from T1. This is because B's answer is not the most conventional answer to A's question.[19] However, as observed by Gibbs, novel, i.e. non-conventional, forms (e.g. novel metaphors, ironies or sarcastic indirect requests) can be accessed more quickly than their literal counterparts as long as *the context is right*. This suggests that sufficient contextual clues (both incoming and stored in long term memory) may counter-balance lack of conventionality.[20] The result is that, given the right context, an utterance which is not the most conventional may be used to explicitly communicate a thought, which would not be explicitly communicated by that utterance outside this particular context.

Accordingly, I argue that in (31) what Elsheikh refers to as T2 may be explicitly communicated – i.e. triggered in the primary inference process, by the utterance of (31) – if A and B have in their common ground (as defined by Horton & Gerrig 2005) the information that B's father is conservative and does not approve of drinking. The use of the discourse marker *well* in this particular example also supports my analysis. The use of *well* here signifies that the communicated information will be in conflict with the hearer's expectations. Clearly, in the context of A's question, the expectations concern the presence of drinks, and not the presence of B's father (put simply, A is asking about the drinks, not about B's father). If *well* signifies that what is communicated by the following *I have invited my father* is in conflict with the hearer's expectations and the hearer's expectations concern the drinks and not the father, then what is communicated by *I have invited my father* has got to be something about the drinks. Thus, the use of

[19] My assumption is that *no* would be the most conventional answer here because the use of *no* would guarantee successful communication irrespective of the interlocutors' personal common ground. It seems plausible to assume that whether an utterance is judged as conventional or not depends on the social, rather than personal, commonality assessment.

[20] As long as the relevant clues stored in long term memory are accessed within the time frame to play immediate role in utterance interpretation (as discussed earlier).

well in the co-text serves as a salient cue as to what is being communicated; together with the assumptions in the common ground, the use of *well* enables Elsheikh's T2 to arise via a primary inference process. In terms of communication efficiency, it is plausible to assume that the product of personal commonality assessment dictated to B that this particular form (i.e. B's answer in (31)) will be sufficient to communicate both that she is not bringing any drinks and a reason for it. If the product of personal commonality assessment were to yield the assumption that A does not know B's father, B would most likely choose a less succinct form in (32).

(32) No, I'm not because I have invited my father.

Given Horton & Gerrig's (2005) approach to common ground, it seems plausible to assume that there will often be misalignments between speakers and hearers as to what is explicitly communicated and what is taken to be explicitly communicated. For example, we can imagine that in scenario (31) the speaker's communicative intentions and the common ground are exactly as discussed earlier, but that due to some normal memory limitation resulting from inattention or tiredness, the relevant memory traces in the hearer's mind were not activated within the time frame to immediately constrain the interpretation of this utterance. In this situation, the thought that B's not bringing any drinks will be derived via the secondary inferential process from the thought that B has invited her father. Such misalignments are a necessary consequence of the local approach to the explicit-implicit distinction.

To summarise, I have argued that once it is acknowledged that interlocutor-specific information constitutes part of the immediate contextual constraints on language processing (production and comprehension), it seems plausible to approach the explicit-implicit distinction from the local perspective of particular conversations in particular contexts between particular interlocutors. I have further argued that the explicit-implicit distinction is an epiphenomenon of the amount of processing effort (expected to be) undertaken by the hearer in the processing of an utterance. I also contended that the distinction between what is explicitly and what is implicitly communicated parallels the distinction between the primary and secondary inferential processes.

5.4 Conclusion

In this chapter, I discussed a wholly pragmatic, inferential approach to the interpretation of conditionals. In line with Horton & Gerrig (2005), I argued that (cognitive) contextual information – including the assumption of common ground – places immediate constraints on utterance interpretation. Using this approach, I was able to explain weak and strong interpretations of conditionals in terms of cognitive contextual – i.e. holistic/individualistic – conditions. I further argued that the existence of such holistic/individualistic constraints on utterance interpretation indicates that what seems objectively messy (i.e. not explainable in terms of some necessary principles) may actually be subjectively predictable.

Finally, I discussed the implications of a wholly pragmatic approach to utterance interpretation for the explicit-implicit distinction. In line with the arguments put forward by Gibbs (2002), I argued that the intuition of the explicit-implicit distinction arises as an epiphenomenon of the amount of cognitive effort undertaken to process a given utterance.

