

**Motion Seen and Understood: Interactions between
Language Comprehension and Visual Perception.**

Lotte Meteyard

PhD Thesis

University College London

UMI Number: U591634

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI U591634

Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.



ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Declaration

This dissertation is the result of original work by the author, where material has been drawn from other sources it has been appropriately acknowledged.

This dissertation is not the same as any that I have submitted for a degree or diploma or other qualification at any other university.

No part of this dissertation has already been or is being concurrently submitted for any such degree, diploma or other qualification.

*For my family and my mother,
as she lives on through us.
Carpe diem.*

Thanks

First I must thank my supervisor, Professor Gabriella Vigliocco. She has mentored me from a nervous undergraduate to a more self-assured scientist. I would not be here without her encouragement and conviction. Thanks also to my secondary supervisor, Professor Bencie Woll, for enlightening conversation and discussion. Special thanks to David Vinson for being a brilliant teacher and colleague and to all the members of the lab for making work worthwhile. Considerable thanks must go to Bahador Bahrami for superior intellectual stimulation and a productive collaboration. I am also indebted to the numerous undergraduate students who helped me collect data. Thank you to all my friends who have kept me company and kept me sane. Finally, I am grateful to Woo for supporting me and making me happy.

Abstract

Embodied theories of cognition state that the body plays a central role in cognitive representation. Under this description semantic representations, which constitute the meaning of words and sentences, are simulations of real experience that directly engage sensory and motor systems. This predicts interactions between comprehension and perception at low levels, since both engage the same systems, but the majority of evidence comes from picture judgements or visuo-spatial attention; therefore it is not clear which visual processes are implicated. In addition, most of the work has concentrated on sentences rather than single words; although theories predict that the semantics of both should be grounded in simulation.

This investigation sought to systematically explore these interactions, using verbs that refer to upwards or downwards motion and sentences derived from the same set of verbs. As well as looking at visuo-spatial attention, we employed tasks routinely used in visual psychophysics that access low levels of motion processing. In this way we were able to separate different levels of visual processing and explore whether interactions between comprehension and perception were present when low level visual processes were assessed or manipulated. The results from this investigation show that: (1) There are bilateral interactions between low level visual processes and semantic content (lexical and sentential). (2) Interactions are automatic, arising whenever linguistic and visual stimuli are presented in close temporal contiguity. (3) Interactions are subject to processes within the visual system such as perceptual learning and suppression. (4) The precise content of semantic representations dictates which visual processes are implicated in interactions.

The data is best explained by a close connection between semantic representation and perceptual systems; when information from both is available it is automatically integrated. However, it does not support the direct and unmediated commitment of the visual system in the semantic representation of motion events. The results suggest a complex relationship between semantic representation and sensory-motor systems that can be explained by combining task specific processes with either strong or weak embodiment.

Table of Contents

1.	SEMANTIC REPRESENTATION	10
1.1.	What is semantic information	10
1.1.1.	Semantic and conceptual knowledge	10
1.2.	Symbolic and Embodied Representation	15
1.2.1.	Computation and Symbols	16
1.2.2.	Perception and Conception	20
1.2.3.	Symbolic versus Embodied Representation.	24
1.3.	Theories of Semantic Representation	25
1.3.1.	Early theories: 1968-1995	26
1.3.1.1.	Early Network Models (Quillian, 1968 and Collins & Loftus, 1975)	26
1.3.1.2.	Attributes of Meaning (Smith, Shoben & Rips, 1974)	30
1.3.1.3.	From Intention to Articulation (Levett, 1989)	32
1.3.1.4.	Parallel Distributed Processing (Rumelhart, McClelland & the PDP Research Group 1987)	33
1.3.1.5.	Multimodality and Category Specificity (Farah & McClelland, 1991)	34
1.3.2.	Summary of Early Theories	36
1.3.3.	1995 – Present	37
1.3.3.1.	Latent Semantic Analysis (Landauer & Dumais, 1997)	37
1.3.3.2.	Conceptualist Semantics (Jackendoff, 2002)	39
1.3.3.3.	Recent Featural theories	41
1.3.3.3.1.	Featural Representations in Word Meaning (McRae, de Sa, Seidenberg, 1997)	41
1.3.3.3.2.	Distributed Conceptual Knowledge (Tyler & Moss, 2001)	42
1.3.3.3.3.	FUSS (Vigliocco, Vinson, Lewis & Garrett, 2004)	44
1.3.3.3.4.	Semantic representations as mappings (Rogers, Lambon-Ralph, Garrard, Bozeat, McClelland, Hodges & Patterson, 2004)	46
1.3.3.4.	Summary of featural theories	48
1.3.3.5.	Embodied, simulation accounts	49
1.3.3.5.1.	The Brain's Concepts (Gallese & Lakoff, 2005)	49
1.3.3.5.2.	Word Webs (Pulvermüller 1999, 2001)	51
1.3.3.5.3.	Perceptual Symbol Systems (Barsalou, 1999; Barsalou et al, 2003)	53
1.3.3.5.4.	The Immersed Experiencer Framework (Zwaan, 2004)	55
1.3.3.5.5.	The Indexical Hypothesis (Glenberg & Robertson, 2000; Glenberg & Kaschak, 2003)	58
1.3.3.5.6.	Summary of Embodied theories	59
1.4.	Key issues in representation	61
1.4.1.	Decomposition	61
1.4.2.	Structure	63
1.4.3.	Content	65
2.	EVIDENCE FOR EMBODIED THEORIES	71
2.1.	Behavioural Evidence	73
2.1.1.	Behavioural evidence with single words and objects	74
2.1.1.1.	The motor domain	75
2.1.1.2.	Visuo-Spatial Information	76
2.1.1.3.	Property Verification	79
2.1.1.4.	Summary of behavioural evidence with single words	82
2.1.2.	Behavioural evidence with sentences	83
2.1.2.1.	The Motor Domain	83
2.1.2.2.	Visuo-Spatial Information	87
2.1.2.3.	Picture judgements	92
2.1.2.4.	Narrative	95
2.1.2.5.	Summary of behavioural evidence with sentences/narrative	98
2.1.3.	Eye-Movements	99
2.1.4.	Summary of behavioural evidence	104
2.2.	Neuroscientific evidence	107
2.2.1.	EEG	108
2.2.2.	TMS	109
2.2.3.	PET/fMRI	112
2.2.4.	Summary of Neuroscientific Evidence	119

2.3.	Conclusions	124	
3.	STRUCTURE OF THE INVESTIGATION		127
3.1.	Strong versus Weak embodiment	127	
3.2.	Questions to be addressed	131	
3.2.1.	Question 1: Are embodied effects present for single words and sentences?		132
3.2.2.	Question 2: Are low level visual processes implicated in interactions between comprehension and perception?	133	
3.2.3.	Question 3: Are visuo-spatial attention and motion perception affected in the same way by comprehension?	133	
3.2.4.	Question 4: For visual motion, are comparable effects found for comprehension on perception and perception on comprehension?	134	
3.3.	Visual processes to be manipulated	135	
3.3.1.	Visuo-spatial Attention	135	
3.3.2.	Motion Perception	136	
3.4.	Experimental Design	136	
3.4.1.	Statistical Conventions	137	
4.	ITEM NORMING, SELECTION AND GENERATION		138
4.1.	Rebus Norming	138	
4.1.1.	Participants	138	
4.1.2.	Method	138	
4.2.	Item Selection	142	
4.2.1.	Item Set One	143	
4.2.2.	Item Set Two	147	
4.3.	Single Word Norming	156	
4.4.	Item Preparation & Filler Items	158	
4.4.1.	Soundfile preparation	158	
4.4.2.	Filler Words and Sentences	159	
5.	COMPREHENSION ON VISUO-SPATIAL ATTENTION		161
5.1.	Blocked Word Comprehension & Shape Categorisation	161	
5.1.1.	Participants	162	
5.1.1.	Method	162	
5.1.2.	Results	165	
5.1.3.	Discussion	166	
5.2.	Single Word Comprehension & Shape Categorisation	167	
5.2.1.	Participants	167	
5.2.2.	Method	167	
5.2.3.	Results	169	
5.2.4.	Discussion	171	
5.3.	Full Sentence Comprehension & Shape Categorisation	172	
5.3.1.	Participants	172	
5.3.2.	Method	172	
5.3.3.	Results	174	
5.3.4.	Discussion	175	
5.4.	Partial Sentence Comprehension & Shape Categorisation	177	
5.4.1.	Participants	177	
5.4.2.	Method	177	
5.4.3.	Results	179	
5.4.4.	Discussion	180	
5.5.	General Discussion	182	
6.	COMPREHENSION ON VISUAL MOTION		187
6.1.	Pilot Experiments	188	
6.1.1.	Participants	188	
6.1.2.	Method	189	
6.1.3.	Results	191	
6.1.4.	Discussion of Pilot Results	193	
6.1.5.	Motion Detection	195	
6.2.	Blocked Word Comprehension & Motion Detection	196	
6.2.1.	Participants	196	
6.2.2.	Method	197	

6.2.3.	Results	200	
6.2.4.	Discussion	202	
6.3.	Single Word Comprehension & Motion Detection	205	
6.3.1.	Participants	205	
6.3.2.	Method	205	
6.3.3.	Results	208	
6.3.4.	Discussion	210	
6.4.	Sentence Comprehension & Motion Detection	212	
6.4.1.	Participants	212	
6.4.2.	Method	212	
6.4.3.	Results	214	
6.4.4.	Discussion	216	
6.5.	Blocked Single Word Comprehension Synchronised to Motion Detection	219	
6.5.1.	Participants	220	
6.5.2.	Method	220	
6.5.3.	Results	222	
6.5.4.	Discussion	229	
6.6.	General Discussion	232	
6.6.1.	Congruency	233	
6.6.2.	Motion Direction	238	
7.	TRANSCRANIAL MAGNETIC STIMULATION	240	
7.1.	TMS during Motion Detection and Passive Comprehension of Blocked Single Words	241	
7.1.1.	Participants	241	
7.1.2.	Method	242	
7.1.3.	Results	245	
7.1.4.	Discussion	250	
8.	PERCEPTION OF VISUAL MOTION ON COMPREHENSION	257	
8.1.	Trial by Trial perception of motion on single word and sentence comprehension	258	
8.1.1.	Participants	258	
8.1.2.	Method	259	
8.1.3.	Results	261	
8.1.4.	Discussion	263	
8.2.	Self-paced reading with motion perception	266	
8.2.1.	Participants	267	
8.2.2.	Method	267	
8.2.3.	Results	269	
8.2.4.	Discussion	272	
8.3.	Visual Lexical Decision with Salient and Ambiguous motion	274	
8.3.1.	Participants	275	
8.3.2.	Method	275	
8.3.3.	Results	278	
8.3.4.	Discussion	283	
8.4.	General Discussion	285	
9.	DISCUSSION AND CONCLUSION	290	
9.1.	Question 1: Are embodied effects present for single words and sentences?	291	
9.1.1.	Visuo-spatial attention	291	
9.1.2.	Comprehension on Motion Perception	294	
9.1.3.	Motion perception on Comprehension	297	
9.1.4.	Summary and Conclusions	298	
9.2.	Question 2: Are low-level visual processes implicated in interactions between comprehension and perception?	300	
9.2.1.	Summary and Conclusions	305	
9.3.	Question 3: Are visuo-spatial attention and motion perception affected in the same way by comprehension?	306	
9.3.1.	Summary and Conclusions	312	
9.4.	Question 4: For visual motion, are congruency effects in the same conditions for comprehension on perception and perception on comprehension?	313	

9.4.1. Summary and Conclusions	315
9.5. Further Discussion	316
9.5.1. How do the present results compare with the available literature?	316
9.5.2. What conclusions can be drawn about the relationship between semantic and visual motion, what implications does this have for theories of semantic representation as a whole and embodied theories in particular?	326
9.5.3. Concluding Statement	329
Appendix 1a – g	330
Appendix 2	338
Appendix 3	343
References	353

Table of Figures

FIGURE 1-1: Schematic of Quillian's Network Model	27
FIGURE 1-2: Schematic of Collins & Loftus Spreading-Activation Theory	28
FIGURE 1-3: Multidimensional Scaling Solutions from Smith, Shoben & Rips (1974)	31
FIGURE 1-4: Representation of the Semantic Relation Between items in FUSS	45
FIGURE 1-5: Schematic of where theories lie along the continuum from amodal to modal	66
FIGURE 2-1: Example of visual stimuli used in Matlock & Richardson	100
FIGURE 2-2: Example lay out from Chambers et al (2002), Experiment 1	101
FIGURE 2-3: Example layout for Chambers et al (2002), Experiment 2	102
FIGURE 3-1: A schematic of direct engagement	127
FIGURE 3-2: A schematic of indirect engagement	129
FIGURE 3-3: Schematic of where theories lie along the continuum from amodal to modal	130
FIGURE 4-1: Example of a rebus sentence	138
FIGURE 4-2: Rebus and single word normed scores for each direction by Item Group	157
FIGURE 5-1: Reaction times for Shape Categorisation by Word Category collapsed across Position	64
FIGURE 5-2: Mean reaction times for (Full) Shape Categorisation by Sentence Category and Position	75
FIGURE 5-3: Mean reaction times for (Partial) Sentence Category	80
FIGURE 6-1: Mean reaction times for Pilot A for each Word Category	192
FIGURE 6-2: Mean d' (a) and c (b) by Word Category and Motion Direction.	201
FIGURE 6-3: D' by Block Section (a) and Trial Number (b)	224
FIGURE 6-4: Beta by Block Section (a) and Trial Number (b)	225
FIGURE 6-5: Reaction time by Block Section (a) and Trial Number (b)	226
FIGURE 7-1: D' for Session 1 (a) and Session 2 (b) by TMS/NO TMS	247
FIGURE 7-2: Inverted U graph from a modal of Stochastic Resonance (from Moss et al, 2004)	252
FIGURE 7-3: An Illustration of Low (a), Moderate (b) and High (c) noise added to a sub-threshold stimulus	253
FIGURE 8-1: Reaction times to judge stimuli for 8.1(a) Lexical decision and 8.1(b) Sentence sensibility	265
FIGURE 8-2: Mean reading times by Phrase and Condition	271
FIGURE 8-3: Normalised reaction times for Experiment 8.3a and 8.3b	280
FIGURE 8-4: Normalised reaction times (after item removal) for Experiment 8.3a and 8.3b by Condition	282
FIGURE 9-1: Schematic of where theories lie along the continuum from amodal to modal	326

Table of Tables

TABLE 1-1: Summary of early theories according to structure, decomposition and content	68
TABLE 1-2: Summary of current theories according to structure, decomposition and content	69
TABLE 1-3: Summary of embodied theories according to structure, decomposition and content	70
TABLE 2-1: Sensory and motor areas found in studies using fMRI or PET and language production	120
TABLE 2-2: Sensory and motor areas found in studies using fMRI or PET and language comprehension	121
TABLE 2-3: Sensory and motor areas found in studies using fMRI or PET and language comprehension (continued)	122
TABLE 3-1: The experiments which address comprehension on different visual processes	134
TABLE 3-2: The experiments which address comprehension on motion perception and motion perception on comprehension	135
TABLE 3-3: Summary of experimental design	136
TABLE 4-1: Scoring system used for picture choices in norming	142
TABLE 4-2: Item Set One – Up Category Verbs	144
TABLE 4-3: Item Set One – Down Category Verbs	144
TABLE 4-4: Item Set One – Control Category Verbs	145
TABLE 4-5: Item Set One Verbs – mean values for each matched variable	146
TABLE 4-6: Item Set One Sentences	148
TABLE 4-7: Item Set Two – Up Category Verbs	149
TABLE 4-8: Item Set Two – Down Category Verbs	150
TABLE 4-9: Item Set Two – Control Category Verbs	151
TABLE 4-10: Item Set Two Verbs – Mean values for each matched variable	152
TABLE 4-11: Item Set Two – Control Sentences	153
TABLE 4-12: Item Set Two – Up Sentences	154
TABLE 4-13: Item Set Two – Down Sentences	155
TABLE 5-1: Mean reaction times for Word Category by Block Half	165
TABLE 5-2: Mean reaction times by Condition for Experiments 5.1 – 5.4	170
TABLE 5-3: Descriptive Summary of Results	182
TABLE 6-1: Mean reaction times for both pilot Experiments for each Condition	193
TABLE 6-2: D prime data for Experiments 6.2 - 6.5 by Condition and Motion Direction	203
TABLE 6-3: Criterion (c) data for Experiments 6.2 - 6.5 by Condition and Motion Direction	209
TABLE 6-4: Beta data for Experiments 6.2 - 6.5 by Condition and Motion Direction	215
TABLE 6-5: Reaction time data for Experiments 6.2 - 6.5 by Condition and Motion Direction	216
TABLE 6-6: The mean difference in C and Beta for each Motion Direction by Word Category	223
TABLE 6-7: Descriptive summary of results	232
TABLE 7-1: Mean d' values by TMS application and Session	246
TABLE 7-2: Mean coherence thresholds (%) by Session and Motion Direction for Experiments 6.2 - 7.2	249
TABLE 8-1: Mean respond time for Sentence Judgement in Experiment 8.1b and 8.2b	262
TABLE 8-2: Mean reading time for each Phrase by Condition	270
TABLE 8-3: Mean reaction times by Condition and Experiment for 8.3a and 8.3b	278
TABLE 8-4: Errors for each Experiment by Condition	280
TABLE 8-5: Mean reaction times (after item removal) by Condition and Experiment for 8.3a and 8.3b	281
TABLE 8-6: Descriptive summary of results	286

1. Semantic Representation

1.1. *What is semantic information*

The focus of this research is semantic representation; therefore, it is necessary to begin with a definition of semantics. In the study of spoken language, information is canonically split into three information types: phonology, morpho-syntax and semantics. Phonology is the particular sound pattern that any given lexical item has as its word form and it includes metrical information. When studying written language phonology is replaced by orthography, the graphic shapes that make up letters or characters of a word form. Morpho-syntax refers to syntax and morphology; the combined term is used here although the two can be studied separately. Syntax governs how words are combined into sentences and marked as nouns, verbs, subject, object and so on. Morphology is the structure of word forms, which can change via inflection (e.g. walk *walked*) or derivation (e.g. nation *national*). In both cases, the information is structural: the word form or sentence structure systematically changes to express grammatical information. Semantic information is linguistic meaning. In psychological terms, it is those representations activated during production or comprehension that allow us to know what words and sentences refer to. In this investigation we will deal both with the semantics of sentences and of single words (lexical semantics).

1.1.1. **Semantic and conceptual knowledge**

"Any utterance is a selective schematisation of a concept" (Slobin, 1996, p.76)

As defined above, semantic information refers specifically to the meaningful representations recruited in the service of language, whereas conceptual information refers to knowledge in a more general, comprehensive, sense. This does not imply that there are no divisions within conceptual information as a whole, for example, the distinction between how to do something, like paint the Mona Lisa (procedural knowledge), what the Mona Lisa looks like (visual knowledge) and the last time I saw the Mona Lisa (episodic knowledge). Semantic information can cut across these different knowledge types. I can tell you when I last saw the Mona Lisa, I can attempt to describe what it looks like, and with great difficulty I could try to explain how to

paint it. Therefore, what is expressible in language varies for different kinds of knowledge. This has been divided in the memory literature into declarative knowledge, that which I can make explicit by using language, and procedural knowledge, which I cannot (Baddeley, 1999). The argument put forward in this section is that in order for semantics to serve language, it has to be treated as distinct from conceptual knowledge, even if that distinction is minimal.

Semantics must be grounded in conceptual knowledge; otherwise language would be useless for communicating our thoughts, feelings and experiences. It is precisely because language has this power, allowing conceptual information to be expressed and shared between individuals, that it is useful at all (Vigliocco & Vinson, in press). Conceptual and semantic information have often been treated as the same thing, for example, object naming has been partitioned into the accessing of visual-structural information (the object's form), 'semantic' information (its function and associations) and its label (Humphreys, Price & Riddoch, 1999). In addition, much work on concepts and categorisation has treated language as simply another access route to conceptual information, akin to using picture stimuli (Vinson & Vigliocco, in press). In this sense, the only thing that language does is attach a label to pre-existing concepts. However, there are good reasons for separating non-linguistic cognition, and its concomitant conceptual information, from information that acts as the 'meaning' for lexical labels. Especially in production, psycholinguistic research separates the 'message' (a prelinguistic conceptual representation that acts as the precursor to an utterance) from the semantic representations recruited in expressing that message (e.g. Garrett, 1975; Levelt, 1989; Bock & Griffin, 2000).

The reasons for separating semantic and conceptual knowledge center around two points. First, there is no steadfast relationship between general conception and lexical semantics. There are things that we can conceptualise of that we are not able to label with one word; this has been called the verbalisation problem. For example, most people have had the feeling that they can't say exactly what they want. Similarly, there are always multiple ways of naming the same concept, also known as perspective taking. For example, a specific horse can be referred to as Dobbin, female horse, mare, animal, or mammal; depending on what I want to convey (Levelt, Roelofs & Meyer, 1999). If language was a perfect mirror of all conceptual information, one would

expect that the relationship between the two would be invariant. Instead, there are variable ways in which the meaning of words maps onto conceptual structure. Polysemy and synonymy illustrate this point further, and are problematic for a one-to-one mapping between words and concepts (Murphy, 2002). If one word can have several meanings (e.g. bank), it must be connected to several concepts. Similarly, if several words can mean the same thing (e.g. sofa, couch, and settee), they must all be connected to the same concept. Unless there is some flexibility in how concepts build lexical semantics, it is difficult to say that one word equals one concept (Murphy, 2002). Second, by separating the two, the semantic system can have a different architecture to conceptual information (e.g. Vigliocco et al, 2004). This is useful because it allows cross-linguistic differences in semantics, such as English and Dutch having two words for 'leg' and 'foot' and Japanese only having one, 'ashi' (Vigliocco & Vinson, in press). In addition, it allows the incorporation of linguistic associations (like that between 'Jack' and 'Jill') and linguistic contexts of occurrence (like the fact that I have come across 'Jack' and 'Jill' in a nursery rhyme, and not in my Psychology Textbook) into semantics, without changing the basis of conceptual knowledge (e.g. Landauer & Dumais, 1997). Without a separation, semantics cannot be generative and productive, expressing novel distinctions across existing conceptual information. Thus, it is sensible to postulate a separation between "the aspect of words that gives them significance and relates them to the world" and "non-linguistic psychological representation[s] of a class of entities in the world" (Murphy, 2002, p.385).