Conclusion

This thesis set out to investigate whether it is plausible and necessary to posit the existence of specifically linguistic semantics. I argued that the notion of linguistic semantics, as well as the cognitive process of deterministic decoding of such content, is implausible and unnecessary to account for meaning in language. I then explored the consequences of this claim for a theory of meaning and utterance interpretation.

In chapter 1, I raised questions about the nature of semantic content and its relation to words. I critically engaged with Fodor's (1998, 2008) notion of externalist/referential semantics and argued that Fodor's account is untenable for several reasons. First, it presupposes internalist content in the guise of mind-dependent properties that our minds attribute to mind-external entities. Furthermore, it is these properties that compose and not, as Fodor argues, reference. Relatedly, I argued that Fodor's lexical-conceptual isomorphism incorrectly predicts one-to-one mapping between words and concepts. Indeed, I showed that referential equivalents like English *shallow* and French *peu profond* indicate that at least some 'atomic' concepts have compositional content. I argued that, in the context of the externalist-internalist debate about the nature of semantic content, the problems with Fodor's referentialism strongly tip the balance in favour of internalism.

I then endorsed Chomsky's (2000a) position on semantics, from which it follows that: (a) conceptual primitives are innately pre-determined, (b) a word's semantics is internally compositional, (c) it is context-variable and (d) it is radically individualistic. However, I argued that there is a tension between the consequences of Chomsky's internalist-individualistic assumptions and his double-interface view of language: from the former it follows that there is no linguistic semantics and from the latter it follows that there is linguistic semantics.

In chapter 2, I sought to show that the tension between the consequences of internalism-individualism, on the one hand, and the double-interface view of language, on the other, should be resolved in favour of internalism-individualism. To this aim, I first discussed some problems with the nature and acquisition of linguistic semantics in Relevance Theory. I argued that the introduction of full-fledged concepts (i.e. concepts with truth-theoretic properties) into the lexical concept repertoire leaves RT with no principled distinction between lexical and ad hoc concepts. Secondly, I argued that even if acquired, lexical concepts (a) are either unnecessary in utterance interpretation because they are post hoc or (b) cannot constitute adequate evidence for the intended

interpretation because they are derived through personal inference. Thirdly, I argued that the posited cognitive process of deterministic decoding of linguistic semantic content is redundant in cases of loose use, cases of so-called concept narrowing and where the communicated concept is the same as the assumed lexical concept. I then argued that there is no linguistic semantics and that utterance interpretation is a wholly pragmatic inferential process, which is immediately constrained by a personal, holistic inference about the communicative intention of the speaker in a given conversational context.

I defended my position by endorsing the Representational Hypothesis' (e.g. Burton-Roberts 2012) definition of meaning-as-relation and Hintzman's (1986) multiple-trace theory of memory. I argued that the RH offers a way of reconciling the non-existence of linguistic semantics with the fact that uttered words are meaningful. I further suggested that Hintzman's radically contextualist account of information retrieval invalidates the criticism that utterance interpretation without linguistic semantics is not sufficiently constrained to allow for successful communication. The general conclusion of chapter 2 was that utterance interpretation does not proceed from what is (controversially) shared – i.e. linguistic semantics – to what sufficiently converges, but from what diverges – i.e. multiple-memory traces associated with particular semiotic labels – to what sufficiently converges.

In chapter 3, I looked at philosophical legacy in thinking about linguistic semantics. I argued that all philosophical notions of shared content which have been posited as linguistic semantics (i.e. causal-externalist wide content, social-externalist wide content and non-truth-theoretic narrow content) are problematic and that holism is the only plausible thesis about mental content. I argued that, contrary to claims made by Fodor & Lepore (1992), holism: (a) *does* allow for mental generalisations (and thus for a theory of *the* human mind) – in terms of quantified laws, (b) *does* offer an accurate account of compositionality – as long as the existence of (immediate) pragmatic constraints is acknowledged; and (c) *does not* give rise to the sense-reference problem – because content externalism (á la Fodor), unlike ascriptive externalism (á la Bilgrami), is impossible to maintain. I further argued that an RH-based wholly inferential account of utterance interpretation, which combines the identity of LOT (i.e. primitive concepts and compositional capacity) among the human species with Bilgrami-type notion of public availability of communicated concepts, allows for successful communication without linguistic semantics.

In the second part of the thesis, I investigated whether my theoretical arguments apply to the analysis of the relation between the linguistic sign *if* and MI. In chapter 4, I argued that the claim that *if* semantically encodes MI cannot be maintained even when this claim is supported by pragmatic explanation. The first reason is that there are conditional belief deviations from MI. Because they arise at the level of belief, they cannot be explained in terms of conversational principles. The second reason is that the problem of pragmatic intrusion into encoded semantics, which was initially identified for Grice (1989), also arises for Relevance Theory (Carston 2002, Noh 2000); this problem is evident in RT's account of the acquisition of linguistic semantics, their claims about the nature of linguistic semantics, their re-definition of the explicit-implicit distinction and their account of the meaning of *if*.

Contrary to claims made in Carston (2002: 99-100), I argued that the purportedly encoded semantic content is not autonomous with respect to pragmatic inference – I showed that if MI applies, it can only ever do so at the level of pragmatically inferred holistic thought. Not only does it demonstrate that MI cannot be the encoded semantic content of *if*, but it also strengthens the RH's claim that thought is the only locus of semantic content and my earlier claim that such content is holistic and accessible solely through pragmatic inference.

In chapter 5, I showed how the interpretation of conditionals can be explained in terms of Horton & Gerrig's (2005) extension of a multiple-trace theory of memory into the study of conversational common ground. In particular, I argued that the distinction between the weak (i.e. modelled by MI) and the strong (i.e. modelled by equivalence) interpretations of conditionals can be explained as long as it is acknowledged that interlocutor-specific information places immediate constraints on utterance interpretation.

I distinguished between basic and extended uses of conditionals. I argued that in basic uses *if* signals a relation between two propositional objects *p* and *q*, whereas in extended uses *if* signals a relation between a proposition and an utterance (or speech act). I argued that, unlike previous approaches (e.g. Noh (2000), Sweetser (1990), Smith & Smith (1988)), the wholly pragmatic (holistic) analysis proposed in this thesis handles the variation found in the interpretation of basic and extended uses. Finally, I discussed the implications of a wholly pragmatic approach to utterance interpretation for the distinction between explicit and implicit communication. In line with Gibbs (2002), I argued that the explicit-implicit distinction is an intuitive epiphenomenon of

the amount of cognitive effort undertaken (and expected to be undertaken) to process a given utterance.

This thesis has indicated a number of areas for further research. Firstly, my investigation of conditionals was limited to indicative conditionals, as this group alone exhibits a degree of variation sufficient to illustrate and support the theoretical arguments put forward in chapters 1-3. However, no study of conditionals is complete without investigating counterfactual (or subjunctive) conditionals and their relation to indicative conditionals.

Drawing the indicative-counterfactual distinction is controversial (e.g. Dudman 1988) and even when drawn, there is no consensus as to why, on some occasions, it is possible to accept an indicative conditional but deny its counterfactual counterpart and, on other occasions, to accept both indicative conditional and its counterfactual counterpart (e.g. Edgington 2008, Jackson 1990). The former situation is illustrated by (1) and (2) below (taken from Jackson 1990). The latter is illustrated by (3) and (4) (taken from Edgington 2008).

(1) If Booth did not kill Lincoln, someone else did.
(2) If Booth had not killed Lincoln, someone else would have.
(3) If you go in, you will get hurt.
(4) If you had gone in, you would have got hurt.

The indicative conditional (1) and the counterfactual (2) are illustrative of the crucial interpretive difference between the two kinds: it is possible to accept (1) but deny (2). However, not all indicative-counterfactual pairs behave in this way. A hearer who accepts (3) and stays outside the building which is just about to collapse will also accept (4) after the building has collapsed. It is important, I believe, to investigate whether this asymmetry can be explained in terms of a wholly pragmatic account proposed in this thesis.

Secondly, it is important to investigate the relation between *and*, *or* and *not*, on the one hand, and logical functors of conjunction, disjunction and negation, on the other. For example, it transpires that even a brief reflection on *or* reveals interesting facts. Consider (5).

(5) Tomorrow I will take the 9am flight to Paris or the 9am flight to New York.