There are numerous ways that theories present the relationship between semantic and conceptual information. For example, they can be seen as separate levels so that semantic representations are symbolic and abstract, independent from conceptual representations. This distinction is made so that the semantic level can carry grammatical information that is not present at the conceptual level (Levelt, Roelofs & Meyer, 1999). The idea that semantic information is abstracted from concepts is developed differently by Rogers et al (2004). Here, semantics is an abstract mapping between concepts and lexical forms; it does not actually carry any meaningful content itself but provides access to the relevant conceptual information. The semantic representations are amodal but influenced by the conceptual structure from which they develop. No mention is made of grammatical or other linguistic information and the theory treats mappings from words forms to concepts as isomorphic with mappings from pictures to concepts. Semantic information is therefore not much more than

sophisticated associations between conceptual content and symbols (i.e. word forms or pictures). However, many theories propose a much closer relationship between the two. For example, semantic representations can be seen as selected subsets of conceptual knowledge (McRae, de Sa & Seidenberg, 1997; Murphy, 2002), in this case the semantic level partitions conceptual information into chunks which serve as word meanings. Semantic information *is* conceptual information that is packaged for language; the two are therefore isomorphic in their content. In similar vein, Jackendoff (2002) argues that there is no separation between linguistic semantics and conceptualisation, there are only differences in how "linguistic forms map into complexes of meaning, not with the content of meaning itself" (p.292). However, he does make a distinction *within* the conceptual system. He postulates a conceptual system that is connected to language and supports abstract concepts, propositional thought and inference, and a perceptually based system that supports multi-modal schemas of objects and action within the environment. Thus, in order to account for certain phenomena, there has to be a distinction that separates information recruited for language versus information that is decidedly 'non-linguistic'. Where Jackendoff departs from other theorists is in placing this distinction deep within conceptual structure, proposing a qualitatively different system of representation, similar in flavour to Paivio's Dual-Coding theory (1986). Vigliocco et al (2004) begin with the principle that semantic information has to have close ties to conceptual knowledge, but propose that it is a separate level that extracts higher order relationships from basic conceptual features. Different conceptual features, such as those coding colour, shape or sound, are bound together into lexico-semantic representations. Whereas the conceptual level may be organized by modality, the semantic level is organized by featural similarity, with items that share many features clustered together. Semantics is therefore dependent on the conceptual level but separate, being organized by different principles.

Despite good reasons for a separation, there are several current theories that make no explicit division between semantic and conceptual knowledge. These are embodied theories, which propose that representations *directly* and *necessarily* recruit sensory and motor systems that are also utilised in conceptual representation, i.e. semantic representations are not abstract, they are intimately tied to experience and action (e.g. Gallese & Lakoff, 2005; Glenberg & Kaschak, 2002, 2003; Barsalou, 1999). Hard line embodied theories of semantics are the only theories that make this strong claim about the involvement of sensory and motor systems. It is unclear where embodied theories

stand on the question of the conceptual-semantic distinction. Some of these theories are only concerned with semantics (Pulvermüller, 1999; Zwaan, 2004), so it is possible that they are comfortable with a division, without being specific about its nature. However, other embodied theories see semantic and conceptual information as isomorphic and motivated by the same principles of organisation, proposing that the structure of sensory and motor information provides the basis for conceptual/semantic and linguistic structures alike (Barsalou, 1999; Gallese & Lakoff, 2005). In some sense, strong embodied theories like these have to assume that the conceptual and semantic levels operate similarly; it is difficult to see how semantics based in embodied simulation would emerge from non-embodied concepts. In contrast, amodal theories of semantic representation are still compatible with a fully embodied conceptual system since they are abstracted/derived from the conceptual level and can therefore ‘lose’ information.

In conclusion, there are cogent arguments for separating representations that serve linguistic (semantic) versus non-linguistic (conceptual) processes. In this investigation we are only concerned with the meaningful representations recruited during language production and comprehension, and we make no claims regarding conceptual representation. In the next two sections, brief summaries of two alternative views of general cognitive function and representation will be outlined. These can be loosely termed the symbolic and embodied traditions. By summarising these alternatives, the following sections, which describe past and current theories of semantic representation, are placed in a historical context. This is necessary for two reasons. First, the experiments reported here were designed to test theories of semantics which are driven by embodied views, and these theories are often motivated by criticisms of the symbolic tradition. Second, the historical context is useful for establishing the key issues in representation that any theory of semantics must address. At the end of the chapter I will outline these key issues, focusing on one in particular (*content*), before introducing the next chapter which reviews the empirical evidence for embodied semantics.

1.2. *Symbolic and Embodied Representation*

There are strong feelings around the term 'representation', as for some theorists it has become inseparable from a symbolic view of the mind (e.g. Varela, Thompson & Rosch, 1991). This is not surprising, since symbols designate things, they *are* representations. Since I am going to be using this term a lot, I need to start by clarifying its use. There are two points. The first is that whether we conceive of the mind as standing in some relation to the world (see Section 1.2.1) or as actively construing it (see Section 1.2.2) we are always concerned with how mental information is processed and transformed. Therefore, we are always concerned with the format of cognitive information and how it is instantiated. In this investigation, this broad definition is what constitutes the term 'representation'. It is a term used for convenience, to express how the mind/brain can come to have informational capacities at all. It is therefore used lightly, without any assumptions about how the mind and the world are related, or about how cognition achieves what it does. Second, this investigation is about cognitive representation in *language*, and however the mind is conceptualised, language is a symbolic process (Jackendoff, 2002). The power of language lies in the fact that words are meaningful, and they are meaningful only because they successfully *refer* to particular things. This is the simple classical definition of representation: for one thing to stand for another. *Without the capacity for representation, semantics is not possible.* The purpose of this first chapter is to explore different theories and their accounts of how semantics achieves representation.

Debates about the nature of cognitive representation have repeatedly focused around one question: Is information about the world embedded with perceptual and motor activity, being grounded in our embodied experience, or is it transformed into a qualitatively different, symbolic, format? These two different views, embodied and symbolic, have existed in various guises within various debates, such as mental imagery (Pylyshyn, 1985 ch.8; Kosslyn, 1996), conceptual representation (Barsalou, 1999; Murphy, 2002) and semantic representation (Fodor, 1987; Lakoff, 87). The symbolic tradition starts with the cognitive domain, what is 'in our heads', whereas the embodied tradition starts with the experiential domain, what is 'in our world'. These different starting points are the product of different philosophical traditions, beyond the scope of this work (see Lombardo, 1987), but it will become quite clear what they have to say about representation.

1.2.1. **Computation and Symbols**

The symbolic approach starts with one general assumption, there is a world 'out there' that needs to be represented by a mental domain 'in the head' (or brain). In order for this representation to take place, information is passed into the cognitive system via the sensory modalities. Perception is the passive reception of information from the world via our senses (Fodor, 1983). This information then has to be cognitively represented so it can be cognitively manipulated. Thus there is a separation between basic perceptual processes and higher “aggregative” processes such as language and problem solving (Simon, 1979, p.385). Pylyshyn (1985) posits three levels, the physical level of the brain, a representational level and symbol processing level. Perceptual information is transformed either at the representational or symbol level into something qualitatively different. The process by which perceptual information is turned into cognitive representation is termed transduction, literally transforming it from one information type (signal) into another (symbol) so that it can be manipulated by cognitive processes. These symbols internally represent the external world (or whatever is needed) and the mind then manipulates these discrete symbols with a set of elementary processes. Information processing provided “our first precise notion of what symbols and symbol manipulation means” (Newell & Simon, 1972, p.6), and when applied in psychology it made cognitive processes and representation tractable, rather than enigmatic, problems (Johnson-Laird, 1993). The symbolic approach was heavily influenced by predicate logic, propositional and computational formalisms (Fodor, 1983; Johnson-Laird, 1993; Simon, 1979; Pylyshyn, 1985). Computation was viewed as an incredibly powerful formalism through which to study cognition (Johnson-Laird, 1993), allowing dense and multifaceted processes to be broken down into elementary units and operations:

“It is a major achievement of modern logic to have shown that computational processes of any complexity whatever are reducible to (or, looked at another way, constructible from) concatenations of surprisingly small collections of basic operations... the postulation of a census of computational elements on the one hand, and of combinatorial elements of the other, the output... being generated by the... application of the latter to the former” (Fodor, 1983, p.29).

The idea of cognition-as-symbol manipulation was powerful in its explanatory promise, providing a means to precisely define and separate psychological processes. Symbols

and structures of symbols can “encode information about any conceivable thing (including an external object), hence [acting] as a surrogate for it” (Fodor, 1983, p.24). These ideas are fully laid out in Newell (1980), where he defines *physical symbol systems* (PhSS) – symbol manipulating devices that are physically realizable (like the computer, or human mind). A PhSS has inputs, outputs, a memory to store symbols and symbol structures, a set of operators (basic processes) and a control system for interpreting active symbol structures so as to guide the next action. Primitive symbols designate external events and objects, *but their nature as internal entities is the same no matter what they symbolize*; they are then able to undergo the same processes regardless of what they designate. This internal consistency allows “the mechanics of processing [to be] determined by the nature of the processor” (Newell & Simon, p.29) rather than by what the symbols represent in the external world. In other words, there is a stability in the internal processes and symbols that is independent of arbitrary variations in content. What is important about this approach is the *structure*, essentially extending predicate logic to the mental domain and developing a computational ‘language of thought’ (Fodor, 1987). Defined like this, the processes that operate on symbols become more important than the content of symbols themselves; symbols don’t change but processes that create, interpret and manipulate symbols do. Similarly, in computing, binary representation in the form of zeros and ones is able to represent whatever is required, but the real trick is in the processing that manipulates that binary code to do everything from providing emails to creating a photo album.

The stability of internal symbols is a necessary characteristic of these logico-propositional theories, in order that symbols can take part in proposition-like processes – comparable to algebraic equations or logical statements. However, one interesting consequence of inert, holistic, symbols is the notion of decomposition. If symbols make up the combinatorial elements of cognition, then cognition *must* be decomposable into those elements – both complex elements made up of simpler units and primitive elements that cannot be broken down any further. The idea of decomposition also applies to the complex and basic processes which manipulate symbols with more complex processes built out of simpler ones. The notion of decomposition is still pervasive in psychological theory, providing a neat way to operationalise theoretical constructs (see section 1.3.2.4 on featural theories of semantics). The search for cognitive primitives, i.e. basic units of cognition, remains controversial (Fodor, 1987; Jackendoff, 2002) but it has taken many forms. One good

example is Chomsky's Universal Grammar, which proposes fundamental parameters for linguistic structure (Chomsky, 1965). Decomposition is perhaps the first example in cognitive science of the search for fundamental cognitive principles (Chater & Vitanyi, 2003).

Finally, it is necessary to briefly discuss designation. For symbol systems, designation is the term for how a symbol comes to stand for something else: the link between the symbol and its referent. As noted above, this is the essential capacity of a symbol (and of representation). Newell (1980) notes that the designatory relation between a symbol and its external referent is left open – how does the symbol 'cat' come to designate an actual cat? Within a symbol system designation is realized by the processes that operate upon the symbols – how the system interprets them. The system interprets what a symbol stands for and this can influence what processes operate on that symbol. For example, it is only when the symbol is interpreted as 'number 2' that numerical calculation can be performed upon it. If the symbol is interpreted as 'cat', the system cannot use it in numerical calculations.

Designation occurs when physical inputs to the system are converted to symbolic units, the process of transduction (Fodor, 1987). For example, the sensory input related to my experience of cats has to be connected to the cognitive symbol for 'cat'. Transduction was seen as a basic operation, it was not necessarily relevant for higher cognitive processes. Recall that Pylyshyn (1985) proposed three levels, the physical level of the brain, a representational level and the symbol processing level. The representational level was argued to have content but no behavioural consequences (ibid p.29); only the symbols that stand for representational content have computational possibilities, and therefore causally influence the system's behaviour. In order for the representational level to exist, transduction has already taken place (i.e. its contents already refer to something). Therefore, transduction does not affect the operations of symbol processing; it is simply a necessary step to get there. Thus, *how* an internal representation does designate an external entity was sidelined in favour of defining higher order process at the symbolic level. This is a non-trivial choice, since theories of conceptual and semantic representation could then provide *symbolic* labels for concepts and semantic units, without having to define exactly how a symbol referred to something. The fundamental problem of reference was avoided. However, designation does matter since the system is semantically constrained: what the symbol *means* has an

influence on processing (Pylyshyn, 1985). This is a simple enough statement, since if functional processes were not semantically constrained, we would constantly see "semantically deviant" behaviours (ibid, p.36), for example, a simple calculation like '1 + 1' would produce the answer 'blue'. Therefore, cognitive symbols must have consistent and singular reference; they can only have one semantic interpretation. This is unlike true functional representations which have many possible interpretations, i.e. the symbol 'X' can refer to 'cat', 'blue' or 'Wednesday'. Basic, primitive, cognitive symbols cannot take this form; there must be a causal link between functional states and what is represented, otherwise any symbolic explanation is "gratuitous" (Pylyshyn, 1985, p.43). Therefore, the symbol system may not be entirely independent of its external referents. This is apparently in conflict with the abstracted nature of symbol systems and it was a problem that was explicitly recognized (Fodor, 1987; Pylyshyn, 1985; Newell, 1980). Elsewhere, it has been termed the 'symbol grounding problem' (Harnad, 1990; Vogt, 2002) and nicely illustrated by the Chinese Room Argument (Searle, 1980), a thought experiment in which an English speaker in a closed room receives Chinese symbols through a hatch and returns other Chinese characters according to strict rules, without ever knowing the meaning of the character strings. This demonstrates that if symbols are not causally linked to their referents (the person in the room does not know what the symbols mean), internal manipulation of those symbols is never enough to establish meaning (the person will never know what messages are communicated). In the end, for symbolic theories, the advantages of having symbols that could operate within a symbol system were ultimately considered to be more important than how transduction might shape symbols. Thus, symbol systems elevated processing structure above representational content.

1.2.2. Perception and Conception

The embodied approach, as the name suggests, starts from the simple fact that the brain developed as a control system for the body (Wilson, 2002; Clark, 1998). For that reason, all the cognitive processes that are performed by the mind/brain are heavily influenced by, and to some extent inseparable from, this relationship. The mental domain is not distinct from the perceptions and actions in which it is involved, and the notion that there are 'mental symbols' which stand apart from real-world interaction is rejected (Clark, 1997; Varela, Thompson & Rosch, 1991). It is no more sensible to talk of the functions of the liver outside the body than it is to talk of the functions of the mind/brain outside the body. By placing the body center stage, the environment also becomes important. In evolutionary terms, the body and the environment are intimately related, since the body acts within an environment in order to survive. Thus embodied cognition is about the primacy of the physical world, the body and its environment for cognition.

For embodied theorists, cognition operates via complex and dynamic interactions between perception and action, and it is therefore inherently dynamic and 'on the fly'; for example, allowing us to quickly catch and throw a ball, with little effort. This is in direct contrast to the basic tenet of symbolic approaches, that there are internal 'mental symbols' which represent the outside world, and upon which cognition operates in order to plan actions or respond to stimuli. Under the symbolic description, the mental world builds a model of the external world by transforming perceptual input (Clark, 1997). Under the embodied approach, there is only limited modeling of the external world and cognition is about real world action rather than symbolic representation (Clark, 1997; Varela, Thompson & Rosch, 1991). In this case, the world is its own best model so the environment becomes an extension of our minds; a simple example is using notes to remind yourself to do something. Rather than keeping check of everything internally, we externalize the burden (Clark, 1997; Richardson & Spivey, 2000). Embodied approaches can be anti-representationalist, rejecting entirely the notion that cognition needs or uses representations of the outside world (Varela, Thompson & Rosch, 1991). For example, when modeled as a dynamical system whose behaviour falls out of the physical parameters of the body and the environment, there is no need for mediating symbols to control action. Particular patterns of activity are a natural consequence of the physical structure of the agent's body and the environment that it is placed in; like

the way reaching for objects is dictated by the way our limbs are structured (imagine how different it would be if we had tentacles). All that is needed is for a sequence (like reaching out) to be set in motion and the parameters of behaviour *emerge* from the system as a whole (Clark, 1997; Lombardo, 1987).

There are various ways in which the embodied framework can be applied to cognition. Here are four approaches from those listed in Wilson (2002), in no particular order of importance: First, cognition necessarily involves perceptual and motor mechanisms since it always takes place in an environmental context, it is *situated*. Second, cognition takes place through on-line interactions with the environment so it is *time pressured*. Third, because the brain developed as a control system for the body, the primary purpose of cognition is to produce actions relevant to a situation. Cognition should be understood in terms of how it contributes to *action*. Fourth, cognitive activity is always intimately linked to perceptual and motor mechanisms, even when the cognition is "decoupled" from the environment (Wilson, 2002, p.626). For embodied theorists, this is also termed 'off-line' cognition, since it means that cognitive processes are independent of any particular environment (Clark, 1997); for example, when we use our imagination to think about things that are not present. In contrast, on-line cognition is anything that makes a connection between cognitive processes and the body or environment; this covers activities such as locomotion, object manipulation or playing a sport. This final point grows out from the first three: if on-line cognition is intimately connected to perception and action, in order to maintain a coherent system *off-line cognition should also be embodied*. It will depend upon perceptual and motor systems. This last claim is the most relevant for our purposes, since it directly links to semantic representation. As noted above, semantics is only possible if a representation can be instantiated that refers to something. This referential relationship has to be stable both when the referent is present and absent. In other words, if I say the word "cat" and you are only able to understand that it means 'cat' when there is a cat present that is not a successful representation; it is just a rather complex way of pointing to something that is there. It only becomes semantically stable if I can say the word "cat" and you always know that it means cat, whether there is cat present or not. Now the word is able to access a stable semantic representation of what a cat is, so you know what the word means. How can a system which is intimately tied to real-world action and dynamic, on-line, processes have stable representations? The answer is through *simulation*:

"In general, the function of these sensory-motor resources is to run a simulation of some aspect of the physical world, as a means of representing information or drawing inferences." (Wilson, 2002, p.633).

Thus, the system internally 'recreates' the environmental objects that it is capable of dealing with. As with the symbolic approach, the environment has to be internalised somehow, but instead of transducing the signal into a symbolic format, the signal is recreated. This idea of simulation has been repeatedly claimed to solve the symbol grounding problem (Lakoff, 1980; Barsalou, 1999; Glenberg & Kaschak, 2002). There is a non-arbitrary relationship between sensory-motor experience and representation. Thus, simulation allows you to keep representations without having to postulate a separation between mind and the world. This point has been expanded by Lakoff (1980), who retains strong representationalist ideas whilst embracing certain elements of embodiment.

"The human conceptual system is a product of human experience, and that experience comes through the body" (Lakoff, 1987, p.206).

For Lakoff (and others, e.g. Barsalou, 1999), the need for embodiment stems from faults within the logico-symbolic approach. According to abstract symbolic approaches, the contents of cognition have to be a veridical internal representation of entities in the external world. It follows from this that they represent the world in a correct (true) or incorrect (false) way. Lakoff calls this 'objectivism': there is an external reality which must be reflected by internal representations/symbols. He argues this is a false framework as it means that "symbols are given meaning independent of the nature of the human organism" (p.177): internal representations are evaluated against external reality, thus, they can either be true or false about how they represent external reality. The body and the physical ways through which we interact with the world play no part in conceptualisation. However, many concepts do not 'mirror' the world; they are products of human experience, for example, the category of things that I like to eat or even money (which depends on our belief that it has value). Human concepts often have no 'objective' counterparts: meaning is internal to human beings, rather than being a relationship of symbols to referents. Lakoff does not deny that there is an objective reality, only that "we have no *privileged access* to it from an external viewpoint" (p.259). In order to understand human conceptual representation, we must first

understand human experience, which is mediated by the body (perception and action). Conceptual structure emerges from our experience of the world and this experience is constrained by the world itself (what objects and events are there to experience) and how are able to experience them (our body's sensory and motor capabilities). In other words, we construct our conceptions of the world based on our experiences, so concepts are not simply an internal mirror of the external world; they are actively created and mediated by our experience. This is an interesting twist to embodied theory. Rather than placing body-environment interactions in the foreground with cognition emerging from it, the whole of experience is pushed inside the mind. There is no need to set up a link between the mind and world, since the world is already inside the mind. This view is adopted by Jackendoff (2002) in his theory of conceptualist semantics (see Section 1.3.2.3 below) and it is also accepted in cognitive neuroscience (Frith, 2007).

The idea that perceptual and motor systems can form the basis for all conceptual representation has been developed in more detail. For example, providing details of how particular neural mechanisms in the motor system could support motor concepts (Gallese & Lakoff, 2005), or how organisational principles in perception could support object representation and inferential processes during reasoning (Goldstone & Barsalou, 1998). As these attempts highlight, the problem for embodied approaches is how to account for abstract, higher-order cognition, such as that implicated in conceptual representation and abstract reasoning. Embodied approaches do very well for on-line action between a body and an environment, and offer some intriguing explanations about how the environment is utilised to benefit cognition (Clark, 1997; Wilson, 2002). However, in order to maintain the embodied framework, simulation is at present the only mechanism that could support semantic and conceptual representations. Simulation easily provides the content of representation with perceptual and motor activity, but how the information is structured and processed is more problematic and less well defined, especially for abstract cognition (Barsalou, 1999; Lakoff, 1980). Thus, for 'higher order' cognition like semantics and concepts, embodied theories raise the importance of representational content above processing structure.

1.2.3. **Symbolic versus Embodied Representation.**

In several respects, the symbolic and embodied traditions are directly opposed to each other. They begin at different places, symbolism asking how the mind can represent the world, and embodiment asking how the mind/body unit can act within the world. The symbolic approach adopts a clear structure for the computational processes of the mind, stating that there are processes which operate over symbols. The embodied approach adopts a clear content for cognitive processes, that which is provided by sensory-motor experience and activity. The symbolic account works very well for abstract reasoning and conceptual representation, embodiment works very well for real-world, real-time actions. Where symbolism falls down is in accounting for cognitive content and for real-time behaviour, where embodiment falls down is in providing a clear structure for cognitive processes and a good explanation for abstract cognition. When it comes to representation, the divide is sharp. Symbolic accounts predict a clear separation between representations (such as semantics) and sensory-motor systems, such that symbols are abstract and amodal. Embodied accounts predict an intimate relationship between the two, such that representations are grounded in perceptual and motor content. It is this prediction that is addressed in the current investigation; if semantic representation shares a common substrate with sensory and motor systems, the two should influence each other.