Let us assume with Grice (1989) that semantically *or* is inclusive, i.e. it is equivalent to the logical functor of disjunction, which is true if at least one of the disjuncts is true. On this assumption, the exclusive (i.e. *p* or *q* and not both) interpretation can be seen as a GCI and explained by reference to semantic scales (e.g. Levinson 2000); by uttering a lower member on the scale (*and* ⊢ *or*), the speaker is conversationally implicating that the stronger one does not obtain. As for Relevance Theory, it is assumed (e.g. Chevallier et al. 2008) that the exclusive interpretation arises as an explicature, derived on the basis of encoded inclusive disjunction and (cognitive) context of the utterance. In the case of (5), it is plausible to assume that the relevant cognitive context would concern the assumption that a single person cannot board two planes at the same time. Both Grice/Levinson and RT have a way of explaining why, even though *or* encodes inclusive disjunction, the utterance of it in (5) communicates the more specific *p* or *q* and not both.

However, it seems that, as was the case with *if*, so with *or*, there is a problem with belief deviations from the logic of inclusive disjunction. A person C can rationally but inconsistently with the logic of inclusive disjunction believe that the next day C will take the 9am flight to Paris or the 9am flight to New York. This exclusive disjunction belief deviation from the logic of inclusive disjunction cannot be explained in conversational terms.

Thirdly, the local account of the explicit-implicit distinction which I pointed towards in this thesis needs a more in-depth investigation. As a particular area of interest I identify research into the extent of correlation between (a) the intuition of explicit communication (E) and a high degree of reliance on communal commonality assessment (CCA); and (b) the intuition of implicit communication (I) and a high degree of reliance on personal commonality assessment (PCA). If the arguments put forward in Gibbs (2005), Hamblin & Gibbs (2003) and in section 5.3 are on the right track, in real life discourses we should often be able to find lack of correlation between E and high degree of CCA and between I and high degree of PCA.

Relevance Theory has shown that the role of pragmatic inference in utterance interpretation is much greater than Grice had envisaged. I hope to have shown in this thesis that the role of pragmatic inference in utterance interpretation is even greater than envisaged by Relevance Theory and that this fact is not as problematic as argued by sceptics.

Bibliography

Adams, E.W. (1965). A logic of conditionals. *Inquiry* 8. 166-97.

Austin, J.L. (1975). *How To Do Things With Words*, second edition. Oxford: Oxford University Press.

Bach, K. (1996). Content: wide vs. narrow. In Craig, E. (ed.), *The Routledge Encyclopedia of Philosophy*. Routledge. Available from http://online.sfsu.edu/kbach/widenarr.html. Accessed on 21/01/2011.

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences* 4(11). 417-423.

Barsalou, L.W. (1982). Context-independent and context-dependent information in concepts. *Memory and Cognition* 10 (1). 82-93.

---------- (1983). Ad hoc categories. *Memory and Cognition* 11. 211-227.

---------- (2005). Abstraction as dynamic interpretation in perceptual symbol systems. In Gershkoff-Stowe, L. & Rakison, D. (eds.), *Building Object Categories*. Carnegie Symposium Series. Majwah, NJ: Lawrence Erlbaum. 389-431.

---------- (2012). *Grounding knowledge in the brain's modal systems*. Paper delivered at 4th UK Cognitive Linguistics Conference at King's College London, 10-12th July 2012.

Beller, S. (2002). Conditional promises and threats – cognition and emotion. In Gray, W.D. & Schunn, C.D. (eds.), *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum. 113-118.

Bilgrami, A. (1992). *Belief and Meaning*. Cambridge, Massachusetts: Basil Blackwell Ltd.

---------- (1998a). Précis of Belief and Meaning. *Philosophy and Phenomenological Research* LVIII (3). 595-605.

---------- (1998b). Why holism is harmless and necessary. *Noûs* 32 [*Philosophical Perspectives* 12: Language, Mind, and Ontology]. 105-126.

---------- (2002). Chomsky and philosophy. *Mind and Language* 17 (3). 290-302.

Blakemore, D. (1992). *Understanding Utterances: An Introduction to Pragmatics.* Oxford: Blackwell.

Block, N. (1986). Advertisement for a semantics for psychology. In French, P. A. et al. (eds.), *Midwest Studies in Philosophy 10: Studies in the Philosophy of Mind*. Minneapolis: University of Minnesota Press. 615-678.

---------- (1993). Holism, hyper-analyticity and hyper-compositionality. *Mind and Language* 8 (1). 1-26.

---------- (1995). An argument for holism. *Proceedings of the Aristotelian Society, New Series* 95. 151-169.

---------- (1998a). Holism, mental and semantic. In Craig, E. (ed.), *The Routledge Enclyclopaedia of Philosophy*. Routledge. Available from http://www.nyu.edu/gsas/dept/philo/faculty/block/papers/MentalSemanticHolism.html. Accessed on 21/01/2011.