1.3. *Theories of Semantic Representation*

In this section a brief history of relevant theories will be provided. Selected theories are those which are explicitly about semantic rather than conceptual representation, as well as those that collapse the two but still refer to semantic representation. The majority of theories are concerned with the representation of lexical, rather than sentential, meaning. Their primary concern is the structure of the lexicon, proposing how single word meaning is established and related to the meaning of other single words. A few theories cover both single word and sentence processes; proposing how syntactic structures are formed during production (Levelt, 1989) or how individual words and syntax are integrated during comprehension (e.g. Zwaan, 1999; Glenberg & Kaschak, 2003). Sentence level semantics must be based on the integration of lexico-semantic representations, so all theories of lexico-semantics make limited predictions about the content of sentential meaning. As will become clear at the end of this Chapter, we are primarily concerned with the content of semantic representation (rather than its structure or decomposition). This is advantageous since content (amodal or modal) should not differ between lexical and sentential representation. Therefore, the particular predictions we are interested in should hold for both sentences and single words. Each theory will be described in some detail, and summarised with reference to three parameters: decomposition (e.g. holistic or featural), content (whether representations have modality specific or amodal/abstracted content) and structure (how semantic representations are organised relative to one another and relative to conceptual representation). At the end of the Chapter, three tables summarise the early theories (Table 1-1), the current theories (Table 1-2) and the recent embodied theories (Table 1-3) with respect to these three parameters.

1.3.1. Early theories: 1968-1995

1.3.1.1. Early Network Models (Quillian, 1968 and Collins & Loftus, 1975)

Quillian's network model is one of the first formalizations of how semantic memory could be structured. The aim was to create a computer program that could approximate semantic memory. The model focuses on the representations that allow "the "meanings" of words to be stored" (p.216), but Quillian does not make a separation between semantic memory and a general memory that is "flexible enough to hold anything stated in language, sensed in perception or otherwise known and remembered" (p.221), thereby assuming that semantic memory is isomorphic with conceptual memory. The model presents a network of nodes that are connected by associative links (see Figure 1-1). Nodes are named (e.g. 'cry', 'plant', 'animal') and take on two forms, depending on how they are connected to semantic information. *Type* nodes are 'parent' nodes and have direct links to the other nodes that constitute its meaning. *Token* nodes are related indirectly to a word's meaning by having an associative link which points to the word's type node. Token nodes therefore make up the constituents of meaning and connect different semantic representations to each other. Representations are cohered in planes, which comprise one type node and its constituent token nodes. A full representation is all the nodes reached by an exhaustive tracing process from the parent type node, together with the total sum of relationships (associative links) within and between planes. In this way, word meaning has both structure (links) as well as content (nodes).

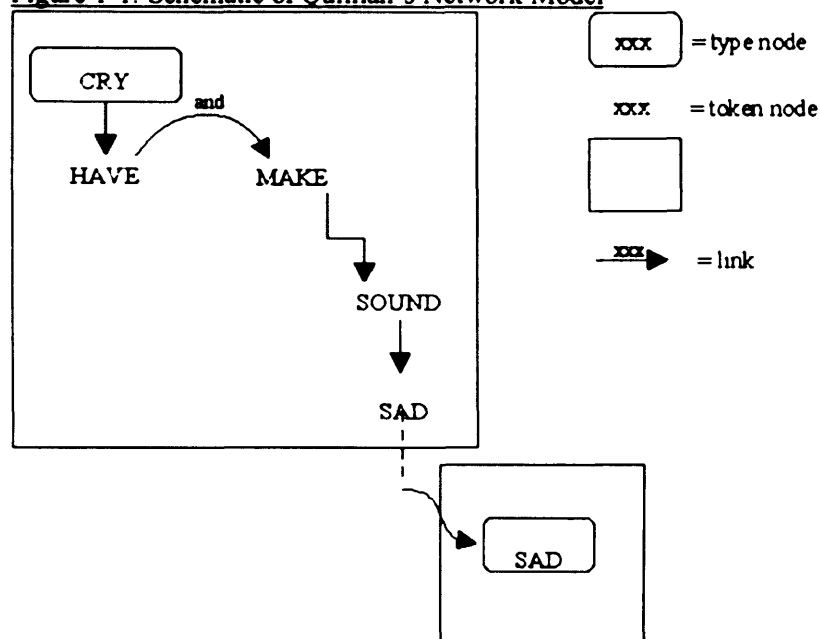
Within a concept plane the associative links are of different types and carry the particular relation between one node and another. For example, Modification (adjectival or adverbial relation) or Conjunction/AND (to link word meanings). Because the links carrying relations between nodes, the network actually has a propositional gestalt, such that when paths are traced from a parent type to token nodes, schematic sentences are formed. For example, if we trace from CRY in Fig. 1 we get the following:

[cry]: make sound sad

For this reason, despite being a network, it is logico-propositional in nature (McClelland & Rogers, 2003).

A key element of Quillian's model is the way in which information is retrieved, for example when concepts are being compared. Beginning at a particular parent node – one that we want to compare with another (e.g. CRY and SAD) – the node is *activated* and that activation spreads outwards via the associative connections. As activation *spreads* each node encountered is marked with a tag and when other tags are encountered, there is an intersection between concepts. After activating one node (e.g. CRY) the other is activated (e.g. SAD) and when activation from those nodes crosses, they can be compared. As we shall see, *spreading activation* becomes more salient in later adaptations of Quillian's model (Collins & Loftus, 1975), as it makes predictions about the speed with which similarity judgments can be made.

Figure 1-1: Schematic of Quillian's Network Model



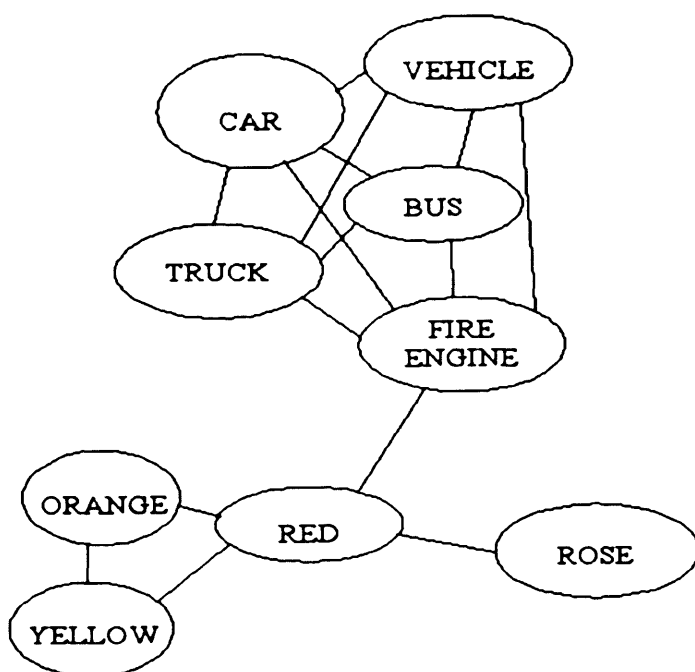
Adapted from Quillian (1968, p.239)

The nodes themselves are assumed to correspond to properties (an attribute with a corresponding value), rather than “visual pictures or words” (p.231), these properties are a base medium by which both linguistic and non-linguistic information can be supported. Quillian notes that “it seems logical to suppose that the same static store of information that underlies semantic reasoning may underlie perception” (p.229). In other words, the properties which form the basis of semantic representation may be of the same form as information that is acquired by perceptual mechanisms. This is a non-trivial point and it is in line with some form of embodiment, although it is not explored by Quillian. This model focuses on the *structure* of semantic memory, rather than its

contents. The representations (nodes) are holistic, localist representations (i.e. they are not decomposed and distributed over smaller parts) and it is the associative connections that provide a way of grouping semantic representations and establishing similarity.

Collins & Loftus (1975) expanded on Quillian's (1968) initial model so that they could encompass relevant behavioural results in semantic representation. They made some additional assumptions about local and global processes so that predictions about the time course of semantic processing could be made. At the local level, only one representation is actively processed at any one time, however, activation from different nodes can occur in parallel because activation spreads from the node of origin. Activation is continuously released as long as a node is being processed; however, the level of activation decreases with time and the distance from the node of origin. In addition, activation is a limited resource: the more nodes activated, the less each will be individually activated. Finally, for intersections between representations to be evaluated, activation at that intersection must pass a threshold. Globally, semantic memory is organized by semantic similarity with similar representations (e.g. vehicles, colours and animals) highly interconnected due to shared properties. Similarity is based on the total number of interconnections between two concepts, see Figure 1-2.

Figure 1-2: Schematic of Collins & Loftus Spreading-Activation Theory
A shorter line represents greater semantic similarity



Adapted from Collins & Loftus (1975, p.412)

When the assumptions about local and global processes are combined, a very powerful prediction emerges. When one member of a given category is activated, e.g. 'fire engine', it will activate 'bus', 'truck' and other vehicles as it is highly interconnected with them. However, 'fire engine' will not prime other red objects, such as 'rose', because there are so few connections between them. As such, vehicles will prime other vehicles, and flowers other flowers. Each representation is holistic and locally defined, as in Quillian (1968), and the links between them define how similar they are. Collins and Loftus used their Spreading Activation Theory to explain various behavioural results from experiments using property verification or naming tasks. Collins & Loftus do not make any further assumptions about the content of the nodes, concentrating instead on inter-connections and the structure of lexical activation.

These early models were primarily concerned with the structure of the semantic system, the way in which individual representations were related to one another by associative links. The links could be meaningful by denoting a specific relationship (Quillian, 1968) or they could be summed in order to define similarity (Collins & Loftus, 1975). The focus on semantic 'units' and their relations made two assumptions. First, the nodes implied *localist* representation. One thing is represented in one place, and that location can only represent that one thing. Second, the representations were self-contained, detached from one another (except for the links) and holistic. The connection between a meaning (e.g. 'fire engine') and its components (e.g. 'red') were provided by associative links rather than 'red' being an element *within* the 'fire engine' representation. In many respects, this was a reflection of their propositional heritage, providing stable symbolic units that could be used in propositional structures. Both theories confound conceptual and semantic information and it wasn't until later, similar theories (e.g. Levelt, 1989) that a separation was introduced. For Quillian, there is a hint that perceptual information may play a role in the content of representations, but for Collins & Loftus, no proposal is made.

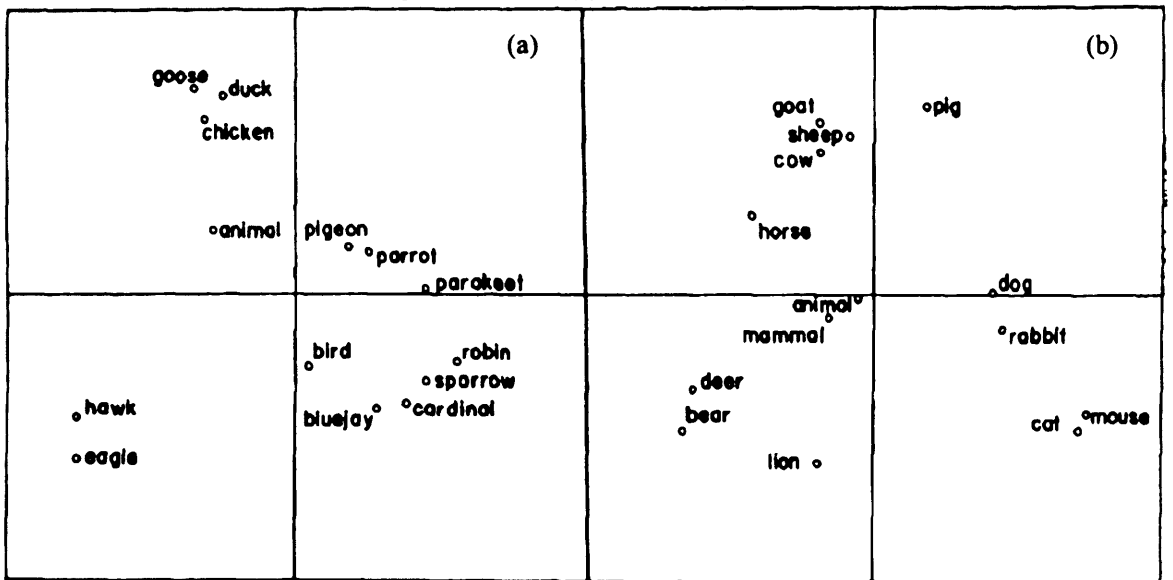
1.3.1.2.

Attributes of Meaning (Smith, Shoben & Rips, 1974)

An alternative to holism is to decompose semantic representations into attributes or features: “The meaning of a word is not an unanalyzable unit but rather can be represented as a set of semantic features” (p. 215). For example, the category of ‘birds’ has features such as ‘has feathers’, ‘has a beak’, ‘lays eggs’ and ‘flies’. For a given item, each feature has a weight that indicates how essential it is for that item. Therefore, the meaning of a word is a list of weighted features. Smith, Shoben and Rips make the assumption that “features associated with a given category vary in the extent to which they define that category... [with] a continuum along which some features will be more defining or essential aspects of a word's meaning, while others will be more accidental or characteristic features” (p. 216). The relative weights of each feature define their importance. In the model, for a given representation, e.g. ‘Robin’, features are ordered from the most essential (e.g. ‘red breast’) to the most incidental (e.g. ‘not domesticated’). However, at some point on this continuum there is a clear bound that separates the essential from the incidental: “This bound therefore creates a distinction with features above the bound being considered defining features and those below the bound being termed characteristic.” (p.216). Smith, Shoben & Rips used typicality ratings to establish how different category members (e.g. ‘Robin’ or ‘Penguin’) were considered as instances of a category (‘Birds’). A further assumption was that superordinate terms (‘bird’, ‘animal’) contained less defining features than their instances (‘robin’). Multidimensional scaling was used to present the ratings in a two-dimensional space that represents the characteristic features of a category. The proximity between an instance and its category represents their similarity on those features (see Figure 3).

In this model, the semantic features are not defined explicitly (e.g. ‘has wings’ or ‘has two legs’), they are assumed. However, the central idea of features as the organizational principle in semantic structure remains influential and has been used in several modern theories (see Section 1.3.3.3). This model was presented in direct opposition to the network models above, taking issue with holism and the use of associative links as providing relational structure. Instead, the features that constitute representations produce a structure where similarity is based on the number of shared features. In this way, featural theories move away from holistic, localist assumptions.

Figure 1-3: Multidimensional scaling solutions for (a) birds and (b) mammals



Taken from Rips, Shoben & Smith (1973).

All featural theories are by definition decompositional, since the features are component parts of the semantic representation. The structure of the semantic system is determined by the way features cluster to form different semantic representations, with similar representations having similar composite features. Like all featural theories, this model comes up against the problem of primitives, i.e. what forms a basic feature? For example, the feature 'has two legs' could be made up of two 'leg' features which are themselves made up of the features 'thigh', 'knee' and 'foot'. It is not clear how you can define a 'basic' feature (but see Jackendoff, 2002, for an interesting solution to this problem). As for the content of particular features, some of the semantic dimensions could be modality specific, such that 'has legs' is perceptual whilst 'non domesticated' is amodal, encyclopedic knowledge. It is not clear whether semantic information is modality specific or amodal; but the use of slightly abstracted terms such as 'has legs', rather than just 'legs', gives the features an amodal quality (c.f. Quillian's (1968) use of associative links such as 'has' or 'and' to define relationships). It appears as though the theory does conflate semantic and conceptual knowledge, since it uses only lexical labels but it is primarily concerned with categories.

1.3.1.3.

From Intention to Articulation (Levelt, 1989)

Levelt's theory is designed to account for language production and I will summarise the parts of the theory that refer to the preverbal message and the representations that mediate between the preverbal message and word forms. For Levelt, non-linguistic cognition is defined as spatial, sensory or propositional and the format that is recruited for language *must* be propositional. This is so that it can support divisions and relations in language such as thematic roles, tense, and categories (e.g. objects, events, places, directions). The preverbal message is built upon this propositional format, and it is the preverbal message that contains semantic representations.

Under this description, semantic representations are of basic types, for example, categories such as events, states, objects, locations, and directions. These categories can fill thematic roles within the preverbal message, such as agent, patient, place and path. Therefore, the semantic divisions map directly onto basic syntactic divisions. Put simply, this theory assumes that there is limited set of semantic categories and thematic roles, based on how "the mind organises the world of experience" (p.74), semantic representations must fulfill one of these types. It is at the level of the preverbal message that cross-linguistic differences are defined, in terms of what needs to be expressed in that language (e.g. the marking of tense). The preverbal message initiates the activation of specific lexical items, lemmas, which are selected from the lexicon (the store of all the words we know). Each item has specifications for its semantic content, syntactic class, morphology and phonological spell out and items are activated when the preverbal message meets their semantic or syntactic specifications. Relations within and between lemmas determine the structure of the lexicon. Within lemmas, relations exist for different specifications of, for example, morphological information. The different verb inflections (e.g. eat, eats, eaten, eating) are all found within the same lemma. Between lemmas, relations can exist because of shared conceptual information (intrinsic) or because of co-occurrence (associative). For example, the lemmas for 'rose' and 'flower' are intrinsically related because a rose is type of flower (hyponymy). Similarly, 'rose' is related to 'daisy' because they are *both* types of flowers (they are co-hyponyms). These intrinsic relations make up the structure of semantic fields. Associative relations exist between 'Jack' and 'Jill' or 'truth' and 'beauty' and they are the product of co-occurrence in language, rather than semantic similarities. Intrinsic and associative relations can influence one another, such that intrinsically related items can become strongly associated (e.g. 'cat' and 'mouse'), and associative relations can operate

over complex semantic relationships, such as that between 'truth' and 'beauty'. Semantic properties are shared on the basis of class membership, e.g. colours, flowers, animals etc., and so the structure of the lexicon is based on these coherent semantic fields. Although it is not spelt out clearly, since a division is made between the conceptual formats (e.g. spatial, sensory and propositional), lexical items are inherently symbolic; gaining their semantic content by reference to the conceptual system (rather than consisting of that content themselves). Thus, lemmas are holistic since they are accessed as wholes by the speech production system. This theory has more recently been implemented in computational model (Levelt, Roelofs & Meyer, 1999) with the assumptions of amodal, holistic and locally defined semantic representations implemented in a similar fashion to Collins & Loftus (1975).

1.3.1.4. *Parallel Distributed Processing* *(Rumelhart, McClelland & the PDP Research Group 1987)*

Parallel Distributed Processing is introduced here not as a theory of semantics, but as a theory of representation (*distributed representation*) which has been widely used in many different specific models of semantics (e.g. Farah & McClelland, 1991; Rogers et al, 2004). In opposition to localist theories where one unit represents one concept, representation of an item is distributed as a pattern of activations over a series of units. These same units are used for the representation of other items by means of different patterns of activation. Therefore the pattern as a whole, rather than the individual unit, is the meaningful level of analysis. It is worth going over the central elements of PDP networks, since this simplifies the explanation of subsequent models that employ them. A simple background is given here, followed by one of the earliest examples of a model of semantics that used a PDP network. Like network models, PDP used nodes/units that are connected by links; however, they are referred to as “simple, neuron like” computing elements (Rumelhart et al, 1987; McClelland & Rogers, 2003). PDP networks are synonymous with Connectionist or Neural Networks, which are more recent terms (Inman, 2000).

In a PDP network there are three layers of units: input units that receive external input, output units that produce external output and hidden units that can only interact with other units. These hidden units typically mediate between the input and output, allowing

representations at the output units to be reinstated given certain inputs. Any unit can be connected to any other by a link with a weighted value; if the value is positive the link is excitatory and if negative the link is inhibitory. The magnitude of the weight indicates the strength of the connection between the two units. The activation of a unit is defined within an equation that takes the (usually) summed inputs to that unit multiplied by the connection weights (e.g. McClelland & Rogers, 2003). Networks learn to represent particular material by being trained – for our purposes with suitably coded conceptual or semantic information. Training supplies activations for the input units (input) and target activations for the output units (output), altering the connection strengths between units so that the desired output is produced. There are a number of different rules that can be used to change the connections in order to achieve this. The simplest is the Hebb rule, in which the connection between two units is strengthened if they have similar activations, i.e. if they both respond positively or negatively to an input. The Delta rule expands on this principle by using the error between the actual activation of output units and the target output to modify connections in a way that reduces this error term (see Inman, 2000 for an excellent summary). In the next section (1.3.3), PDP models will appear in later formulations of semantic representation (McRae, de Sa & Seidenberg, 1997; Tyler & Moss, 2001; Rogers et al, 2004). This highlights the fact that these models can accommodate a number of different assumptions about the make-up of semantic organization. The focus on distributed representation encourages feature type inputs (and therefore decomposition) since the constituent parts of a representation can be presented at different input nodes. However, the features can be defined as modality specific (Rogers et al, 2004; Vigliocco et al, 2004) or they could simply correspond to more abstract classes of information, such as functional versus visual (Farah & McClelland, 1991; Tyler & Moss, 2001).

1.3.1.5. *Multimodality and Category Specificity (Farah & McClelland, 1991)*

Farah & McClelland (1991) used a PDP network to model the featural representation of living and non-living objects. These were defined as patterns of activation across ‘visual’ and ‘functional’ input units. Visual inputs encode visual features, e.g. colour and shape, whereas functional inputs encoded ‘use’ information, e.g. that mice are kept as pets or that a broom is used to sweep. In a norming procedure, participants were given passages describing various living and non-living things and they had to identify

visual or functional features in descriptions of different objects. For each object, the relative frequency of visual and functional features was then used to define the frequency of particular inputs into the model (rather than specifying individual features such as 'red' or 'kept as a pet'). Thus, each item was entered in to the model as a distribution of visual and functional inputs. The output units were the words corresponding to the objects ('naming units'). The model was motivated by the idea that *distributions* of feature types (such as visual versus functional features) could underlie object representation without a-priori category specific organization. Living things contain more visual features and non-living things (such as tools) contain more functional features, therefore, distinctions between the two fall out of their featural make-up. Thus, they assume that some features (visual or functional) are more characteristic for certain categories; note that this is slightly different to Smith, Shoben & Rips (1974) who used characteristic features *per se* to define categorical similarity. Farah & McClelland used the network to model category specific deficits for living versus non-living things, as found in patients with semantic impairments. By selectively damaging the visual feature system, living things were more impaired; conversely by selectively damaging the functional features, the non-living things were more impaired. The features for this model were extracted from verbal passages, but by defining categories from verbal passages it conflates conceptual and semantic information. However, it is assumed that features are grounded into a particular modality of representation, for example, the visual features would be represented by the visual systems and the functional features represented by the motor system. Therefore, the content of semantic representation is multimodal, rather than amodal, with combinations of features from different modalities aggregated into semantic representations for particular entities (in this case objects); this is in line with some form of embodiment. In line with PDP theory, localism is rejected and the representations are distributed over these features; the distribution of features supporting category specific organisation and the relationship between representations (i.e. semantic similarity).

1.3.2. Summary of Early Theories

Table 1-1 summarises where each of the early theories stand on the descriptive parameters in use (structure, decomposition and content). Early theories of semantic representation centered on network type models where lexical concepts could be represented either by individual nodes in the network (Quillian, 1968; Collins & Loftus, 1975; Levelt, 1989), or as distributions over units that represent semantic features (Smith, Shoben & Rips, 1974; Farah & McClelland, 1991). With the advent of PDP models, 'semantic networks' were advanced by the idea of distributed representation and the addition of features, elements of semantic content that could combine to form different representations (Smith, Shoben & Rips, 1974; Farah & McClelland, 1991). This introduced the distinction between localist and distributed representation. Feature based theories appeared early, and along with it a more explicit place for modal content as part of semantic structure (Farah & McClelland, 1991), rather than abstract symbolic representations with amodal content (Levelt, 1989). Holistic representation is favoured by early theories that use the connections between local nodes to quantify semantic similarity (Collins & Loftus, 1975, Levelt, 1989). Holistic representations also allow a propositional component to the structure of semantics (Quillian, 1968; Levelt, 1989), in line with the symbolic tradition. Distributed representation is favored by featural theories which employ decomposition; the focus is how a global structure emerges from individual elements, with little or no discussion of how the propositional elements of language are produced (Smith, Shoben & Rips, 1974; Farah & McClelland, 1991). The majority of early theories confound the global structure of semantics with that of conceptual information, making no explicit proposals about their relationship (Quillian, 1968; Collins & Loftus, 1975; Smith, Shoben & Rips, 1974; Farah & McClelland, 1991). As we shall see in the next section, features of one kind or another come to dominate in theories of semantic representation, along with more articulated ideas about the relation between conceptual and semantic content. In addition, recent years have seen the development of both moderate and extreme ideas about the dependence of representation on modal content.

1.3.3. 1995-present

1.3.3.1. Latent Semantic Analysis (Landauer & Dumais, 1997)

“The psychological similarity between any two words is reflected in the way they co-occur in small subsamples of language, that the source of language samples produces words in a way that ensures a mostly orderly stochastic mapping between semantic similarity and output distance.” (p.215)

As computational modeling developed, new approaches to semantic representation were introduced. Latent Semantic Analysis (LSA) uses the distribution profile of a given word within a large corpus of text to ‘extract’ the semantic profile of that word. This can then be compared to the profile for a different word to generate a measure of semantic similarity between those two words. LSA is primarily a mathematical treatment of semantic organization, one central claim being that distributional information combined with particular mathematical transformations is enough to successfully induce semantic content. LSA focused on knowledge acquisition and representation; since the model uses text as the source of distributional information, conceptual and semantic representation were treated as the same thing. The “contextual statistics of usage” (p.211) is the basis for inducting the semantic content of a given word: an Encyclopedia was used as the text and 4.6 million words were extracted for analysis. The text was broken down into samples, consisting of either a complete entry from the encyclopedia or the first 2000 characters, whichever was smaller. A 2 dimensional table was established, with 60,768 rows encoding one word each and 30,473 columns encoding the text samples. The cells in the table were marked with the frequency with which a particular word appeared in a particular text sample. These raw frequencies were then divided by the ‘entropy’ for that word, i.e. its frequency across many samples. The more samples a word appears in, the less informative any given sample is for the meaning of that word. Finally, the data was transformed using ‘singular value decomposition’, a technique similar to factor analysis, in which a complex data set is reduced to a number of dimensions that capture the most variance within the data set. These dimensions are represented within a multi-dimensional abstract space as a number of vectors that represent each word and each context: “The final output is a representation from which one can calculate similarity measures between all pairs consisting of either event types or contexts (e.g., word-word, word-

paragraph, or paragraph-paragraph similarities).” (p.216). Thus words that appear in similar contexts are judged to be similar. The similarity measure used was the cosine (angle) between two vectors; this value increases as the angle between two vectors decreases, thus, a larger value indicates a closer relationship.

LSA is an excellent example of semantics as embedded within a network: the meaning of a word is defined by its relation to other words, rather than by what it refers to. Recall that this is similar to the Early Network Models which define similarity as associations between individual items; LSA represents these associations as probabilities of co-occurrence in text. It is a classically symbolic approach (see Section 1.2.1) since the meaning of a word is defined as a set of abstracted symbols (vectors). The model is constrained by the real world referents of the words, their designation, only as far as the text sampled by the model faithfully encodes those referents. Consequently, the model has been criticized for being too abstract and has been shown to perform less well than theories of semantic representation which directly encode the real world properties of referents, such as their modal qualities (Vigliocco et al, 2004), or which take account of the way in which the referents can be employed in real situations (Glenberg & Robertson, 2000). However, it is still a powerful demonstration that distributional information is semantically informative; operationalising the straightforward assumption that the context of usage reflects and informs what a word means. It is not clear whether semantic representations utilize this distributional information but concepts do not (i.e. the two levels are separated). The article stresses that this mechanism is useful for several cognitive domains so we can infer that the two are isomorphic: this allows distributional information to support the acquisition of all knowledge, rather than being ‘limited’ to semantics. In line with symbolic approaches the focus here is on the process, proposing a mechanism that extracts the similarity of words. As with the Early Network Theories, the content of semantic representation is provided by quantified relationships between holistic, amodal units; LSA is a statistical reincarnation of these early theories.

1.3.3.2.

Conceptualist Semantics (Jackendoff, 2002)

In contrast to theories motivated by abstract representation, Jackendoff begins by arguing against the traditional view of reference that requires a link between linguistic symbols and 'the world'. This problem is exactly the same as that of designation outlined above (Section 1.2.1), if I am using language (symbols) to refer to something, how is that reference achieved? Jackendoff solves this problem by 'pushing' the world inside the mind, what he terms the 'mentalist enterprise' (see Lakoff, 1987 and Frith, 2007, for a similar view). Linguistic terms refer to the *concepts* that the speaker has about the world, rather than to the things in the world themselves. In order to purposefully understand and engage with things in the world, the mind constructs "cognitive structures in response to [these] inputs from the senses" (p.299) which form the basis of conceptual representation. These internal concepts *are* meaning since they are what the mind has constructed in response to experience with the external world - their very existence is the referential link between the mind and the world. For Jackendoff, it is the job of perception researchers to answer how such rich perceptual experience is established, providing a store of entities to which language can then refer.

As noted above (Section 1.1.1), Jackendoff argues that "we must consider the domain of linguistic semantics to be continuous with human conceptualisation as a whole" (p.282), so the distinction between meaningful linguistic and non-linguistic representation is made within the conceptual system. A propositional system that directly supports language is termed Conceptual Structure (CS) and the non-linguistic system is termed Spatial Structure (SpS). Language interfaces with the two systems in different ways. CS supports the propositional and abstract elements of language, whereas SpS supports the interface between language and visuo-spatial percepts (Jackendoff states that other modalities such as sound, taste, smell and feelings have their own modal counterparts). CS is hierarchical and componential, built out of discrete entities. It supports processes that use predicate argument structure, taxonomic structure (e.g. for categorical reasoning) and type-token relations. It is a propositional, combinatorial system that is not specifically linguistic but it does interface directly with linguistic processes (phonology and syntax), such that its structure is reflected in language. It is the divisions within CS that are 'visible' to grammar, for example, whereas the difference between 'yellow' and 'red' lies in their perceptual properties, for the purposes of CS they

are both labeled as 'kind-of-colour' and therefore undifferentiated grammatically. For abstract words that have no sensory-motor correlates and therefore no SpS components (e.g. *fairness*, or logical operators such as *and*, *if* and *not*), the proposal is that they have only CS components. SpS is spatial knowledge of the world integrated across several modalities, incorporating haptic, proprioceptive, visual and motor information. By allowing CS and SpS to be united under conceptual structure, removing language leaves you with "a non-linguistic association of cognitive structures in memory, much of which could be shared by a non-linguistic organism" (p.349). What produces a lexical item rather than just a concept is the addition of language as an extra modality into which the concept extends. Lexical semantics incorporates both elements from CS and SpS (being isomorphic with them), for example, the representation of 'cat' has both CS components that identify it as a mammal and a household pet, and SpS components that identify what cats look like, how they move and what colours they can be. One powerful consequence of the separation into CS and SpS components is that perceptual features such as 'has legs' or 'has a tail' can be attributed to SpS (i.e. supramodal content), rather than causing problems for a completely amodal system (cf. the grounding problem, section 1.2.1).

Jackendoff is firmly devoted to decomposition within all areas of conceptual structure, supporting compositional accounts of everything from object representation to verb meaning. It appears that this commitment to decomposition stems from the fact that it allows continuity to be found across domains, providing a structured account of how concepts are created, and it also allows the productive nature of combinatorial systems to be manifest across the whole 'meaning' system (CS and SpS). The components proposed by Jackendoff are best considered as features, although he prefers abstract features to modality specific ones (see below). Primitives of meaning do not have to be perceptual primitives, and for Jackendoff any decomposition will necessitate abstract primitives (such as 'cause' being a primitive for events). It is unclear what the relationship of these primitives would be to the CS and SpS systems, although presumably they would underpin the conceptual structure of both, with mostly perceptual primitives providing the basis for SpS and abstract primitives providing the basis for CS.

In sum, Jackendoff sees semantic and conceptual information as isomorphic, but he distinguishes between conceptual information that supports abstract and propositional components of language from perceptual information that supports non-linguistic, perceptual representations. Language has access to both these information types, with a more immediate connection to the CS system. In terms of content, the CS system is fundamentally abstract, and the SpS system is also abstracted from any one modality. However, he does note that features of modal information such as taste, smell and feeling have access to their own modal counterparts. Therefore, the larger conceptual structures that support meaning (CS and SpS) are one step removed from perceptual information *per se*, but it appears that there is still the possibility that a more direct connection between language and perceptual information is possible through specifically modal features. Although this theory is dependent on features (due to its decompositional nature) it is not included with other featural theories below, as it has not been computationally implemented for empirical purposes. However, it is assumed that in line with all featural theories, similarity between items is defined by the similarity of their featural make-up.

1.3.3.3. *Recent Featural theories:*

1.3.3.3.1. ***Featural Representations in Word Meaning*** ***(McRae, de Sa, Seidenberg, 1997)***

A descendent of featural theories such as Farah & McClelland (1991), the model put forward by McRae et al looked specifically at featural representations within a distributed (connectionist) network. McRae et al note that “all distributed representations incorporate featural representations of some sort” (p.100), since a representation is presented as a pattern of activation over a series of units (i.e. features). Whereas Farah & McClelland focused on functional versus sensory features, McRae et al instead looked at how different measures of feature distribution influence representation. Their aim was to look at how correlations amongst features would affect semantic judgments for different items (this idea is similar to Smith, Shoben & Rips (1974) characteristic features that are shared between items of the same category).

They looked exclusively at nouns from 10 different categories (e.g. mammals, fruits, clothing, furniture, vehicles and weapons) and asked participants to list physical properties, functional attributes and encyclopedic facts for a subset of those nouns. These speaker-generated features were then used to generate two different semantic representations: one derived from weighted individual features and one derived from correlated feature pairs. For the individual representation, each feature was weighted according to how many participants listed it for that noun. For the correlated representation, correlations were calculated between pairs of features (limited to 240 features shared by 3 or more nouns). Vectors were computed across these correlated pairs; for the features in a given noun, it's unit was set to 1 if the noun contained both items in a pair, 0 if it contained neither and -1 if it contained one but not the other (a violation of the correlation). Thus, individual and correlated representations were independently computed. Similarity in individual features predicted priming in artifacts but not living things, whereas similarity in correlated pairs predicted priming in living things but not artifacts. As in Farah & McClelland (1991), a division between living things and artifacts falls out of the feature composition; this time without a definition of the modal content of particular features. McRae et al proposed that living things have more correlated features because the constraints on them are less arbitrary than those for artifacts, because of evolution all animals have some similar properties (see Tyler & Moss (2001) below for a theory that unites visual and functional modes with shared and individual featural distributions by extending this argument). McRae et al note that "all feature-based theories of semantic memory assume that individual features are represented" (p.104), although nothing more is said about how this might be implemented. It is assumed that features are represented in a distributed, decompositional network, but the content of the features themselves is left open.

1.3.3.3.2. *Distributed Conceptual Knowledge (Tyler & Moss, 2001)*

The distributed account of Tyler & Moss deals explicitly with conceptual representation, however the authors operationalise features in a simple connectionist network, and use verbally generated features (words) as the basis for their model. They therefore collapse conceptual and semantic representation, treating them as isomorphic. For this reason, the model is summarised here. Tyler & Moss propose that

representations are structured around shared and distinctive features. Shared features are those that are common to different concepts, and distinctive features are those that are limited to less than three different concepts. This is similar to the correlated and individual feature analysis done in McRae et al (1997), except that here the two are computationally dependent (the more distinctive a feature, the less it is shared). When operationalised in a connectionist network, correlated features (those that co-occur and predict one another) support each other in representation, and are therefore less prone to damage. The model focuses on neuropsychological literature, and seeks to account for various category specific deficits. Using a feature generation task, features were collected for 93 different object concepts from animate (e.g. animals and plants) and artifact categories (e.g. tools and vehicles). It is assumed that information is “randomly distributed without category/domain organization” and the feature distribution means the connectionist model will organize itself into “graded, overlapping regions in semantic space” (p.249-250). A PDP model was produced with feature vectors (distributions of different features) entered via the input layer; training was complete when the same vector pattern was reproduced at the output layer. The features themselves are categorized as, for example, perceptual or functional (implying some modality specific content). The model was then ‘lesioned’ by randomly removing connections between the input, hidden and output layers. Results showed that the distinctive features of animate concepts were more vulnerable than those for artifacts, but the shared properties were better preserved. By categorizing the features as perceptual or functional, this theory links the arguments in Farah & McClelland (1991) and McRae et al (1997): living things have more shared features because they are perceptually similar. The theory takes on the standard assumptions that come with features: decomposition, distribution and similarity dictated by featural composition. Also in line with other featural theories (e.g. Farah & McClelland, 1991), conceptual and semantic information are isomorphic and there is a hint that features are modally defined.

1.3.3.3.

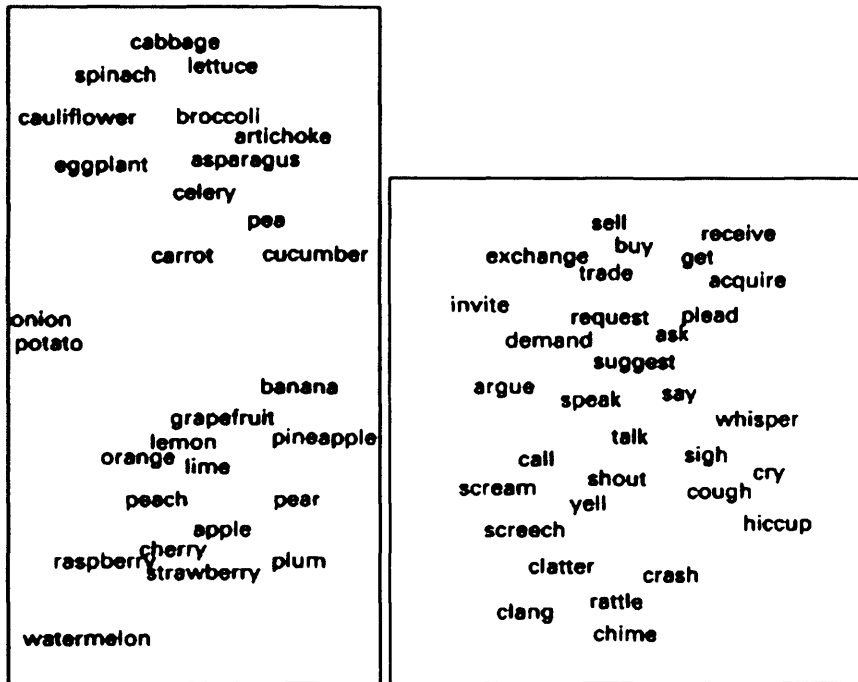
FUSS (Vigliocco, Vinson, Lewis & Garrett, 2004)

The Featural and Unitary Semantic Space hypothesis (FUSS) assumes that conceptual structure is organized by feature type, such as those that are modality specific (visual, motor etc.). A separate semantic level is derived from conceptual structure as features are bound together into lexico-semantic representations. To obtain features, native English speakers were given individual words and asked to generate features “sufficient to define and describe different words.. [whilst avoiding] free associations and “dictionary style” definitions” (p.7-8). For example, features for ‘Strawberry’ would be ‘red’, ‘sweet’, ‘fruit’ and ‘edible’. Features were collected for 230 nouns and 216 verbs, moving away from previous theories by modeling action as well as object representations. The same organizing principles were used for object and action words with lexico-semantic representations that share many features being grouped together.

Features were weighted by how many participants had produced that feature for a given item, and then combined across words. A self-organising map (SOM) modeled the feature space. SOMs, which are composed of an input and output layer, respond to statistical regularities in the input without other structural assumptions. All output units are connected to all input units and they are also laterally connected to each other. Feature vectors were presented in a random order via the input layer and the unit in the output layer that responded maximally to that vector (the ‘winner’) came to represent that collection of features. The winning unit inhibited all other units and changed its connection weights with the input so that it would respond the next time that feature vector was presented (i.e. it would strengthen its winning response). In addition, the lateral connections allow winning output units to alter neighbouring output units so that they respond more similarly. In this way, after training, a map organized by neighbourhoods of similar words was produced in which individual output units corresponded to individual lexical items. The output units are organized across feature vectors and are a best approximation of lexico-semantic representations, rather than a firm commitment to a processing structure. It is notable that rather than measure a first-order variable (e.g. correlations between individual features) the output units are organized so that second-order relationships (i.e. similarity across all features) dictate semantic structure. Similarity between two words was taken as being the Euclidian distance between their corresponding units on the SOM (see Figure 1-4). The model was used to successfully account for graded effects in priming, picture-word

interference and error induction tasks. Here, items that were more similar according to FUSS produced a greater number of errors in naming, greater priming and greater interference as distracters in picture naming. These results held for both action and object words

Figure 1-4: Representation of the semantic relation between items as depicted by a 2D representation of the multi-dimensional space
(Taken from Vinson & Vigliocco, 2002)



FUSS is put forward as both localist (one best responding unit) and distributed (a neighbourhood of responding units). There is clear commitment to decomposition at the conceptual level (i.e. conceptual features that are used to create semantic representations). These conceptual features are then bound together into more holistic semantic representations. This is a good example of separation between the conceptual and semantic allowing different organisational principles: decomposition at the conceptual level and a greater degree of holism at the semantic level. Theoretically, by assuming that modality specific features are an essential part of conceptual (and therefore lexico-semantic) representation, FUSS makes a clear statement about the semantic representation of words with sensory and motor content. These words are at least partly grounded in sensory and motor representations (operationalised as sensory or motor features, e.g. 'red' or 'used for kicking'), and as such, have embodied content. By separating lexico-semantic and conceptual levels, FUSS allows specifically

linguistic information, such as syntax, to be integrated with the lexico-semantic level. Lexico-semantic representations can also be repeatedly and differentially derived from conceptual structure, allowing different representations across languages. Similar to other featural theories, the number of shared features provides the structure of semantic organization and the way in which similarity is established. However, the SOM means that second-order relationships replace first-order measures that are more directly dependent on individual features.

1.3.3.3.4. *Semantic representations as mappings between verbal and experiential representations. (Rogers, Lambon-Ralph, Garrard, Bozeat, McClelland, Hodges & Patterson, 2004)*

"Semantic representations are defined with reference to the function that they perform and not the content that they encode" (p.216)

The following theory was developed to account for the deterioration of semantic memory in patients with "selective and progressive deficits in conceptual knowledge" (p.205). The theory is that semantics is the stable *mapping* between modality-specific information from different domains and as such it does not "code explicit semantic content" (p.206). The mapping allows that content to be accessed, given a particular task (e.g. comparing two concepts) or input (e.g. visual input corresponding to a named entity). The model is implemented in a simple PDP network that has two sets of input units (which also act as output units) corresponding to the verbal (words) and visual (sensory) modalities. A set of hidden units has connections to both these input pools, and these hidden units stand for the semantic representation. Mappings in these hidden units emerge when the model is presented with perceptual inputs for a given entity, for example the structured visual information produced by objects. In order to generate the input for the visual and verbal unit, the authors employed both verbal and visual feature generation tasks. In the verbal task participants list as many attributes (features) as possible for a given lexical item, in the visual task participants draw the items and features were extracted (e.g. tail, hooves, and legs). It is worth noting that the paper only deals with objects, so it is unclear how it would deal with more abstract entities (like events, or abstract concepts) which do not have such clear perceptual input.

Attributes were classified and feature vectors were generated for input to the model. The model was trained on visual (image), verbal (descriptions) and single name (label) inputs, producing representational structure in the hidden units that was constrained by, but not reflected in, the visual or verbal modalities. Objects that have similar perceptual inputs will also have similar semantic representations, for the simple reason that the perceptual inputs - and therefore the mappings from them - are structured in a similar way. Commonalities in the verbal labels used to refer to objects (for example in statements about objects: "The cat is my pet" and "The rabbit is my pet") will also constrain the semantic mappings. Similarity is thus a product of the "deep" (p.206) structure across and between modalities; it is a second-order variable that extends over features. This is distinct to most featural theories where measures of similarity are derived from first order variables fixed to individual features, such as correlations or the number of participants that produced that feature (e.g. McRae et al, 1997). However, one other featural theory also uses second-order measures to structure semantic representations. FUSS (Vigliocco et al, 2004) creates a space that supervenes over the individual features that define it, modeling lexico-semantic representation and similarity geographically as areas and distances in that space. Also in line with FUSS, the model of Rogers et al (2004) is compatible with decomposition at the conceptual level (although they do not commit to it) and holistic, unitary representation at the semantic level (a stable set of mappings). But where Vigliocco et al (2004) specify modal content and a dependence on semantic features, Rogers et al (2004) are strict about there being only amodal mappings. Thus, Rogers et al (2004) diverge from all featural theories since they do not use first or second-order relationships to demonstrate how effective features can be in accounting for semantic representation. Instead, the second-order relationships *are* the representation itself. One further difference is that unlike all other featural theories, no commitment is made to decomposition; the structured modal input is operationalised as attributes/features only for the purposes of the model¹.