---------- (1998b). Conceptual role semantics. In Craig, E. (ed.), *The Routledge Enclyclopaedia of Philosophy*. Routledge. Available from http://www.nyu.edu/gsas/dept/philo/faculty/block/papers/ConceptualRoleSemantics.html. Accessed on 21/01/2011.

Brown, C. (2011). Narrow mental content. In Zalta, E. N. (ed.), *The Stanford Encyclopedia of Philosophy*. Available from http://plato.stanford.edu/entries/content-narrow. Accessed on 21/01/2011.

Burge, T. (1979). Individualism and the mental. In French, P. et al. (eds.), *Midwest Studies in Philosophy*, vol. 4. Minneapolis: University of Minnesota Press. [Reprinted in Burge, T. (2007). *Foundations of mind*. Oxford: Clarendon Press. 100-150]

Burton-Roberts, N. (2000). Where and what is phonology? A representational perspective. In Burton-Roberts, N., Carr, P. & Docherty, G., (eds.) *Phonological Knowledge: Conceptual and Empirical Issues*. Oxford: Oxford University Press. 39-66.

---------- (2005). Robyn Carston on semantics, pragmatics and 'encoding'. *Journal of Linguistics* 41. 389-407.

---------- (2007). Varieties of semantics and encoding: negation, narrowing/loosening and numericals. In Burton-Roberts, N. (ed.), *Pragmatics*. Basingstoke: Palgrave-Macmillan. 90-114.

---------- (2009). The Grounding of Syntax - and More. *Newcastle Working Papers in Linguistics* (15). 21-39.

---------- (2011). On the grounding of syntax and the role of phonology in human cognition. *Lingua* 121 (14). 2089-2102.

---------- (2012). Meaning, semantics and semiotics. In Capone, A. (ed.), *Perspectives on Pragmatics and Philosophy*. Springer.

---------- (manuscript). *Natural language and its conventional representation*.

Burton-Roberts, N. & Carr, P. (1999). On speech and natural language. *Language Sciences* 21. 371-406.

Burton-Roberts, N. & Poole, G. (2006a). Syntax vs. phonology: a representational approach to stylistic fronting and verb-second in Icelandic. *Lingua* 116. 562-600.

---------- (2006b). "Virtual conceptual necessity", feature-dissociation and the Saussurian legacy in generative grammar. *Journal of Linguistics* 42. 575–628.

Carr, P. (2000). Scientific realism, sociophonetic variation, and innate endowments in phonology. In Burton-Roberts, N., Carr, P. & Docherty, G. (eds.), *Phonological Knowledge: Conceptual and Empirical Issues*. Oxford: Oxford University Press. 67-104.

Carston, R. (1988). Implicature, explicature, and truth-theoretic semantics. In Kempson, R. (ed.), *Mental Representations*. Cambridge: Cambridge University Press. 155-181. [Reprinted in Davis, S. (ed.), *Pragmatics: A Reader*. New York: Oxford University Press. 33-51.]

---------- (1996). Enrichment and loosening: complementary processes in deriving the proposition expressed? *UCL Working Papers in Linguistics* 8. 61-88.

---------- (1998). The semantics/pragmatics distinction: a view from relevance theory. *UCL Working Papers in Linguistics* 10. 1-30.

---------- (2002). *Thoughts and Utterances: The Pragmatics of Explicit Communication.* Oxford: Blackwell.

---------- (2010). Explicit communication and 'free' pragmatic enrichment. In Soria, B. & Romero, E. (eds.), *Explicit Communication: Robyn Carston's Pragmatics*. Basingstoke: Palgrave Macmillan. 217-287.

Chapman, S. (2005). *Paul Grice, Philosopher and Linguist*. Basingstoke: Palgrave Macmillan.

Chevallier, C., Noveck, I., Bott, L., Lanzetti, V., Nazir T. & Sperber, D. (2008). Making disjunctions exclusive. *Quarterly Journal of Experimental Psychology* 61(11). 1741-1760.

Chng, S. (1999). *Language, Thought and Literal Meaning*. PhD thesis, University of Newcastle.

Chomsky, N. (1975). *Reflections on Language*. New York: Pantheon.

---------- (1981). *Lectures on Government and Binding*. Dordrecht: Foris.

---------- (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.