The authors are explicit that the model is an abstract and amodal semantic store. However, the system is interactive so it builds a semantic architecture reliant on structured modal input. Thus, category specific deficits can arise if non-linguistic modal information is damaged as this disrupts the mappings from those sources. In sum,

¹ Note that mappings across different domains are difficult to implement if there are no features, i.e. no individual sources that can be mapped repeatedly for different representations. Therefore, I think Rogers et al (2004) have to make a commitment to some kind of decomposition, at least at the conceptual level.

semantic representations are taken to serve a particular function which in turn defines their structure. The semantic system consists of abstract representations whose content is the higher-order structure extracted from mapping across modal inputs. The theory does not explicitly discuss the relationship between semantic and conceptual information; however, since semantics is the link between conceptual information (non-linguistic modal input) and the linguistic modality, it is abstracted from conceptual information whilst maintaining a partial dependence on it.

1.3.3.4. *Summary of featural theories*

These theories all use featural stimuli to tap into the structure of semantic representations. Features are typically gathered from feature generation tasks, in which participants are presented with a word (e.g. apple) and told to generate as many features that define that word/concept as possible (e.g. green, red, edible, fruit, food, grows on trees, sweet, crunchy etc.), the features are then compiled. All theories using featural representations must assume at some level that the features are represented (McRae, de Sa & Seidenberg, 1997), and if those features are taken from verbal generation tasks and assumed to be a window onto underlying conceptual content, the theories can be criticized for ignoring how different features are acquired from different modalities (e.g. verbal, perceptual or motor) (Rogers et al, 2004). However, some models make explicit statements about the content, and origin, of particular feature classes (Vigliocco et al, 2004). The majority of theories are also committed to decomposition, since the features are component parts of semantic representations (the one exception being Rogers et al, 2004, but see footnote 1). All theories are concerned with the distributional power of featural representations and they employ some computational model to implement the features and extract these structural relationships. Therefore, although features usually commit theories to certain assumptions (decompositionality for example), they are primarily a tool that enables researchers to explore the basis of semantic relationships (how two semantic representations are similar and dissimilar, e.g. McRae, de Sa & Seidenberg, 1997), the structure of semantic representations (whether shared or distinct features across concept domains are important, e.g. Tyler & Moss, 2001), and the structure of lexico-semantic representation (whether certain properties of lexical representation, like syntactic class, naturally emerge from featural distributions, see Vigliocco et al, 2004). Again, the notable exception is Rogers et al (2004), where

features are used to define an amodal semantic system that maps between conceptual and symbolic information. Most featural theories collapse conceptual and semantic information (McRae et al, 1997; Tyler & Moss, 2001; Farah & McClelland, 1991) and in doing so they assume that both semantic and conceptual information are decomposed. Those theories that use second-order relationships between features to define semantic structure are only partially dependent on conceptual information (Rogers et al, 2004; Vigliocco et al, 2004) and are therefore compatible with decomposition at the conceptual level and a degree of holism at the semantic level (note that Vigliocco et al (2004) is the theory that makes this distinction explicit).

Table 1-2 summarises the above featural theories as well as Landauer & Dumais (1997) and Jackendoff (2002) along the descriptive parameters of structure, decomposition and content.

1.3.3.5. *Embodied, simulation accounts*

1.3.3.5.1. *The Brain's Concepts (Gallese & Lakoff, 2005)*

This is perhaps one of the strongest formulations of embodied conceptual and semantic representation, drawing heavily on the performance and representation of motor actions. It focuses on the precise neural implementation of embodied content, making no distinction between conceptual and semantic information. A key element of this theory is the idea that "*understanding is imagination*" (p.456) and the same neural substrates are used for perceiving/doing, imagining and linguistic understanding. A system which fuses information from different modalities (e.g. vision, audition and motor action) underpins cognition, and is also utilised by language, so that linguistic processes are not isolated from general cognition (i.e. language is not modular). The system combines modalities within the particular sensory and motor systems themselves, so it is multi-modal². As such, the sensory-motor system is necessary for representation. This is a strong claim, made by most embodied theories, but spelt out nicely here. Modality specific information could still, in theory, be recruited by association areas that integrate the modal information and have access to it, but do not

² This is in contrast to 'supra-modal' integration that combines information in a separate area that is distinct from the modal system itself.

directly or necessarily recruit them for representation. The evidence for this multi-modality comes from monkey studies which show neurons in premotor areas selectively responding to motor, visual and somatosensory information for the purpose of controlling movements in peripersonal space. Therefore, distinct modal information can be integrated within the modal areas themselves, rather than requiring a separate association area. The authors define areas which perform multi-modal integration (within a given modal area) as *functional clusters* that perform as a unit for a given computational function, for example, transforming the spatial location of objects into appropriate motor information for that object to be grasped or viewed. The operations performed by functional clusters are defined as *action simulations*; multi-modal integration happens for the purpose of performing relevant actions, and is therefore a simulation of, for example, interacting with an object at a given location. The integration is a simulation which acts as a plan for a given action. For example, neurons that fire when the head is turned to a particular location also fire when an object is seen or heard in that location; for the authors, this neuron is simulating the head being turned towards the stimulus. The structure of representations is inherent in these functional clusters, for example, they define a 'general-purpose subcluster' that is concerned with the general goal of an action, rather than the detailed manner or timing by which it is carried out (these have their own subclusters). In conceptual simulation, if this subcluster is activated on its own, it can represent an imagined general action with a particular goal: "We can conceptualise a generalised grasping without any particular manner being specified" (p.461). There are also subclusters that code for parameter values, for example, the role of who is performing a given action (me or someone else) or the level of force required to perform an action. These different subclusters are taken from the domain of motor actions and presumably similar clusters underpin perception and the other senses. With the addition of 'control clusters', the subclusters form the basis of conceptual representation, producing networks that represent "schemas" (p.467) (i.e. concepts). In sum, this theory proposes a heavily neural account of how embodied content underpins conceptual representation. Semantics representations, since they are isomorphic with concepts, have components afforded by the preexisting structure of the sensory and motor systems, allowing decomposition to be clearly maintained. Sensory and motor information is the necessary content of concrete concepts, and simulation using multi-modal systems underpins action, perception, imagination and conceptual representation. Presumably the similarity in modal content between two items dictates their semantic

similarity (working along the same principles as the visual/functional features of Farah & McClelland, 1991) and representations are clearly distributed (between different modal systems).

1.3.3.5.2. *Word Webs (Pulvermüller 1999, 2001)*

This account of word meaning is also neurally motivated, seeking to provide neural mechanisms for representation that are grounded in Hebbian learning (Pulvermüller, 1999). Assemblies of neural populations in distinct cortical areas act together to achieve representation, producing distributed networks of co-activated regions. The principles of Hebbian learning explain how these assemblies come into existence: "Coactivated neurons become associated... Associations can occur between adjacent or distant neurons...[and if] neurons become associated, they will develop into a functional unit, a cell assembly." (p.254, 1999). The activation of these cell assemblies is defined spatio-temporally across participating cortical regions and it forms the basis for representation. Neural activity that is stimulus specific and spatio-temporally defined has been found, suggesting a particular assembly of neurons that represent that stimulus: "A web of strongly connected neurons, each of them contributing to specific sensory and motor processes related to an object, may thus become the cortical representation of this object" (2001, p.517). Thus, there is a suggestion that conceptual representation (e.g. of an object) is also based in cell assemblies, with the addition of word form areas to produce semantic representations.

For semantics, the essential process is the association of areas relating to the word form (for spoken languages Pulvermüller locates this to the left hemisphere perisylvian area) with areas relating to the perceptions and actions to which the word refers. The coactivated neurons in the different cortical regions "develop into a higher-order assembly" (p.260, 1999), named 'Word Webs', that encompasses these word form specific and semantically specific cortical activations. If these coactivations occur frequently then the cell assembly of the word is changed to reflect the correlated activation of the different regions. Semantic content determines which cortical regions will be part of a word's cell assembly and once different cortical regions are bound into an assembly, their activation becomes interdependent. The different areas within an assembly will show stimulus specificity only if other areas of the assembly are intact (Pulvermüller, 2001). This suggests that the representations are not particularly robust

given damage in one part of the assembly, but no further details are given as to the dependence of assembly components. The activation of lexical and semantic elements of the word web are 'near simultaneous', with the first activation of the web, known as ignition, followed by active reverberation (maintenance) of activity within the web (Pulvermüller, 2001).

For concrete words, which have a clearly identified perceptual modality or experience, word webs involve sensory and motor areas. The word form is experienced in the presence of the percept to which it refers, the percept causes particular activations in the sensory and/or motor cortices and these become associated to the perisylvian activation related to the word form. Categories of words which follow this process are, for example, those corresponding to experiences in vision, motor action, smell, taste, pain, touch, sound and emotion. "Members of these word classes should be represented in assemblies with specific cortical topographies [which correspond to the modality of the referent.] For example, whereas an assembly representing a pain or touch word may include substantial numbers of neurons in somatosensory cortices, sound words may have exceptionally high numbers of neurons in bidirectional auditory cortices included in their assemblies" (p.262, 1999). Since the referents of concrete words recruit regions bidirectionally in sensory and motor cortices, these words are predicted to be less left lateralised than abstract open class words or function words, whose context of experience is more heavily linguistic. In sum, every word is represented by a Word Web that activates lexical areas related to the word form (e.g. perisylvian areas) and other cortical areas defined by the semantic character of the word's referent, and therefore the neural traces of how that referent was/is experienced. The content is influenced both by sensory-motor and linguistic experience: the structure of semantic representations is dependent on their neural organisation, i.e. which areas are recruited for the representation. As for Gallese & Lakoff (2005) concepts and semantics are based in similar organizational principles and semantic similarity is probably defined by the similarity of the cortical regions that represent particular words. However, this theory is not clearly decompositional, aside from the distinct cortical areas; but it is clearly distributed, since semantic representations are dynamic patterns across different cortical regions.

1.3.3.5.3. *Perceptual Symbol Systems (Barsalou, 1999; Barsalou et al, 2003)*

Using a nominal twist on the Physical Symbol Systems account put forward by Newell (1980, see Section 1.2.1 above), the Perceptual Symbol Systems (PcSS) account is the most wide-ranging theory of how simulation might drive human cognition. For example, the theory has been applied to conceptual processing (Pecher, Zeelenberg & Barsalou, 2004), social cognition (Niedenthal, Barsalou, Winkielman, Krauth-Gruber & Ric, 2005) and propositional reasoning (Barsalou, 1999). The theory makes a firm connection between perceptual and conceptual systems (Goldstone & Barsalou, 1998). There is direct modulation of perception by conceptual processing and the genesis of conceptual abilities lies in the pre-existing structure of the sensory and motor systems (Barsalou, 1999, p.598-590; Goldstone & Barsalou, 1998). In this sense it shares the strong assumptions of Gallese & Lakoff (2005) that higher-order cognitive processes are the product of lower-order, sensory-motor activity. The theory collapses semantic and conceptual processes, since both are based in simulations (Barsalou, 1999), whilst maintaining that 'shallow' processing of word associations, at the lexical level, circumvents conceptual (read semantic) activation (Solomon & Barsalou, 2004; Niedenthal et al, 2005).

This theory is more classically cognitive, making predictions about cortical regions rather than using particular neural substrates as a starting point. The focus is on a coherent system with component processes and functional units. The PcSS is based upon Perceptual Symbols, Frames, Simulations and Simulators. Perceptual Symbols (PcS) are based in the sensory and motor neural systems that are active when a percept is experienced, thus they are multi-modal, being established in the relevant sensory (visual, haptic, gustatory, olfactory, auditory) or motor area (introspection is also included as a modality of experience). PcS differ from perceptual 'images' or 'pictures' which might act as representations for a number of reasons. First, PcS are schematic, rather than holistic, re-presentations of perception. They do not recreate complete neural activations generated by perceiving a particular entity, rather, they are attentionally selected elements of that activation: "the symbol formation process selects and stores a subset of the active neurons in a perceptual state" (Barsalou, 1999, p.584). It follows from this that they can also be indeterminate, providing schematic representation rather than having to recreate every detail of a particular concept.

Second, PcS exist for the elements of a concept, as well as for the whole concept itself. These constituent PcS are separate and they make the system componential, combinatorial and productive (a necessary capacity for conceptual processing). Third, they are activated unconsciously, separating them from conscious imagery (although the two may be based in a similar recruitment of sensory and motor areas) and fourth, they are dynamic rather than static. PcS do not only stand for specific individual, fixed, entities. Whether the PcS designates an individual (a token) or a category of individuals (a type) is defined by the context of activation in combination with the PcS perceptually based content. Since they are established by dynamic neural activity, their subsequent activations can vary depending on context (which may emphasize a particular aspect of the symbol's content) or be changed by other PcS that are stored in the same area. In order to represent a concept, frames combine related PcS, for example, PcS for various components (doors, windows, wheels etc.) come together for the concept 'CAR'. The simulations that this frame and its relevant PcS perform constitute a simulator for a concept ('CAR'). The simulations themselves are never a comprehensive, complete, simulation of the entire concept. Instead, through *specific simulations* these simulators represent whatever aspect of the concept is under focus (e.g. whether I want to think about driving a car, washing a car or buying a car). Simulators represent types, and the simulations are the tokens.

For semantics, the implication is that on comprehension of a word the relevant simulator is accessed (e.g. 'CAR'). This simulator then performs a simulation relevant to the sentential context: "As comprehension proceeds, representations of individuals develop, as in the perception of a physical scene." (Barsalou, 1999, p.605) For example, "I drove the small car up the large hill" would produce a different simulation than "I drove the large car up the small hill", or even "I drove my car up that hill", as the simulations of 'CAR' and 'HILL' would vary depending on the specifications in the sentence. Barsalou sees the power of language in its effective construction of simulations, productively allowing novel and creative situations to be communicated; there is thus no difference between the simulations for semantic and conceptual entities. In this sense, lexical labels are another access route to the simulation mechanisms that underpin human cognition. Representation is decompositional, since different perceptual symbols are available for different features of a concept, and it is distributed, being realized across different modal areas. In this sense, it can be compared to featural theories that contain modality specific features (Vigliocco et al, 2004; Jackendoff,

2002), although they do not invoke simulation as the central mechanism. As for the other embodied theories (Gallese & Lakoff, 2005; Pulvermüller, 2001) we must assume that semantic similarity falls out of shared modal content, in this case, defined as shared constituent PcS.

1.3.3.5.4. *The Immersed Experiencer Framework (Zwaan, 2004)*

"Language is a set of cues to the comprehender to construct an experiential... simulation of the described situation... the comprehender is an immersed experiencer... and comprehension is the vicarious experience of the described situation." (p.36)

The Immersed Experiencer Framework (IEF) is a comprehensive theory of how embodied processes might work during language comprehension. As with other embodied theories, comprehension involves the simulation of whatever the language describes, and this simulation necessarily recruits sensory and motor representations. The IEF offers a detailed account of the comprehension process, stressing the 'vicarious experience' that is the product of comprehension. Since comprehension is based on experience, it has similar constraints, for example, only one perspective can be taken at any one time. The simulated representations are schematic, because language focuses on particular aspects of a situation, attentional capacity is limited and comprehenders use economy of processing, i.e. they do not activate more information than is necessary for comprehension. There are three stages to comprehension: Activation, Construal and Integration. Each stage is part of the incremental process by which a simulation is built, moving from individual entities (words), to scenes (clauses or intonation units), to extended episodes (connected discourse). The three stages are therefore differentiated by the amount of information that they combine and there is a large degree of temporal overlap between them.

During *Activation*, individual words activate functional webs which represent the object, action or quality to which the word refers. Representation is achieved because these functional webs are active when the referential entity is actually experienced, therefore they are necessarily comprised of the sensory and motor input that an entity produces.

Pulvermüller's Word Webs (section 1.3.3.5.2.) are cited as support for these functional webs, so it is possible that lexico-semantic representations are similar for these two theories. Just as there are many ways to experience an entity, for example I can see one object from many view points, several webs are activated by one word. These webs map onto each other diffusely and experience biases the base level of activation as more common experiences (e.g. canonical viewpoints) have more active webs. When there is no bias from experience (e.g. no canonical viewpoint) there is no dominantly active web. The webs are constrained by the preceding representations, which act as a context into which the webs have to fit (this is similar to the constraint of context on Barsalou's perceptual symbols). The more constraining the context, the stronger the bias for one particular web and the higher it's activation. This constraint satisfaction is termed 'articulation' and it happens during the second stage of comprehension: Construal.

Construal occurs over clausal or intonation units, integrating the webs accessed during Activation into coherent event representations. Linguistic cues, such as prepositions which specify locations, constrain the activation of different webs to those that are consistent with the appropriate perceptual representation. For example, "the teapot is on the shelf" would activate a side viewpoint, whereas "the teapot is on the floor" would activate a view from above. Construal occurs rapidly, resulting in an articulated simulation. Representations can be articulated by previous information, e.g. "the broken teapot" and by subsequent information, e.g. "the teapot that I broke". A construal is built from several components, each of which is an element of the simulated situation. There is a distinct temporal and spatial region; these define, for example, whether the event is ongoing and whether it occurs near or far from the immersed experiencer (IE). There is a distinct perspective and orientation of the IE relative to the entities (e.g. "I looked at the clouds" or "I looked at the clouds through the window"). There is a focal entity and a background entity, and a relation between the two (e.g. "The cloud moved across the sky"). Where these things are not made explicit by the language, they are implied by the entities and their real-world referents. For example, "The woman stared at the spider" implies different distances and orientations compared to "The woman stared at the skyscraper". Finally, the focal and background entities are further articulated by features which can be defined by adjectives (e.g. 'blue', 'slowly' and 'huge'). If the described situation is dynamic, e.g. "The woman ran away from the spider", then the construal is also dynamic.

Integration is the final, grand, stage of comprehension during which transitions are made between the construals produced from connected discourse. The transitions are "experientially based" (p.46) in as much as they are constrained by our real experiences of situations. Transitions can involve a change in focus, zooming in on a detail or zooming out to a wider scene, or they can involve a change in modality, e.g. when the IE sees an object and then hears it make a noise. The ease of integration is determined by the continuity in space, time and perspective between different construals. If subsequent construals jump around with respect to the time, place and perspective of a situation, then comprehension will temporarily require more resources. The overlap between construals in terms of the number of items, the focal and background entities, their features and so on, will also determine how easily transitions are made. Linguistic cues, such as the difference between definite ('the') and indefinite ('a') articles, effect integration since the former maintains a previous entity ("the bird on the grass") whilst the latter activates a new entity ("a bird on the grass").

In sum, the goal of comprehension is to construct a situation model, influenced by the perceptual and spatio-temporal characteristics of the referential situation. Perceptual and motor representations are necessary since comprehension is proposed to be the "reconstruction of an experience with its referent [that requires the] integration and sequencing of traces from actual experience cued by the linguistic input" (p.38). At the lexico-semantic level, functional webs may be similar to Word Webs (Pulvermüller, 2001) so similarity would be defined by the similarity in modal content/cortical areas that are recruited. Thus, decomposition is not an important aspect of lexico-semantic structure with distributed representations for each word that diffusely map onto one another. However, since this theory is primarily concerned with the comprehension of sentences and narrative, the focus is shifted away from lexico-semantics. Note that the later stages of comprehension (Construal and Integration) have defining parameters (e.g. perspective, orientation, overlap) that give comprehension its structure and a method for establishing similarity between larger, integrated semantic representations. In line with all embodied theories, there is a clear commitment to embodied content that can engage in the simulation of experience that is necessary for comprehension. The relationship between conceptual and semantic information is left open (but see the discussion in the penultimate paragraph of section 1.1.1. about the improbability of deriving embodied, simulation semantics from a conceptual level that does not work in the same way).

1.3.3.5.5. *The Indexical Hypothesis (Glenberg & Robertson, 2000; Glenberg & Kaschak, 2003)*

The indexical hypothesis seeks to link language, and specifically semantics, to the preparation of action within the environment. As with other simulation theories, the symbol grounding problem (see Section 1.2.1 above) is invoked as a damning argument against abstract symbolic theories, requiring an alternative theory of meaning that does not have this problem. In the indexical hypothesis the "mental representations of language are the representations of a situation (or the affordances of a situation) rather than a mental representation of the language itself" (Glenberg & Robertson, 2000, p.384), that is, a simulation of the situation is performed and this is how comprehension is possible. This is therefore similar to the IE framework (Zwaan, 2004) where comprehension involves the vicarious experience of the described situation. By taking as its starting point the non-controversial idea that cognition developed to "coordinate effective action" (Glenberg & Robertson, 2000, p.383), the theory proposes that "the meaning of a situation to an individual consists of set of potential actions available to that individual in that situation" (Glenberg & Kaschak, 2003, p.100). The set of possible actions is constrained by the body of the actor and the objects that can be acted upon (the affordances), the learning history of the individual and the goal of the action. The affordances dictate how objects can be acted upon, for example, bicycles afford riding by humans but not by fish, and bicycles can be used to get from place to place, but not to cook food. The individual's experience (their learning history) introduces cultural and social norms into the constraints: I can ride the bicycle that I own, but not the one that belongs to a stranger (unless I want to steal it). The goal constrains the action for obvious reasons, if I want to get to the shops I will use my bicycle, and I will also use my bicycle when I want to do some exercise. I will not use my bicycle when I want to cook some food (unless I have no food and have to go to the shops). In order to understand a sentence, the Indexical Hypothesis proposes three stages. First, the words are *Indexed* to objects in the environment or to perceptual symbols (Barsalou, 1999), which stand for real world objects. *Affordances* are then derived from the perceptual symbols, which lend themselves easily to this since they simulate the situation described in the sentence, necessarily recruiting sensory and motor information in a context dependent way. The derivation of affordances is based on previous experience and is also dependent on the other perceptual symbols already in use (e.g. the actions involved in 'riding'). Thus, as for other embodied theories (Barsalou, 1999; Zwaan, 2004) the

integration of perceptually based representations provides an easy way to sensibly constrain semantic interpretation: what is possible in the real world is possible in semantics. The Indexical Hypothesis makes the clearest exposition of this point. Finally, the affordances of the various symbols are then *meshed* based on the syntax of the sentence; for example, "Sarah rode my bike" allows that Sarah (the subject) is acting upon the bike (the object). In addition, the use of the verb 'ride' allows that Sarah is sitting on the bicycle controlling its movement, rather than carrying it or lying under it. Meshing occurs when the potential actions in a situation can be successfully combined to accomplish a goal. For example, if 'Sarah' was indexed to a boat, rather than a person, affordances could not be derived and the sentence would be nonsensical (i.e. impossible in the real world and therefore impossible in semantics). The three steps (indexing, derivation and meshing) are not strictly ordered, since meshing can influence which affordances are derived. The content of semantic representations is decidedly embodied, but as for Zwaan (2004) the focus is on sentential and narrative comprehension, so lexico-semantic representations are dealt with more briefly. However, Barsalou's PcSS (1999) is cited as the basis for lexico-semantics, so we can assume that the same theoretical parameters are applied: an isomorphism with concepts, decomposition, distributed organization and item-by-item similarity based on shared modal content.