---------- (2000a). *New Horizons in the Study of Language and Mind*. Cambridge: Cambridge University Press.

---------- (2000b). *The Architecture of Language*. New Delhi: Oxford University Press.

---------- (2002). *On Nature and Language*. Cambridge: Cambridge University Press.

---------- (2003). Reply to Horwich. In Antony, L.M. & Hornstein, N. (eds.), *Chomsky and His Critics.* Malden, MA: Blackwell. 295-304.

---------- (2004). Beyond explanatory adequacy. In Belletti, A. (ed.), *Structures and Beyond: The Cartography of Syntactic Structure*, vol. 3. Oxford: Oxford University Press. 104-131.

---------- (2005a). Three factors in language design. *Linguistic Inquiry* 36 (1). 1–22.

---------- (2005b). On phases. (manuscript). [Published (2008) in Freidin, R., Otero, C. P. & Zubizarreta, M. L. (eds.), *Foundational Issues in Linguistic Theory*. Cambridge, MA: MIT Press. 133-166.]

---------- (2007). Approaching UG from below. In Sauerland, U. & Gärtner, H-M. (eds.), *Interfaces + Recursion = Language?: Chomsky's Minimalism and the View from Syntax-Semantics*. Berlin: Mouton de Gruyter. 1-29.

Clark, H.H. (1994). Discourse in production. In Gernsbacher, M.A. (ed.), *Handbook of Psycholinguistics*. San Diego: Academic Press. 985-1021.

Comrie, B. (1986). Conditionals: a typology. In Traugott, E. et al. (eds.), *On Conditionals*. Cambridge: Cambridge University Press. 77-99.

Dancygier, B. & Sweetser, E. (1997). *Then* in conditional constructions. *Cognitive Linguistics* 8(2). 109-136.

Dąbrowska, E. (2004). *Language, Mind and Brain: Some Psychological and Neurological Constraints on Theories of Grammar*. Edinburgh: Edinburgh University Press.

Dudman, V.H. (1988). Indicative and subjunctive. *Analysis* 48. 113-22.

Edgington, D. (2008). Conditionals. In Zalta, E. N. (ed.), *The Stanford Encyclopedia of Philosophy*. Available from http://plato.stanford.edu/entries/conditionals. Accessed on 25/09/2012.

Elsheikh, E. (2010). Meaning versus semantics: a representational perspective. *Newcastle Working Papers in Linguistics* 16. 44-55.

---------- (2011). *The Myth of Explicit Communication: A View from the Representational Hypothesis*. Lambert Academic Publishing.

Fodor, J. (1970). Three reasons for not deriving "Kill" from "Cause to die". *Linguistic Inquiry* 1 (4). 429-438.

---------- (1980). Methodological solipsism considered as a research strategy in cognitive psychology. *The Behavioural and Brain Sciences* 3. 63-109.

---------- (1998). *Concepts: where cognitive science went wrong*. Oxford: Oxford University Press.

---------- (2001). Language, thought and compositionality. *Mind and Language* (16). 1-15.

---------- (2003). Is it a bird? Problems with old and new approaches to the theory of concepts. *Times Literary Supplement*, 17 January 2003. 3-4.

---------- (2007). Semantics – an interview with Jerry Fodor. *Revista Virtual de Estudos da Linguagem* – ReVEL 5 (8). 1-13.

---------- (2008). *LOT2: Language of Thought Revisited*. Oxford: Oxford University Press.

Fodor, J. & Lepore, E. (1992). *Holism: A Shopper's Guide*. Oxford, Cambridge: Blackwell Publishers.

Frege, G. (1960). On sense and reference. In Geach, P. & Black, M. (eds.), *Translations from the Philosophical Writings of Gottlob Frege*. Oxford: Basil Blackwell. 36-56.

Geiss, M.L. & Zwiky, A.M. (1971). On invited inferences. *Linguistic Inquiry* 2 (4). 561-566.

Gibbs, R. (2002). A new look at literal meaning in understanding what is said and implicated. *Journal of Pragmatics* 34. 457-486.

Grice, H.P. (1989). *Studies in the Way of Words*. Cambridge, Mass: Harvard University Press.

Groefsema, M. (2007): Concepts and word meaning in relevance theory. In Burton-Roberts, N., (ed.) *Pragmatics*. Basingstoke: Palgrave-Macmillan. 136-157.

Hamblin, J.L. & Gibbs, R. (2003). Processing the meanings of what speakers say and implicate. *Discourse Processes* 35 (1). 59-80.