1.3.3.5.6.

Summary of Embodied theories

These theories explicitly build on the embodied approach, applying the ideas to semantic representation. In addition, most theories deal in comprehension rather than production. Presumably in production, the simulations are active prior to the selection of word forms. There are some theories which attempt to provide a structure for embodied content within semantic representations, either by drawing on structure within perceptual and motor systems and thereby establishing decomposition (Barsalou, 1999; Gallese & Lakoff, 2005) or by referring to cortical organisation which gives a looser description of semantic structure (Pulvermüller, 1999). There are also theories that focus on the comprehension process and how a meaningful representation is built from the language input (Zwaan, 2004; Glenberg & Kaschak, 2003). For these theories, more structure is provided for the integration of representations into a coherent situation model, and other embodied theories are cited as possible implementations of lexico-semantic organization; therefore assumptions from those theories are adopted. The

unifying theme is the *necessary* inclusion of sensory-motor content in representation and the direct modulation of sensory and motor areas by semantic content (in order to re-activate perceptual and motor experience). Therefore, in all cases, semantic similarity must be established by the similarity in modal content; this is comparable to featural theories that allow modality specific features (e.g. Farah & McClelland, 1991; Vigliocco et al, 2004). Embodied theories are very clear about semantic content, so another consequence of emphasizing modality specific content is that semantic information is probably isomorphic with conceptual information; this is because whatever is present in semantic content must be present at the conceptual level too (as long as it cannot be confined to language alone, e.g. syntactic markers). Table 1-3 summarises the embodied theories along the three descriptive parameters of structure, decomposition and content.

1.4. *Key issues in representation*

As we have seen, there are several ways to theorize about semantic representation. I have used three descriptive parameters: decomposition, structure and content to provide a common framework for comparing across theories. Tables 1-1, 1-2 and 1-3 summarise where each theory stands on these three parameters and I will now treat each one in turn, ending with content, which provides the motivation for the current work.

1.4.1. **Decomposition**

The idea of decomposition, that representations can be separated into distinct components, came from the symbolic tradition but it has become an essential part of many theories of semantics, whether they propose that semantics contains abstract, amodal or modality specific representations (Barsalou, 1999; Jackendoff, 2002; McRae, de Sa & Seidenberg, 1997). Decompositional theories stand in opposition to holistic theories which propose that semantic representations cannot be broken down into smaller sub-components (Levelt, 1989; Collins & Loftus, 1975). Decomposition implies two things, first, that individual components are represented, and second, that the components combine to form 'higher order' representations. Featural theories are the best example of how decomposition can be applied to semantic representation. They assume some level of decomposition, for example a dichotomy between perceptual/visual and functional features (Farah & McClelland, 1991; Tyler & Moss, 2001) or abstract and visuo-spatial features (Jackendoff, 2002). There are also those which are more diverse, with features from all modalities as well as more abstract features, such as encyclopedic facts or category labels (McRae, de Sa & Seidenberg, 1997; Vigliocco et al, 2004). Featural theories concentrate on how the distribution of features can explain the structure of the semantic system, for example whether features are shared/correlated between items or idiosyncratic (McRae, de Sa & Seidenberg, 1997; Tyler & Moss, 2001; Vigliocco et al, 2004). It is also possible to propose that decomposition (features) is present at the conceptual level, being bound into more holistic representations at the semantic level (Vigliocco et al, 2004).

Theories that propose holistic representation at both the conceptual and semantic levels are less common (Levelt, 1989; Levelt, Roelofs & Meyer, 1999; Landauer & Dumais,

1997) although some of the recent embodied theories appear to be more holistic than decompositional. Here, a distributed web of activity is proposed as the semantic representation so there little internal structure to the representation beyond that activated for a particular context (Zwaan, 2004) or modality (Pulvermüller, 1999). However, these theories do not make an explicit commitment to holism, so later versions may specify a method of decomposition as other embodied theories have done (Barsalou, 1999; Gallese & Lakoff, 2005). The advantage of decomposition for embodied theories is that there can be a direct contact between semantic and sensory-motor content. If the conceptual level consists of sensory-motor features (i.e. component parts) and semantic representation is isomorphic with those conceptual features, then semantic content is directly linked to sensory-motor information.

Holism implies that semantic representations are independent but connected; therefore decomposition allows one to grasp the elements of representation where holistic accounts do not. For decompositional theories the relationships between representations are given by their internal similarity; for holistic accounts the associations between representations provide similarity (with similar concepts being linked together). However, these associations are typically defined by what the theorist assumes to be similar (e.g. categories such as flowers and vehicles, Quillian, 1968) making their definition at bit circular. However, later theories developed ways to extract associations from existing corpora (Landauer & Dumais, 1997), providing a better motivation for them. At present, the majority of theories subscribe to some level of decomposition, but arguments against 'semantic primitives' are still used to refute their plausibility (Fodor, 1987). The argument is that decomposition is difficult to stop in a sensible way, at what point does something become a semantic primitive (i.e. no longer divisible)? For example, the feature 'red' could be a primitive, but it can be applied to many different shades or hues – do these need their own primitives? In addition, typicality effects in categorization suggest that the boundaries between concepts are graded and fuzzy (Murphy, 2002), therefore clean decomposition is problematic. For some, the benefit of decomposition is too valuable in explaining semantic representations (Jackendoff, 2002); the argument is made that semantic primitives may not behave in the same way as lexicalized features. For example, the lexicalized feature 'red' may well be made up of semantic primitives that have a different organization, allowing fuzzy boundaries whilst maintaining decomposition. Drawing comparisons with the successful use of primitives in phonology, semantic primitives may not be consciously accessible or they

may not combine in the same way as components at higher levels. For Jackendoff (*ibid*) an important part of the research program into semantics is the progressive refinement of decompositional theories so that primitives can be realized.

1.4.2. **Structure**

Two aspects of structure will be reviewed: first, whether theories are distributed or localist (this has close ties to decomposition) and second, how semantic and conceptual information is related (this was also dealt with in section 1.1.1 and will be discussed briefly here). Beginning with the distributed/localist aspect, featural theories are typically distributed. This is because distribution (e.g. shared or individual features) is the most important element in establishing semantic similarity and structure when it is based on features. Thus, coherent representations emerge from these feature distributions. In contrast, holistic theories are typically localist because individual representations have to be coherent entities that stand alone in relation to other entities. The semantic network has been around since the first computational models (Quillian, 1968) and the principles of a connected set of nodes has been used in both localist (Levelt, Roelofs & Meyer, 1999; Collins & Loftus, 1975) and distributed (Farah & McClelland, 1991; McRae, de Sa & Seidenberg, 1997; Tyler & Moss, 2001) models, reincarnated for distributed theories through PDP models (Rumelhart, McClelland et al, 1987). Therefore, localist theories can integrate some form of distribution in higher order semantic structure. For example, local entities can be connected by associative links so that the global structure of the semantic system is a distributed network of local nodes (Quillian, 1968; Collins & Loftus, 1975; Levelt, 1989). Alternatively, the distributional relationships between local entities in a large linguistic corpus can be used to model semantic similarity (Landauer & Dumais, 1997). However, at the single word level decompositional theories are distributed and holistic theories are localist. Early theories employed localist representations (Levelt, 1989; Quillian, 1968; Collins & Loftus, 1975), but for several reasons distributed representations are currently the most popular way to conceive of semantics. Distributed representations are more compatible with how the brain appears to work (Pulvermüller, 1999); they can accommodate higher order relationships across the same set of units, allowing the representation of many things to emerge from the same 'space' (Rogers et al, 2004; Vigliocco et al, 2004; Tyler & Moss, 2001); and they also manifest how statistical information about distribution

can be used in semantic representation (McRae et al, 1997; Landauer & Dumais, 1997). A few embodied theories propose an internal structure to lexical semantics based on traditional notions of decomposition (Barsalou, 1999; Gallese & Lakoff, 2005) or geographically define structure by the cortical areas that are implicated (Pulvermüller, 1999). Embodied theories typically employ distributed representations that are grounded in modality specific content; the neural basis of a multi-modal system is impossible without distribution. If different geographical areas support individual representations, it is nonsensical to say that those representations are locally defined (although it might be possible to argue around this by separating functional and neural descriptions, as the original Symbolic approach did: Pylyshyn, 1985).

The relationship between semantic and conceptual information is absent in several theories because no explicit statements are made. For example, some theories simply neglect the distinction by modeling concept or 'knowledge' representation through linguistic stimuli (Tyler & Moss, 2001; Landauer & Dumais, 1997; Farah & McClelland, 1991). Some conflate the two by using the terms 'semantic' and 'conceptual' interchangeably (Smith, Shoben & Rips, 1974) or by carrying over assumptions from a previous theory (Collins & Loftus, 1975) without an explicit statement that the two are isomorphic. However, for those theories that do make clear assumptions, there is a continuum from a semantic system which is completely separate to the conceptual level (e.g. Levelt, 1989) to one that is completely subsumed by it (e.g. Jackendoff, 2002; Quillian, 1968). In between there are theories which propose that the semantic system does not carry any meaningful content per se, but acts as a mapping between conceptual content and symbolic forms such as words and pictures (Rogers et al, 2004). The mappings are connected to and influenced by the structure of both the symbolic forms and the conceptual content; therefore the relationship is not one of complete independence. Vigliocco et al (2004) use similar principles, stating that the semantic system captures higher order relationships (i.e. collections of features) over conceptual content. However, semantic content is not an abstraction so it is dependent on these conceptual features; therefore a closer tie between semantic and conceptual information is retained. Moving further along the continuum towards isomorphism, for McRae, de Sa & Seidenberg (1997) semantic representations are subsets of conceptual knowledge which act as word meanings. This also appears to be the view taken by some embodied theories (Barsalou, 1999; Gallese & Lakoff, 2005) where semantic and conceptual access is one and the same. Others concern themselves only with linguistic

stimuli (Pulvermüller, 1999) but hint that conceptual knowledge has the same foundations (see section 1.3.3.5.2). To clarify their position I have argued that embodied approaches must assume some isomorphism between semantic and conceptual content, simply because if semantics are derived from conceptual content, they can only lose information relative to the conceptual level (i.e. become amodal) rather than gain it (i.e. become modal). The distinction between semantic and conceptual knowledge is often neglected despite the fact that is important in understanding the formation of semantic representations. In future, it is hoped that more theories will be explicit about the assumptions they make on this matter.

1.4.3. **Content**

The issue of what is actually *contained within* a semantic representation has recently become a hot topic. This is related to the emergence of embodied theories, which propose that semantic content is sensory and motor information (e.g. Barsalou, 1999). This goes against the symbolic tradition, from which some semantic theories have been motivated (e.g. Levelt, 1989); however, it is worth noting that modality specific content appeared early (e.g. Quillian, 1968) but was never pushed as the main motivation for any particular theory until now. Currently, several theories employ modality specific content within semantic representation, without postulating that the content is necessarily or directly accessed (Rogers et al, 2004; Jackendoff, 2002; Vigliocco et al, 2004).

It is interesting to note that the continuum in the above section, detailing how semantic and conceptual information are related, is often correlated with the continuum from amodal to modal semantic content. There are of course those exceptions where no assumptions about modality are made (Tyler & Moss, 2001; McRae, de Sa & Seidenberg, 1997) but typically, independence between semantic and conceptual content occurs with amodal theories of semantics and isomorphism (or increased dependence) occurs with modality specific theories of semantic content. This is not that surprising since conceptual information has to have its genesis in sensory and motor experience and is therefore more likely to be modality specific and based in sensory and motor systems. Thus, semantic representations can only be amodal if they are independent of conceptual representations (or if some concepts are themselves amodal,

Jackendoff, 2002). I would like restate the point that in this investigation we are only concerned with word and sentence meaning, and we do not make any assumptions about conceptual content.

Figure 1-5: Schematic of where theories lie along the continuum from amodal to modal

**Collins & Loftus; McRae, de Sa & Seidenberg and Smith, Shoben & Rips not included as no clear assertions are made*

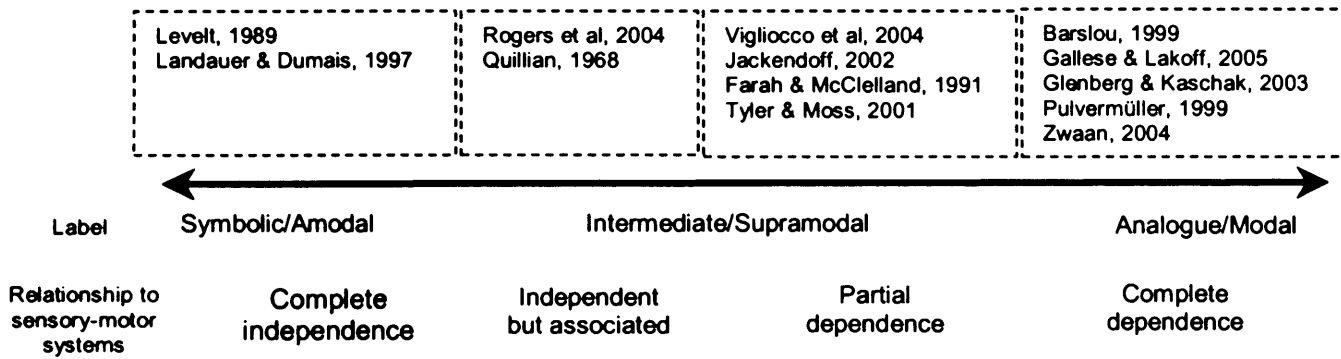


Figure 1-5 demonstrates where theories lie along on the continuum from amodal to modal. At the extreme left there are amodal theories which propose that semantic representations are abstract and symbolic, separate from modality specific information and therefore independent of sensory-motor systems (Levelt, 1989; Landauer & Dumais, 1997). Then there are theories which propose a largely symbolic/abstract semantic system (Rogers et al, 2004; Quillian, 1968) whilst retaining a close association to modality specific information. Moving towards to a closer link between modality specific and semantic information, some theories explicitly (Jackendoff, 2002; Vigliocco et al, 2004) or implicitly (Farah & McClelland, 1991; Tyler & Moss, 2001) propose that modality specific conceptual information is important for semantic content. In this case modality specific content is contained within semantic representations but it does not dictate the structure of the semantic system, therefore there is only a partial dependence. Finally we come to embodied theories. Here, sensory and motor systems are the primary force in semantic content (Barsalou, 1999; Gallese & Lakoff, 2005) or an important organizational principle (Pulvermüller, 1999) and there is a complete dependence of semantic representations on sensory and motor systems.

Abstract, amodal theories predict that there should be no direct interaction between semantic access and sensori-motor systems, since the two use distinct and separate

domains. Intermediate theories propose that interactions can occur, depending on the task and situation, but that semantic representation is still possible without access to sensori-motor information. Embodied theories predict that sensori-motor information is a necessary part of semantic representation, so interactions between the two should be consistent and prevalent. The following chapter summarises the evidence, both behavioural and neuroscientific, for embodied content in semantic representation. Some gaps in this literature will be formalized and used to motivate the experiments presented here.

Table 1-1: Summary of Early Theories according to Structure, Decomposition and Content

Theory		Structure							Decomposition	Content			
		Isomorphism with concepts			Basis of similarity between items		Organisation			Holistic	Componential / Featural	Amodal	Supra-modal
Reference	Section	None	Some	Total	External	Internal	Localist	Distributed					
Quillian (1968)	1.3.1.1			✓	✓		✓		✓		✓	✓	
Collins & Loftus (1975)	1.3.1.1			✓	✓		✓		✓		✓		
Smith, Shoben & Rips (1974)	1.3.1.2			✓		✓		✓		✓	✓?		
Levelt (1989)	1.3.1.3	✓			✓		✓		✓		✓		
Farah & McClelland (1991)	1.3.1.5			✓		✓		✓		✓		✓?	✓?

? Indicates that no clear statement was made, therefore the classification within the table is inferred

Table 1-2: Summary of Current theories according to Structure, Decomposition and Content

Theory		Structure							Decomposition		Content		
		Isomorphism with concepts			Basis of similarity between items		Organisation						
Reference	Section	None	Some	Total	External	Internal	Localist	Distributed	Holistic	Componential / Featural	Amodal	Supra-modal	Multi-Modal
Landauer & Dumais (1997)	1.3.3.1			✓	✓*		✓		✓		✓		
Jackendoff (2002)	1.3.3.2			✓		✓		✓?		✓	✓	✓†	✓
McRae, de Sa & Seidenberg (1997)	1.3.3.3.1			✓		✓		✓		✓	✓?		
Tyler & Moss (2001)	1.3.3.3.2			✓		✓		✓		✓		✓?	✓?
Rogers et al (2004)	1.3.3.3.3		✓			✓		✓	✓?	✓?	✓		
Vigliocco et al (2004)	1.3.3.3.4		✓			✓		✓	✓	✓		✓	

? Indicates that no clear statement was made, therefore the classification within the table is inferred

* Distributional information across independent, localist items is equivalent to defining similarity externally

† CS structure is amodal, SpS structure is supramodal and there are also modality specific features.

Table 1-3: Summary of Embodied theories according to Structure, Decomposition and Content

Theory		Structure							Decomposition		Content		
		Isomorphism with concepts			Basis of similarity between items		Organisation						
Reference	Section	None	Some	Total	External	Internal	Localist	Distributed	Holistic	Componential	Amodal	Supra-modal	Multi-Modal
Gallese & Lakoff (2005)	1.3.3.5.1			✓		✓		✓		✓			✓
Pulvermüller (1999)	1.3.3.5.2		✓			✓		✓	✓ ?				✓
Barsalou (1999)	1.3.3.5.3			✓		✓		✓		✓			✓
Zwaan (2004)	1.3.3.5.4		✓ ?*			✓		✓	✓ ?*				✓
Glenberg & colleagues (2000; 2003)	1.3.3.5.5			✓ ?†		✓		✓		✓ ?†			✓

? Indicates that no clear statement was made, therefore the classification within the table is inferred

* Cites Pulvermüller so is classed similarly

† Cites Barsalou so is classed similarly

2. Evidence for Embodied Theories

This chapter is a literature review of the empirical work testing embodied semantics. The review is organised by methodology. Individual sections review behavioural and neuroscientific evidence, and within behavioural evidence single word and sentence studies are separated. This is because different processes underlie single word recognition/comprehension and the integrative processes across lexical items that are involved in sentence/narrative comprehension. In addition, whereas individual words without a sentence context represent types representations (being underspecified), sentences present token representations (being specified). This point may have consequences for the way in which we interpret the evidence since it means that the semantic representation of single words is necessarily schematic and less substantive in comparison to that for sentences. Within single word and sentences, the behavioural literature is further broken up into modality and task; for example, studies that looked at language referring to motor or visual events and studies that used property verification tasks (tapping conceptual as well as semantic access). A separate section is devoted to eye-movement studies, a method with different assumptions to manual/verbal response tasks. To allow ease of exposition, the neuroscientific literature is broken up by technique (EEG, TMS, PET/fMRI) and studies with single words and sentences are not separated. This is because there were very few studies with sentence stimuli and the key finding for all neuroscientific investigations is the activation of sensory or motor areas during semantic access. Neuropsychological evidence is not reviewed here as it is beyond the scope of this chapter (and this investigation) to explore the heterogeneous findings from this discipline³. Finally, there is an increasing literature on the embodied representation of abstract concepts such as time (e.g. Boroditsky, 2000; 2001). However, this work will not be reviewed exhaustively as the current investigation focuses on language referring to concrete events. There will be coverage of abstract language only where it is investigated in conjunction with concrete objects and events or where it can be easily integrated with a concrete domain, as in the case of fictive motion⁴.

³ For a critique, see Mahon & Caramazza (2005).

⁴ Even this marginal treatment of abstract language shows that results are inconsistent (see, for example, section 2.1.2.2) and whilst it is clear how the processing of concrete language can be interpreted in relation to sensory and motor systems, this is not so for abstract language. For an alternative view, see

The majority of behavioural studies have the following structure: one independent (IV) variable manipulates the modality specific content in the presented linguistic items by varying their referents. For example, the language may refer to events or objects that share a particular sensory or motor ingredient: a type of bodily movement or action, a location in space, a sensory quality (colour, orientation, motion) or simply a modal class (e.g. visual or motor). A second IV then manipulates the task which assesses the behavioural state of the participant. Critically, this behavioural task must tap into the relevant sensory or motor systems that the language ought to engage if its content is simulated. Another paradigm that is often employed is the congruence or incongruence between the semantic and sensory/motor variables. For example, sentences refer to an event with a particular direction, towards or away from the participant. If the sentences describe an action, e.g. “You closed the drawer” or “You opened the drawer”, a task could be used requiring a motor action towards or away from the participant, e.g. pressing a button near or far to the body. If the sentences describe an object moving towards or away, e.g. “Harry threw you the ball” or “You threw Harry the ball”, a visual stimulus that depicts movements towards or away from the participant could be displayed. If the sentence and the sensory-motor variable are both ‘towards’ this is a congruent condition; if the sentence is ‘towards’ and the sensory-motor variable ‘away’, this is an incongruent condition. For neuroscientific evidence, the majority of studies look at activity in sensory and motor cortices during language comprehension or production. Embodiment provides the clear prediction that semantic access should activate the areas that are involved in the experience of the word’s or sentence’s referents. Crucially, neuroscientific studies can look for these activations without manipulating the sensory or motor domain; if sensory-motor activity is found during comprehension or production it supports the strong claim of embodied theories that activation is an obligatory part of semantic access.

The strong prediction is the direct engagement hypothesis: to achieve representation, semantic content *necessarily* and *directly* recruits the sensory and motor systems in a simulation of the on-line experience of the referents. A weaker prediction is that semantic content recruits sensory and motor systems through association, rather than simulation. Here, the recruitment of sensory-motor content may not be necessary but it may be still be direct. Interestingly, this still predicts consistent interactions between semantic and sensory-motor information. Sensory and motor information may be

recruited routinely during semantic access because of intimate ties that develop as a result of experience. However, this is a very different proposal to embodied simulation (which also has the necessity constraint). The only way to firmly assess the necessity constraint is through neuropsychological evidence or directly suppressing sensory or motor information (e.g. through TMS). Therefore, the majority of behavioural evidence actually explores the directness of the connection. This still allows us to distinguish between non-embodied associative connections, which would be open to a great deal of mediation, from embodied simulation, which would not. Therefore, the directness of the connection still allows us to distinguish somehow between stronger and weaker versions of embodiment.

2.1. *Behavioural Evidence*

2.1.1. **Behavioural evidence with single words and objects**

2.1.1.1. *The motor domain*

In these studies, words were presented that referred to particular motor actions or objects with particular affordances associated to them. Affordances are potential interactions between the body and an object, for example a cup handle on the left of a cup has the affordance that I grasp it with my left hand (Tucker & Ellis, 1998). There are a variety of tasks that measure how these words engage the motor system.

In a classic study, Tucker & Ellis (1998) presented pictures of objects, with affordances on the left or right. Participants judged whether the objects were upright or inverted, and used their left or right hand for the different response options. They found that responses were faster, and fewer errors were made, when the hand making the 'upright' response and the affordance were congruent. This supports the inference that simply seeing a picture of an object activates the motor actions associated to using it.

Richardson, Spivey & Cheung (2001) extended this paradigm into semantics by presenting participants with a rapid serial visual presentation (RSVP) of 8 pictured objects, with left or right affordances, followed by the name of an object. Participants had to judge if the named object was or was not in the sequence. In contrast to Tucker

& Ellis (1998) they found that responses were faster when the hand making the 'yes' response was opposite to affordance. Again, this supports the inference that name of the object activated a representation of the pictured object and the motor affordance that was active when it was perceived. The task used in Richardson et al (2001) implicates memory processes as the participants have to recall whether the object was presented; this was a possible explanation for why Tucker & Ellis obtained congruent facilitation, whereas Richardson et al find congruent interference. The modulation of the online motor behaviour (response hand) by the affordance associated to the object supports the automatic activation of motor actions when object pictures are perceived, but it is not clear whether the semantic content of the object name also activates that affordance, or whether affordances are only accessed through object perception. Myung, Blumstein & Sedivy (2005, Experiment 1) addressed this issue by using an auditory lexical decision task in which primes did or did not share affordances (referred to as manipulation features) with the target word. For example, a typewriter can be interacted with in a similar way to a piano; with fine pressing movements of the fingers. A typewriter does not, however, share affordances with a blanket. Lexical decisions were faster when the prime shared affordances with the target, supporting the automatic activation of motor actions upon semantic access. Siakaluk, Pexman, Aguilera, Owen & Sears (2007) provided further support for this with a lexical decision task in which target words had been rated on how easy or hard they were to physically interact with using a human body, producing a Body-Object Interaction score (BOI). The words were matched for various lexical variables so that the only manipulation was whether they had a high BOI (i.e. referred to objects that were easy to interact with, e.g. a hammer) or a low BOI (i.e. were difficult to interact with, e.g. a planet). The authors hypothesised that high BOI words would have a richer semantic representation, as the affordances provide an extra modality of semantic information. A richer semantic representation would result in faster lexical decision times as there is more feedback from semantics to orthography and phonology. Participants performed either a lexical or phonological decision task and results showed that in both tasks, decisions were faster for high BOI words as compared to low BOI words. The authors concluded that semantic representation includes information about sensori-motor experience, on the assumption that high BOI have 'more' of this information.

Presenting signed rather than spoken language, Tseng & Bergen (2005) used a response box was placed with the buttons aligned perpendicular to the body. This meant that,

depending on the button to be pressed, participants had to move their hand towards or away from their body when judging if two consecutively presented signs were the same or different. The task was completed with both native signers of American Sign Language (ASL) and non-signers. In the experimental trials, signs were presented that referred to literal towards/away actions (e.g. 'bowl' or 'catch'), metaphorical towards/away movement (e.g. 'tell' or 'yesterday'). For both these categories, the hand movement that made the form of the sign was also towards/away from the body. A critical control group was used where the hand movement was towards/away, but the referent did not have any movement; for example 'girl' where the hand moves away from the face. Signers responses were faster when the response action (towards/away) was congruent with the semantic or metaphorical signs, but not for the hand movement only signs. Non-signers did not show this effect, having similar response times for all signs. Thus, this study provides support for the automatic activity of motor information when words referring to motor actions are comprehended. However, Tseng & Bergen (2005) collapsed their analysis across signs that referred to literal movements (e.g. 'catch') or metaphorical movements (e.g. 'tell'), so it is unclear whether literal or metaphorical signs behave in the same way, as separate analyses were not conducted. The effect across both these groups may have been driven by one group alone. For experiments that use sentence comprehension, there is some evidence that literal and metaphorical/abstract language do not behave in the same way (see Section 2.1.2.2), so collapsing across groups is unwise. Clearer evidence comes from Boulenger, Roy, Paulignan, Deprez, Jeannerod & Nazir (2006) who used a lexical decision task in French. Verbs referring to hand/arm, leg or mouth/face actions, e.g. paint, cry, and nouns referring to non-manipulable concrete objects, e.g. cliff, star, were presented. Participants performed the lexical decision by reaching and grasping a cylindrical object. The thumb and forefinger were placed on a pad, a go signal was then presented on screen and participants left the pad and reached toward the cylindrical object. The onset of the movement triggered the appearance of the item for lexical decision; if the item was a word participants had to grasp the object, if the item was non-word they had to stop reaching and return to the start-pad. Parameters of the movement were recorded (reaching time and peak acceleration). Acceleration peaks were smaller and later for action verbs than for nouns (experiment 1); acceleration peaks are the result of initial muscle contractions during reaching movements so delays in the peak suggest interference. In contrast, when the item was presented as the go signal, peak acceleration were earlier for action verbs than for nouns (experiment 2). Thus, when the

lexical decision and reaching movement were made concurrently, there was interference. When the lexical decision was made before the reaching movement, there was facilitation. The early interference supports comprehension of motor words activating motor cortex, and thereby interfering with motor execution. The later facilitation also supports the same conclusion, although the authors note that it could be due to motor imagery following comprehension. Crucially, the interference appeared between 120-210 msec following movement onset, supporting the early activation of motor information during comprehension.

The evidence for motor information in lexical semantics supports the activation of 'affordances' or manipulation features associated to an object, when the object's name is heard or seen. The evidence for object names activating their associated motor manipulation is quite consistent, although studies have used tasks that implicate memory processes (Richardson, Spivey & Cheung, 2001), indirect measures of motor manipulation (Siakaluk et al, 2007) or priming between lexical items, rather than interactions between comprehension and actual motor action (Myung, Blumstein & Sedivy, 2005). However, two studies demonstrate that single-word comprehension of action verbs interacts with the motor system (Tseng & Bergen, 2005; Boulanger et al, 2006). These two studies reinforce the inference that motor information accessed during semantic processing is based in the motor system itself. However, the majority of studies support only indirect access to motor information⁵.

2.1.1.2. Visuo-Spatial Information

There are only a few studies that have looked at single word comprehension and visuo-spatial location (see Section 2.1.2.2 for similar studies with sentence comprehension). Zwaan & Yaxley (2003) presented pairs of words on which speeded similarity judgements were made. The critical items were pairs which referred to objects with a canonical spatial relation where one is above the other, for example, root-branch or floor-ceiling. The visual presentation of the words was either congruent with the canonical relation (e.g. root at the bottom and branch at the top) or incongruent (e.g. root on the top and branch on the bottom). Reaction times were faster when the presentation was congruent with the canonical relation, but not such effect was found

⁵ For a broad discussion of motor action and simulation with reference to language processing, see Fischer & Zwaan (in press).

when the words were presented horizontally and reading order (left to right) was manipulated. Along with some additional evidence which showed that this effect was limited to the right hemisphere (i.e. left visual-field presentation) which is more involved in visual imagery and therefore more likely to be involved in perceptual simulations, the results were taken to support facilitation when the actual location of the words was congruent with the simulated location of their referents.

A series of experiments with a similar motivation explored whether visual attention was similarly affected by comprehension (Estes, Verges & Barsalou, in press). In two experiments, participants were presented with a category word (e.g. cowboy) and a part word (e.g. boot or hat) that was located at the top or bottom of the object. Following the presentation of the part word, participants identified a target letter presented in the top or bottom of the visual; therefore, the location of the target letter was congruent or incongruent with the preceding part word's canonical location. They found that letter identification was slower when the target appeared in the congruent location (Experiments 1 and 2). This finding was still reliable when part words were presented alone, without the preceding context word (Experiment 3). Therefore, words referring to concrete nouns with a canonical location impair the allocation of visual attention to objects in that location. The authors discussed this result in relation to visual imagery evoked by the words. One of the earliest studies of visual imagery showed that participants were slower to detect a real image when they were actively imagining a similar image in the same location (Perky, 1910). Thus, conscious visual imagery interferes with visual perception, presumably because the two involve the same substrate (Kosslyn, 1996). Estes et al extend the arguments from conscious visual imagery to semantic access, which is a logical leap unless one assumes that either (a) conscious visual images are automatically evoked during single word comprehension or (b) the processing of semantic representations is very similar to conscious visual imagery. It is this latter possibility that is in line with embodied theories since they propose that the semantic representation of visual information is based in the visual system, as is visual imagery (Kosslyn, 1996). This is not made explicit in the paper itself and is in contrast with some embodied theories which explicitly separate themselves from visual imagery (Barsalou, 1999). In order to expand on the imagery evoked by semantic content, the authors proposed that the interference was due to dissimilarity between the target object (an X or O) and the word's referent (e.g. a hat or boot). If the presentation of the word 'hat' was followed by a picture of a hat, they

hypothesized that facilitation would result. This is inconsistent with the original finding (Perky, 1910) which showed interference between visually similar target images and conscious imagery, but it is consistent with the finding that language quickly indexes attention to referents in the environment (e.g. Chambers et al, 2002). In sum, although the results present some evidence for embodiment it is not clear exactly what can be inferred, since any explanation is inconsistent with either theory or data.

In a different manipulation of visual properties, Pecher, Zeelenberg & Raaijmakers (1998) manipulated the relationship between the prime and target in a semantic priming task according to the similarity of their referents' visual form, e.g. pizza-coin or honey-glue. In the first experiments, they also included items that were associatively related, e.g. start-end or silver-gold. In the first two experiments, only associative priming was found for lexical decisions and word naming on the targets. However, in two further experiments the lexical decision and naming tasks were preceded by a decision task in which all the items were judged on their perceptual form, i.e. whether the word referred to an oblong object or not. Under these conditions, a small perceptual form priming effect was found for word naming with faster responses when the prime referent shared visual form with the target. In a final experiment, the form-judgement task was followed by visual lexical decision, this time with the associated pairs removed. Perceptual-form priming was found for perceptually related pairs. It was concluded that perceptual priming was only found when the perceptual features of the referents had been made salient, and associative pairs did not invoke association-checking strategies that masked the perceptual-form priming. This was taken to be in line with context specific priming effects, where the current context modulates the availability of semantic features that are active for a given word.

In summary, there is limited evidence that single words activate visuo-spatial information. Object words appear to activate their canonical location, although it is unclear what form the location information takes; for example, it could be a cue to shift attention to that location or a visual simulation of the referent (or both, Estes et al, in press). For objects, the representation of a canonical location interferes with perception at that location (Estes et al, in press) and this finding agrees with the evidence from sentence comprehension (see Section 2.1.2.2). The apparent conflict between orienting towards a congruent location and impaired perception at that location is reconciled by arguments about the how consistent the visual object is with the imagined or simulated

object (Estes et al, in press), although this explanation has its own problems and it remains to be tested. If location information is a visual simulation of the referent, it is not clear how the location of the words themselves interacts with the simulated location of the referents (Zwaan & Yaxley, 2003), beyond the inference that it helps when they are 'congruent'. For other perceptual information, it appears that the modality has to be made salient before it can affect word to word priming, weakening the idea that these features are automatically accessed (but see Kellenbach, Wijers & Mulder, 2000 in section 2.2.1). Therefore, the results for visual information are opaque and do not provide clear support for visual information in the semantic representation of single words. In a similar vein to the findings for single words and the motor domain, it is not always clear what form the visual information takes. Therefore, there is no strong evidence for a direct connection between visual processes and semantic information.

2.1.1.3. Property Verification

Several studies have used property verification and it is assumed that conceptual representations are accessed for this task to be performed. In order to judge if a trees have branches I have to access what I know about trees and branches. When the stimuli are words, we assume that the semantic representation of the word has to be accessed too.

In an on-line task, concept and property words were presented in a sentence like structure and participants had to judge as quickly as possible whether the property was true of that concept, e.g. Sheet can be Clean; Leaves can be Rustling (Pecher, Zeelenberg & Barsalou, 2003). The critical manipulation was the modality of the properties, consecutive properties could either be the same modality, e.g. Blender-Loud Leaves-Rustling (both auditory) or of different modalities, e.g. Soap-Perfumed Television-Noisy (one gustatory one auditory). Reaction times were faster when the current modality was the same as the previous property rather than different. This effect was not produced by any associative relationship between consecutive properties, and was not dependent on the SOA between the concept and property names appearing. The authors concluded that this supports simulation during semantic and conceptual access, since there is a cost to changing the modal information that is simulated on successive trials, similar to the cost of attending to events in different modalities during

an online task (Spence, Nicholls & Driver, 2001). This result was replicated in Portuguese by Marques (2006) who showed that the modality switching cost was present even when category (living vs. non-living) was kept constant, but no similar switching cost was present when modality was constant but the category was switched; indicating that modal rather than category specific organisation supports conceptual representation. Extending these findings, Pecher, Zeelenberg & Barsalou (2004) used the same task but presented each concept twice, with properties from either the same or different modalities. The first presentation, the prime, could be one of two modalities, e.g. apple-shiny (visual) or apple-tart (taste). The second presentation, the target, could be the same or different to that first presentation, e.g. apple-green (visual). Reaction times to the target were faster, and less errors were made, when the same modality was used, further supporting the simulation of modality specific information. When the same modality is used, that information will be more available to simulation the second time. The lag between the first and second presentation was 12, 18, 24 or 100 trials with the facilitation present at all lags. However, pairwise comparisons showed that the effect was strong at short lags (12-18) but not for long lags (24-100), indicating that access to the modal information is mediated by recent experience as well as the current context.

In a slightly different approach, Borghi, Glenberg & Kaschak (2004) presented single words as properties to be verified for concepts referred to in a previous context sentence. The sentences referred to an action that was being performed with the object that implied a particular perspective, for example driving a car implies you are inside it; filling it with petrol implies that you are near the boot of the car. Responses were faster when the location of the property coincided with the perspective in the sentence. For example, participants responded faster to 'trunk' than 'hood' when the sentence was "You are filling a car with petrol". When the perspective was irrelevant to the location of the properties, e.g. being inside the car, no difference in reaction times was observed. In a final experiment, sentences were used that did not imply an interaction with the object, e.g. "There is a doll standing on the table". Properties were presented that were either at the top or the bottom of the object, e.g. hair versus ankle. A button box was arranged vertically so that responses were made by moving the hand up or down. When the property location was congruent with the end-point of the movement, the response time was faster. This effect was removed when no movement was required (i.e. fingers were always on the response buttons). From this series of experiments, the authors

concluded that accessing semantic/conceptual information necessarily involves simulated interaction with the objects; therefore activating motor and spatial information congruent with that interaction. Properties that were congruent with these simulations, by being in a similar spatial location, were responded to faster as the simulation made them more available.

Finally, Solomon & Barsalou (2004) presented a concept word followed by a property word and used regression analysis to look at factors that affect reaction times to make the verification. A between subjects manipulation used fillers that were false but highly associated, e.g. owl-tree, or false but not associated, e.g. owl-wheel. They found that when fillers were highly associated more variance of the reaction times was accounted for by perceptual variables, especially the size of the property relative to the concept, the property 'nose' was previously judged to comprise about 5% of a fox, whereas the property 'face' comprised 15% of the gorilla. The larger the property, the longer the reaction time; this was interpreted as larger properties taking longer to simulate.

However, when the fillers were not associated, little of the variance was accounted for by perceptual variables, suggesting that a word-association strategy (not simulation) was being employed to solve the task. This was further supported by more variance being accounted for by linguistic variables, such as frequency. Further support for the word-association strategy was found in a separate group who were told to imagine the concept and search their visual image for the property. These subjects always took longer and showed the same pattern: when fillers were not associated, less of the variance was accounted for by perceptual variables, suggesting that a word-association strategy was being employed despite the instruction. The authors concluded that simulation is only employed when deep conceptual processing is required, not when other strategies (such as word-association) can be used.

The data from property verification is quite consistent, showing effects of the perceptual modality and spatial location of the property. Property verification requires conceptual as well as semantic access and as noted in Chapter One (section 1.1.1) there are good reasons to separate conceptual and semantic knowledge. A weak criticism of this work is that the evidence from property verification supports modality specific information for conceptual representation, but it is not so easily applied to semantics (unless the two are considered isomorphic). A stronger criticism is that all property verification tasks may invoke imagery or more conscious processing than is typically required for

semantic access: participants have to make an explicit judgement. Therefore these tasks may not be representative of normal semantic processing. However, for single words, it is the best evidence for an automatic connection between semantic and sensory-motor information, and it does provide support for a somewhat direct connection.

2.1.1.4. *Summary of behavioural evidence with single words*

The evidence for the motor domain consistently shows that motor information is activated during single words comprehension, particularly if that word refers to a manipulable object. However, only two studies show direct interactions with motor actions, so most of the evidence taps motor processing indirectly. The evidence for visual/visuo-spatial information is weaker. Object words appear to activate their canonical location but it is not clear whether this is because of the simulation of the referent at a particular location or because the word creates expectations that influence visuo-spatial attention (e.g. by directing attention to a particular location). The evidence from property verification is more consistent and does support the automatic activation of modality specific information; however, it requires an explicit judgement from the participant and may reflect more effortful conceptual access rather than simple semantic activation. As noted at the beginning of this Chapter, single words present type information that is necessarily underspecified; effects may be weaker or harder to tap because less substantial semantic information is available. The evidence from single words does support the automatic activation of some modality specific information but it is often indirectly accessed and the precise nature of that information is unclear (e.g. if it is actually motor information or associated features, if it is visual information or attention). So far, we cannot draw any strong conclusions about the directness of the link between semantic and sensory-motor information.

2.1.2. Behavioural evidence with sentences

2.1.2.1. The Motor Domain

In a series of experiments, Klatzky and colleagues looked at how a symbolic cue to prepare a particular hand-shape affected sensibility judgements for sentences describing motor actions. In all studies, associations for the symbols were established during a training procedure where one of four cues was presented and participants had to make a hand shape in response to that cue: pinching with the thumb and forefinger, poking with the forefinger, a clenched fist or a flat palm. In the first set of experiments (Klatzky et al, 1989) presentation of the cue was followed by a written phrase that described possible or impossible interactions with objects, e.g. e.g. “rub your stomach” vs. “climb a newspaper”. Each sensible phrase was paired with a cue that mimicked the hand shape used in the phrase, e.g. a flat palm for “rub your stomach”, or a clenched fist for “squeeze a tomato”. Presentation of a congruent cue primed sensibility judgements for the phrase, as seen by faster reaction times. This priming was relative to a neutral baseline, i.e. a cue not associated with hand shape. In follow up studies (McClosky, Klatzky & Pellegrino, 1992) the same priming effects were found when subjects made a button press or verbalised their response for the sensibility judgement. To assess whether the cue engaged motor preparation for the hand or motor preparation generally, a dual-task method was used. The hand-shape cue was followed with equal probability by the sensibility judgement (i.e. a phrase) or a second cue to perform a motor task. This second task, manipulated between subjects, was to tap out a sequence on the button box (hand motor task) or say a sequence of four syllables (verbal motor task). In the group whose second task was motor tapping, the hand-shape cue no longer primed sensibility judgements. Thus, the possibility of preparing for the tapping task removed the effect of preparing the hand-shape. This supports a semantic system that has access to effector specific motor information that is relevant to particular referents (in this case, phrases describing certain hand actions).

In a set of experiments that looked at the role of on-line motor action during sentence comprehension, Glenberg & Kaschak (2004) presented participants with written sentences, also for sensibility judgements, that implied action towards or away from the participant, e.g. “Open the drawer”. The sentences were imperative (as above), concrete transfer (e.g. “Sarah gave you the gift”) or abstract transfer (e.g. “Sarah told you the

story”) constructions. Participants made their response using a three button box arranged linearly away from participant. Participants kept their finger on the middle button, and positive responses could be made by pressing the button nearest to them (yes-is-near) or furthest from them (yes-is-far). For all sentences describing motion towards the body, faster responses were made in the yes-is-near condition as compared to the yes-is-far. For away sentences, the yes-is-far condition produced faster responses for concrete and abstract transfer sentences, but not imperatives. The priming effect was removed when fingers were kept on the near and far buttons, so responses did not require an actual movement towards or away from the body. The authors named this the Action Sentence Compatibility effect (ACE) and concluded that motor systems are active during comprehension of ‘motor’ sentences. Borreggine & Kaschak (2006) replicated ACE effect, using a go-no go paradigm in which participants responded to auditory sentences only when sensible. Similar to the original study, responses were made on keyboard arranged perpendicular to the body, with the P key (away) and Q key (towards) as the ‘go’ response keys. Crucially, participants did not know which key to press on ‘go’ trials until a cue was presented during the sentence. In a between subjects manipulation, the cue was presented at sentence onset or after sentence offset at delays of 50, 500 and 1000msec. The ACE effect (i.e. faster responses for movements congruent with the sentence) was found for both concrete and abstract transfer sentences but only in the sentence onset group; suggesting that preparation of the response has to be concurrent with sentence comprehension to get the ACE. Taken together with the results from Klatzky and colleagues, it appears that interactions occur when motor preparation precedes or is concurrent with comprehension.