Hintzman, D. (1984). Episodic versus semantic memory: A distinction whose time has come – and gone? *Behavioral and Brain Sciences* 7(2). 240-241.

---------- (1986). "Schema abstraction" in a multiple trace memory model. *Psychological Review* 93. 411-28.

---------- (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review* 95. 528-551.

---------- (2008). Memory from the outside, memory from the inside. In Gluck, M.A., Anderson, J.R. & Kosslyn, S.M. (eds.), *Memory and Mind*. New York: Lawrence Erlbaum. 15-30.

Horn, L. (1989). *Natural History of Negation*. Cambridge: Cambridge University Press.

---------- (2000). From *if* to *iff*: conditional perfection as pragmatic strengthening. *Journal of Pragmatics* 32 (3). 289-326.

Horton, W.S. (2008). A memory-based approach to common ground and audience design. In Kecsekes, I. & Mey, J. (eds.), *Intention, Common Ground and the Egocentric Speaker-Hearer*. Berlin, New York: Mouton de Gruyter. 189-222.

Horton, W.S. & Gerrig, R.J. (2005). Conversational common ground and memory processes in language production. *Discourse Processes* 40 (1). 1-35.

Hussein, M. (2008). The truth-conditional/non-truth-conditional and the conceptual/ procedural distinctions revisited. *Newcastle Working Papers in Linguistics* 14. 61-80.

---------- (2009). *Relevance Theory and Procedural Meaning: The Semantics and Pragmatics of Discourse Markers in English and Arabic*. PhD Thesis, Newcastle University.

Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.

Jackson, F. (1990). Classifying conditionals I. *Analysis* 50. 134-47.

Kjøll, G. (2009). *Where 'meaning' comes from – internalism and externalism about semantic content*. Paper delivered at the CSMN-CASTL workshop on semantics at University of Oslo, 6th October 2009.

Lahav, R. (1989). Against compositionality: the case of adjectives. *Philosophical Studies* 57. 261-279.

Lalumera, E. (2009). More than words. In De Prabanter, P. & Kissinem, M. (eds.), *Utterance Interpretation and Cognitive Models*. Bingley: Emerald Publishers. 75-99.

Laurence, S. & Margolis, E. (1999). Concepts and cognitive science. In Laurence, S. & Margolis, E. (eds.), *Concepts: Core Readings*. Cambridge, MA: MIT Press. 3-81.

Lepore, E. (1999). Semantic holism. In Auri, R. (ed.) *The Cambridge Dictionary of Philosophy*, second edition. Cambridge: Cambridge University Press.

Levinson, S. (1983). *Pragmatics*. Cambridge: Cambridge University Press.

---------- (2000). *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. Cambridge, MA: MIT Press.

Ludlow, P. (2003). Referential semantics for I-languages? In Antony, L.M. & Hornstein, N. (eds.), *Chomsky and His Critics*. Malden, MA: Blackwell. 140-161.

MacFarlane, J. (2010). What is assertion? In Cappelen, H. & Brown, J. (eds.), *Assertion - New Philosophical Essays*. Oxford and New York: Oxford University Press. 79–96.

Mauri, C. & van der Auwera, J. (2012). Connectives. In Allan, K. & Jaszczolt, K.M. (eds.), *The Cambridge Handbook of Pragmatics*. Cambridge: Cambridge University Press. 377-401.

Noh, E-J. (1996). A relevance-theoretic account of metarepresentative uses in conditionals. *UCL Working Papers in Linguistics* 8.

---------- (2000). *Metarepresentation: A Relevance-Theory Approach***.** Amsterdam: John Benjamins Publishing Co.

Pagin, P. (2006). Meaning holism. In Lepore, E. & Smith, B. (eds.), *Handbook of Philosophy of Language*. Oxford: Oxford University Press. 213-232.

Pateman, T. (1987). *Language in Mind and Language in Society*. Oxford: Oxford University Press.

Perry, J. (1977). Frege on demonstratives. *Philosophical Review* 86. 474-497.

Putnam, H. (1975). The meaning of 'meaning'. In *Philosophical Papers, vol. 2: Mind Language and Reality*. Cambridge: Cambridge University Press. 215-271.

Quine, W.V.O. (1951). Two dogmas of empiricism. *Philosophical Review* 60. 20-43. [Reprinted in W.V.O. Quine (1961), *From a Logical Point of View*. Harvard: Harvard University Press, second edition.]

Recanati, F. (1991). The pragmatics of what is said. In Davis, S. (ed.) *Pragmatics: A Reader*. New York: Oxford University Press. 97-120.

---------- (1993). *Direct Reference: From Language to Thought*. Oxford: Blackwell Publishers Ltd.

---------- (1998). Pragmatics. *Routledge Encyclopedia of Philosophy*, vol. 7. London: Routledge. 620-633.

---------- (2005). Literalism and Contextualism: Some Varieties. In Preyer, G. & Peter, G. (eds.) *Contextualism in Philosophy*. Oxford: Oxford University Press.

Rey, G. (2010). The analytic/synthetic distinction. In Zalta, E.N. (ed.), *The Stanford Encyclopedia of Philosophy*. Available from http://plato.stanford.edu/entries/analytic-synthetic. Accessed on 21/01/2011.

Searle, J. (1978). Literal meaning. *Erkenntnis* 13. 207–224.

Smith, E. & Medin, D. (1999). The exemplar view. In Margolis, E. & Laurence, S. (eds.), *Concepts: Core Readings*. Cambridge, MA: MIT Press.

Smith, N. and Smith, A. (1988). A Relevance-theoretic Account of Conditionals. In Hyman, L. & Charles, N. (eds.), *Language, Speech and Mind: Studies in Honour of Victoria A. Fromkin.* London, New York: Routledge. 322 -352.

Sperber, D. & Wilson, D. (1987). Précis of Relevance: Communication and Cognition. *Behavioural and Brain Sciences* 10. 697-754.

---------- (1995). *Relevance: Communication and Cognition*, second edition. Oxford: Blackwell Publishers Ltd.

---------- (1998). The mapping between the mental and the public lexicon. In Carruthers, P. & Boucher, J. (eds.), *Thought and Language*. Cambridge: Cambridge University Press.

Strawson, P. F. (1986). If and '⊃'. In Grandy, R.E. & Warner, R. (eds.), *Philosophical Grounds of Rationality: Intentions, Categories, Ends*. Oxford: Oxford University Press. 229-242.

Sweetser, E. (1990). *From Etymology to Pragmatics: Metaphorical and Cultural Aspects of Semantic Structure*. Cambridge: Cambridge University Press.

Swiatek, K. (2012). The solipsist mind and reality. In Frath, P. et al. (eds.), *Res-per-nomen III: La Référence, la Conscience et le Sujet Énonciateur.* Reims: Éditions et Presses Universitaires de Reims. 469-484.

Tulving, E. (2002). Episodic memory: from mind to brain. *Annual Review of Psychology* 53. 1-25.

Urquiza, C. (2011). Lexical pragmatics and memory traces. *UCL Working Papers in Linguistics* 23. 47-68.

Van der Auwera, J. (1985). *Language and Logic. A Speculative and Condition-Theoretic Study*. Amsterdam: Benjamins.

---------- (1986). Conditionals and speech acts. In Traugott, E. et al. (eds.), *On Conditionals*. Cambridge: Cambridge University Press. 197-214.

---------- (1997). Pragmatics in the last quarter century: The case of conditional perfection. *Journal of Pragmatics* 27. 261-274.

Wedgwood, D. (2007) Shared assumptions: semantic minimalism and RT. *Journal of Linguistics* 43. 647-681.

Wikforss, A.M. (2004). Externalism and incomplete understanding. *Philosophical Quarterly* 54. 287-294.

Wilson, D. & Sperber, D. (1993). Linguistic form and relevance. *Lingua* 90. 1-25.

---------- (2004). Relevance Theory. In Ward, G. & Horn, L.R. (eds.), *Handbook of Pragmatics*. Oxford: Blackwell. 607-632.

Young, D.G. (2005). Encoding and linguistic semantics. *Newcastle Working Papers in Linguistics* 11.

---------- (2006). *The Problem of Semantic Underdeterminacy: A Representational Approach.* PhD thesis. University of Newcastle.

URL

The British National Corpus [Online], version 3 (BNC XML Edition), 2007. Distributed by Oxford University Computing Services on behalf of the BNC Consortium. Available from: http://www.natcorp.ox.ac.uk/. Accessed on 15/12/2011.

Cambridge Dictionaries Online. Cambridge University Press. Available from: http://dictionary.cambridge.org/. Accessed on 02/08/2011.