Two studies in Italian have manipulated the effector used to respond to sentences describing actions with particular effectors, e.g. the hand, foot or mouth. Buccino, Riggio, Melli, Binkofski, Gallese & Rizzolatti (2005) presented participants with auditory sentences describing hand and foot actions, e.g. “He played the piano” or “He kicked the ball” as well as abstract content sentences, e.g. “He hated the sea”. In a go-no go paradigm, participants had to respond if the sentence referred to a concrete action (i.e. hand or foot sentences) but not if it referred to an abstract event. Response effector was manipulated between subjects, with one group pressing a button with the hand and the other group using their foot. On critical trials the ‘go’ signal was presented during the verb. Reaction times showed that the group responding with their hand were slower for hand action sentences whereas the group responding with their foot were slower for

foot action sentences. Using a similar method, Scorolli & Borghi (2006) presented participants with pairs of nouns and verbs that described actions and asked them to judge whether the pairing made sense. If the pair did not make sense, they did not make a response (a go-no go paradigm). They presented the pairs visually and in blocks that comprised hand and mouth actions or hand and foot actions. For example, a mouth action would be “to suck” followed by “the sweet” and a hand action would be “to unwrap” “the sweet”. For the other block a foot action would be “to kick” and “the ball” whereas a hand action would be “to throw” and “the ball”. In this way, the manipulated object was constant and comparisons could be made for the same noun paired with a different action verb. The hand action pairs were used as a baseline, and participants responded by pressing a foot pedal or by making a verbal response. The verbal response group were significantly faster for mouth and foot actions compared to hand actions, although the difference for the foot actions was smaller. The pedal response group were significantly faster for foot versus hand actions, with no difference between mouth and hand actions. Both studies therefore show an interaction between the response effector and the effector implicated in action sentences. Buccino et al (ibid) found interference whereas Scorolli & Borghi (ibid) found facilitation, however, in Buccino et al the go-no go signal was presented during the verb whilst in Scorolli & Borghi participants made a judgement once the noun was presented (i.e. after the whole sentence had been processed). It is possible that the difference between interference and facilitation is due to whether the response action is prepared during the sentence or after it has finished. It is also likely that different tasks are important. Sensibility judgements require attention to the sentence as a whole whereas responding to concrete actions but not abstract events requires a focus on the specifics of a described event.

Other experiments have refined the interaction between motor action and comprehension further, to the point in the sentence when the motor action becomes definite. Zwaan & Taylor (2006) generated sentences which *implied* manual rotation that was clockwise, e.g. “Jane started the car”, or counter-clockwise, e.g. “Eric turned down the volume”. Sensibility judgements were made about auditory sentences by turning a horizontal knob. Half the subjects turned the knob clockwise to respond ‘sensible’, and half the subjects turned the knob counter-clockwise; the opposite direction was used for a ‘not sensible’ response. Results showed that sensible judgements were faster when the direction of manual rotation matched that in the sentence (experiment 2). Longer sentences were then generated and used in a self paced

reading task (experiment 4). This time the knob was turned smoothly in a clockwise direction to progress through the sentence, with a turn of 5° presenting the next phrase in the sentence. Sentences were split into four regions: preverb (the seven phrases preceding verb), target (the verb), postverbal region (the phrase after the verb) and the last phrase. For example, clockwise: “Before/the/big race/the driver/took out/his key/and/started/the/car” and counter-clockwise: “He/realised/that the/music/was/too loud/so he/turned down/the/volume” (slashes indicate phrase boundaries). Reaction times were faster when there was congruence between the implied rotation in the sentence and the actual rotation made in the task, but only for the target region, i.e. the effect was localised to the verb. The authors concluded that an interaction was present *after* sentence offset in Experiment 2 because the sensibility judgement required a re-simulation of the sentence which influenced the actual manual rotation, whereas in Experiment 4 the manual rotation interacted with the immediate pairing of the verb’s orthographic form to its meaning. Other experiments in the same series explored the effect of visual rotation on the comprehension of manual rotation sentences. Following an experiment in which manual rotation was shown to be slower when incongruent visual rotation was perceived (e.g. clockwise visual rotation and counter-clockwise manual rotation) (experiment 1), other experiments replaced actual manual rotation with manual rotation sentences. Participants monitored a rotating cross for a colour change (which only occurred during filler items) whilst making sensibility judgements on auditory sentences. Judgements to the manual rotation sentences were faster when the visual and implied rotations were congruent (experiment 3). In support of this, a self paced reading task showed that the verb region in manual rotation sentences was read quicker when a visual stimulus portrayed illusory motion congruent with the rotation in the sentence (i.e. clockwise or counter-clockwise). Sentences were presented centrally, phrase by phrase, with a surrounding circle of shaded ovals that created the illusion of rotating visual motion (experiment 5). In sum, these experiments demonstrate that motor as well as visual information can affect the comprehension of sentences with implied manual rotation. The authors concluded that ‘motor resonance’ was produced by the visual rotation, activating/resonating with the motor programs that would bring about the observed visual rotation. This motor resonance was shown to affect not only actual manual rotation, but also simulated manual rotation produced during sentence comprehension. See Fischer & Zwaan (in press) for an extensive discussion of motor resonance in comprehension.

In sum, there is consistent evidence for an interaction between the comprehension of sentences referring to motor actions, and the performance of actions that are congruent or incongruent with what the sentence describes. These interactions are effector/movement specific and dependent on motor preparation preceding or being concurrent with comprehension. In all but one case (Buccino et al, 2005) congruence between the implied and actual action leads to speeded responses, and it is clear that the timing of the comprehension and motor action is crucial (Borreggine & Kaschak, 2006). The results may be more consistent since sentences present specific events/actions (tokens) with an agent performing an action; therefore, the motor content is more complete and possibly produces stronger motor activity. However, as there are no control groups, facilitation in one condition is not separable from interference in the condition used for comparison. That minor criticism aside, the evidence from motor sentences is consistent in demonstrating that motor information is automatically activated during comprehension; this does support a direct connection between motor sentences and motor information.

2.1.2.2. Visuo-Spatial Information

This section reviews evidence for interactions between visuo-spatial processes and sentence comprehension. Noordzij, van der Lubbe, Neggers & Postma (2004) used sentence-picture or picture-picture verification in combination with a secondary spatial tapping or verbal suppression task. Participants were first trained on spatial tapping, tapping out the four corners of a horizontal square, or verbal suppression, repeatedly counting from one to four. For the verification task, a sentence or a picture was presented that provided the relative location of two objects (a circle, square or triangle) that was always 'to the left/right of'. For example, "The circle is to the right of the triangle" or a picture depicting that relation. The secondary task was performed during the presentation of the first stimulus. A probe picture was then presented which displayed the two objects in the correct or the opposing relation. Participants had to decide if the probe matched the preceding stimulus. Spatial tapping, but not verbal suppression, interfered with sentence-picture verification, resulting in longer reaction times. In experiment 1, when participants responded to the verification task manually, interference was only found for participants who had used a visuo-spatial strategy – translating the sentence into a pictorial format and therefore tapping visuo-spatial

working memory. Participants were categorised as using a verbal strategy if their reaction times were significantly longer for sentence-picture verification as compared to picture-picture verification. However, in experiment 2, where participants gave their verification response verbally, interference was found across all participants. This study provides support for a visuo-spatial representation that is created during comprehension, with the caveat that participants can control whether or not this representation is used.

Alternative tests of interactions between comprehension and visuo-spatial processing have used experimental sentences that describe motion upwards, downwards, towards or away from the observer. Comprehension of these sentences was paired with attention to a spatial location or to motion that was congruent or incongruent with the movement described in the sentence. Richardson, Spivey, Edelman & Naples (2001) normed 30 verbs for spatial properties. Ratings were gathered for the axis of motion, and verbs were then classed as horizontal (e.g. pull, point, want, offend) and vertical (e.g. sink, float, hope, respect). Sentences were then created from these verbs. Concrete sentences referred to actual objects moving, for example “The ship sinks in the ocean” whereas abstract sentences referred to emotional states or abstract quantities, for example “The store owner increases the price” or “The girl hopes for a pony”. In one experiment, the sentences were presented binaurally and pictures were synchronously displayed that depicted the actors in the sentence. For example, e.g. “The balloon floats through the window” was accompanied by a picture of a balloon when the word “balloon” was heard and so on. During sentence comprehension, pictures were presented centrally. In a test phase, two pictures were presented in either a horizontal or vertical arrangement and participants had to judge if those pictures had been presented together during comprehension. The concrete and abstract sentences were analysed together, and faster reaction times were observed for vertically arranged pictures for the ‘vertical’ sentences; although the horizontal sentences did not show comparative effects (Experiment 2, Richardson, Spivey, Barsalou & McRae, 2003). In addition, post-hoc analyses showed that the effect was significant for concrete, but not abstract, sentences (see footnote 2).

In another experiment (Experiment 1, *ibid*), sentences containing these verbs were presented binaurally and at sentence offset a shape appeared in the vertical (top/bottom) or horizontal (left/right) meridian of the screen; participants had to categorise the shape as a circle or square. To compare incongruent and congruent conditions, the analysis

was collapsed across top/bottom and left/right location, as well as collapsing across concrete and abstract verbs for the sentences. Results showed that when the shape location (vertical or horizontal) was congruent with the sentence, categorisation was slower than when it was incongruent; this difference was stronger for abstract sentences (see footnote 1, *ibid*). Whilst demonstrating that sentence comprehension can activate 'image schemas' that facilitate memory processes (Experiment 2) but interfere with visuo-spatial attention (Experiment 1) there are inconsistent patterns between concrete and abstract sentences. However, a recent study has replicated and extended the findings from Richardson et al (2003), clarifying the effects for concrete and abstract sentences.

Bergen, Lindsay, Matlock & Narayanan (2007) used the visuo-spatial location task, where sentence comprehension is followed by categorisation of a shape that appears on screen. Recall that the original finding was one of interference when the motion in the sentence was congruent with the location of the target, i.e. reaction times were longer when a vertical sentence was followed by a shape in the vertical meridian. Bergen et al (*in press*) improved on the previous findings by not collapsing across location or sentence type. The vertical dimension was chosen and both verbs that refer to motion (upwards or downwards) and nouns that have a canonical spatial location (high or low) were used. In the critical trials, the shape for categorisation appeared in the top or bottom of the screen; therefore being congruent or incongruent with the sentence. A series of experiments investigated whether concrete and abstract sentences containing these items would produce congruent interference as observed in Richardson et al (2003); e.g. when an 'up' sentence is heard and the shape appears at the top of the screen. Concrete sentences with verbs referring to intransitive motion, e.g. "The mule climbed" or "The chair toppled", were found to produce interference (Experiment 1) as were sentences with concrete nouns, e.g. "The ceiling cracked" or "The cellar flooded" (Experiment 2). However, sentences that referred to metaphorical motion, e.g. "The prices climbed" or abstract motion, e.g. "The percentage increased" did not produce interference (Experiments 3 & 4). Both Richardson et al (2003) and Bergen et al (*in press*) discuss the results in terms of the 'image schema' or visual imagery that the sentence creates (in line with Perky, 1910). It is confusing to implicate imagery over and above semantic space or motion, as it implies isomorphism between semantic and imaginative representations. As discussed in the section 2.1.1.2, saying that visual imagery is evoked during comprehension has two interpretations: either semantic access

and conscious visual imagery are on a continuum or visual images are always produced during semantic access. It is more parsimonious to infer that semantic representations of motion and object location, for concrete sentences, are visuo-spatial in nature, i.e. grounded in systems that support the perception of motion and location. This would put 'visual' semantic information on a continuum with conscious visual imagery. I find this confusing as it appears to conflate automatic semantic activation with more conscious, effortful processing. If this is the interpretation, then the relationship between semantic processing and conscious visual imagery needs to be defined.

In a study that used actual motion, rather than a spatial or visual approximation, Kaschak et al (2005) presented sentences aurally whilst black and white graphics were displayed. In separate experiments, grammaticality or sensibility judgements were made on sentences that described vertical (upwards or downwards) or egocentric motion (towards or away). Similarly, the graphics depicted motion in two planes; vertical motion that was upwards or downwards and egocentric motion that was towards or away from the observer. In a blocked design, several sentences were presented whilst a 30 second visual display was observed. The sentences were either congruent or incongruent with the visual motion, and analysis was collapsed into a comparison between congruent and incongruent conditions. Judgements of both kinds were slower when the visual display was congruent with the motion in the sentence. The authors concluded that the neurons involved in perceiving a particular direction of motion were less available to simulate the direction in the sentence, resulting in slower reaction times. To reconcile this with other findings that show congruent facilitation, the authors proposed that differences in the timing and integratability of the simulated and actual events were critical. If simulation of the sentence occurs at the same time as the perception of a congruent stimulus, interference will occur. However, if simulation and perception are separated in time, facilitation will occur as the perceptual system is primed for one event by the other. In addition, if the stimuli closely represent what the sentence describes (e.g. pictorially) facilitation is more likely as the two can be more easily integrated. However if the stimuli only approximates what the sentence describes, as black and white lines for motion or an abstract stimulus, interference is more likely as the two cannot be easily integrated by the perceptual systems.

In an extension of the original experiment, Kaschak, Zwaan, Aveyard, Yaxley (2006) looked at the modality of presentation for both the linguistic and motion stimuli. Beginning with the assumption that language processing is a demanding task, the authors predicted that if perceptual and linguistic items are presented concurrently but in a *different* modality both can be processed at the same time so interference will result when they are congruent. However, when perceptual and linguistic items are presented in the *same* modality they will be processed serially as the two items compete for resources. This should then produce faster responses for congruent stimuli as the same modal system is primed between each event. Sentences describing egocentric motion (towards or away) and vertical motion (up or down) were created that emphasised the auditory aspect of the event. For example, “The surfer heard the next wave crash towards him” or “The hawk screeched as it descended on its prey”. Directional auditory stimuli were created by filtering white noise. The filtering produced the impression of something moving towards or away (similar to the Doppler Effect) and upward or downward; stimuli were normed to confirm the subjective effect of the filtering. Using the same blocked paradigm as Kaschak et al (2005) a series of sentences were judged while a 30 second perceptual stimulus was presented; as before, the sentences could either be congruent or incongruent with the motion stimuli. When the sentences were presented visually (via RSVP) with the auditory motion stimuli, shorter reaction times were found when the two were incongruent. However, when both the sentences and the motion stimuli were presented aurally, shorter reaction times were found when the two were congruent. This result was confirmed for between and within subjects manipulations of the presentation modalities (although it was weaker for the visual presentation of the sentences).

Evidence for visuo-spatial qualities (object location or directional motion) is less consistent than for motor sentences, but some common findings emerge. When sentences describe the directional motion or location of objects, this interferes with the categorisation of a shape in the congruent location on a screen (Richardson et al, 2001; Bergen et al, in press), although this is only consistently the case for concrete objects and events. This supports a visuo-spatial representation built during comprehension that interferes with real-time visuo-spatial perception or allocation of attention. The main problem for these studies is the ambiguity of the visuo-spatial representation; how exactly does it interfere with attending to a particular location? The easiest interpretation is that representation is akin to visual imagery, but no studies have yet

controlled for, or explored, visual imagery during the comprehension of such sentences. For visual motion, there is one study showing that incongruence between the sentence and percept produces shorter response times for comprehension (Kaschak et al, 2005); with the argument that this is a result of the different presentation modalities of the stimuli (Kaschak et al, 2006). Visual motion provides a promising avenue for exploring non-motor embodied effects, as the perceptual systems involved in motion processing are well characterised (Blake, Sekuler & Grossman, 2004) so any interactions between perception and comprehension can be understood in terms of those perceptual systems.

2.1.2.3. Picture judgements

A series of experiments, mostly conducted by Zwaan and colleagues, have used picture judgements to explore perceptual simulation in language comprehension. All experiments present target pictures following the comprehension of sentences. These pictures are either congruent or incongruent with an object described in the preceding sentence. For example, Stanfield & Zwaan (2001) presented sentences that implied a particular orientation for the sentence's object. For example, "John put the pencil in the cup" or "John put the pencil in the drawer"; in the first sentence the pencil is vertical (standing in the cup) and in the second it is horizontal (lying in the drawer). After a sentence had been read, a line drawing was presented and participants had to judge if the pictured object had appeared in the sentence. Filler items had comprehension questions to ensure that participants read the sentences accurately. The experimental sentences described objects that were vertical or horizontal, and the pictures presented objects upright or were rotated 90 ° to present the object horizontally. Thus, the pictures were either congruent or incongruent with the implied orientation in the sentence. Responses to the picture were faster when the orientations were congruent compared to incongruent. A second series of experiments supported these results. Using the same method as the previous experiment, sentences were presented that manipulated the implied visual form of an object, rather than its orientation. For example, "The ranger saw the eagle in the sky" implies that the eagle is flying, so will have outstretched wings; conversely, "The ranger saw the eagle in its nest" implies that the eagle is sitting with its wings folded. Responses were faster, and participants made fewer errors, when the pictures were congruent with the visual form implied in the sentence, rather than incongruent (Zwaan, Stanfield & Yaxley, 2002, Experiment 1). In another experiment,

participants named the pictures rather than judge if they had been in the sentence. This removes the explicit comparison between the picture and the sentence, being a more implicit test of the congruency effect. In addition, a set of control sentences were used that did not imply a particular form, e.g. "The ranger heard the eagle in the forest". Compared to the incongruent condition, naming was faster when visual form was congruent between the sentence and picture. Naming latencies for the control sentences fell halfway between congruent and incongruent times; being marginally faster than the incongruent condition by subjects and not different from the congruent condition (ibid, Experiment 2). These results support the inference that, when reading, we represent subtle differences in visual form and orientation implied by the text; and that this representation is in a visual, rather than propositional, format.

Connell (in press) extended these findings by manipulating the implied colour of the object. As in the original method, participants judged if a pictured object had appeared in the preceding sentence. For example "John looked at the steak on his plate" implies that the steak is cooked and brown; "John looked at the steak in the butcher's window" implies that the steak is raw and red. Pictures were line drawings naturally coloured from photographs and could be congruent or incongruent with the implied colour (e.g. a red or brown steak). In contrast to previous results, responses to the pictures were slower when the colours were congruent as compared to incongruent. This result was tentatively explained by colour being a unimodal property that is represented with less stability than visual form (which can be perceived by multiple senses). Therefore, when the depicted colour is incongruent with the simulation of the sentence, it can be effectively ignored (being a unimodal, rather than multimodal, mismatch) and does not interfere with the object recognition required to perform the task. However, in the congruent condition the neurons already active in perception are additionally needed to simulate the sentence, producing interference. Despite the tenuous explanation, this data still supports the inference visual information about objects (colour, form and orientation) is perceptually simulated during text comprehension.

Finally, similar effects have also been found for auditory comprehension. Sentences were presented binaurally, with fillers followed by comprehension questions. On each trial the sentence was followed by two pictures, separated by a brief mask, and participants had to judge if the second picture was the same as the first. On critical trials

the sentences described the motion of a ball towards or away from the comprehender; for example, “The pitcher hurled the softball at you” and “You hurled the softball at the pitcher”. The pictures were of a similar ball to that described in the sentence (e.g. for the above sentences, a softball) and the second picture was slightly smaller or slightly larger than the first; thereby approximating the motion of the ball towards (larger) or away (smaller). When the motion of the ball in the sentence was congruent with the pictured motion, judgements were faster, replicating the congruency effect seen in the studies cited above (Zwaan, Madden, Yaxley, & Aveyard, 2004).

A key issue with these studies is how important explicit comparison between the pictured object and the sentence is in producing these results. The participant has to actively recall the sentence, with the implication that it may be this active recall that produces the congruency effects rather than the automatic processes involved during comprehension. However, the naming (experiment 2, Zwaan, Madden & Yaxley, 2002) and picture comparison tasks (Zwaan, Madden, Yaxley & Aveyard, 2004) argue against this conclusion, as no explicit comparison is necessary⁶. In all experiments a large number of filler items are used to disguise the experimental manipulations, lending further weight to the automatic/implicit activation of visuo-perceptual information during sentence comprehension. Picture recognition is not a simple process, requiring the perception of the picture and successful mapping to a conceptual representation. The evidence clearly supports facilitation of such processes when the picture is congruent with a particular linguistic description, and the embodied interpretation is that the simulation creates a particular visual form for the target object. It is also possible that rather than congruent facilitation, incongruent interference drives the results. Only one study has used control items (experiment 2, Zwaan, Stanfield, Yaxley, 2002), and these did not clearly differentiate between these two options. Whether the effects are due to facilitation, interference or both, the strong interpretation places the interaction between picture perception and simulation *within* the visual processing system (because simulations are modality specific). However, no evidence has yet been found for comprehension affecting early visual processing, e.g. effects that show improvements in the detection of visual information. Therefore, it may be later stages of object recognition that are influenced rather than the object representation per se. Thus, the

⁶ The naming paradigm has been successfully used elsewhere; see e.g. Kaup, Ludtke & Zwaan, 2005, for an extension of this task to the representation of negated entities and events.

perceptual systems that are affected, and whether they are facilitated or inhibited, remains approximate.

2.1.2.4. Narrative

Situation models are “mental representations of the state of affairs described in text rather than of the text itself” (p.15, Zwaan, 1999) and there is a rich literature demonstrating that the representations built during text comprehension affect the availability of particular concepts, supporting the idea that comprehension is a vicarious experience of what the text describes (Zwaan, 1999). This version of situation models, based in ‘real-world’ perceptual and motor information rather than a propositional or associative format, is in line with an embodied account in which simulation underpins comprehension. Rather than exhaustively reviewing the literature, I will present a selection of studies that demonstrate how situation models are visuo-spatially constrained. These studies show how access to information from the narrative is modulated in a manner consistent with an embodied model of the narrated events (for reviews see Zwaan, 1999 and 2004; Zwaan & Radvansky, 1998). However, there is also some evidence of interactions between narrative comprehension and sensory processes.

Visuo-spatial, but not verbal, memory load has been shown to interfere with narrative comprehension. In a between subjects manipulation, participants had to remember a low or high load letter string (4 or 6 letters) or a low or high load pattern of black dots (3 or 5 dots). These tasks taxed verbal and visuo-spatial working memory respectively. Participants were presented with the stimuli to memorise, and then read passages sentence by sentence. After they had read the passage they had to recall the memorised stimulus. The measure of comprehension was the extent of the Contradiction Effect (CE, Albrecht & O’Brien, 1993; cited in Fincher-Keifer, 2001). The CE occurs when the final sentence of a passage is not globally coherent with information presented earlier. For example, participants read a narrative about Sarah who is a vegetarian and is in a restaurant and deciding what to eat. She spends some time deciding on the wine and the starter, and then orders a beef-burger for her main course. The CE is evidenced by longer reading times for final sentences that contradict the global coherence of the narrative (a vegetarian ordering a beef-burger) as compared to consistent sentences (a

vegetarian ordering vegetable curry). The results showed that the visuo-spatial task, but not the verbal task, disrupted the CE; that is, reading times to contradictory sentences were *shorter* when the participant had to memorise a visuo-spatial layout. In addition, the CE was significant under low visuo-spatial load (i.e. longer reading times for contradictory versus consistent sentences) but not under high visuo-spatial load. These results suggest that narrative comprehension requires visuo-spatial working memory in order to construct a stable representation of the situation model. If we accept that visuo-spatial working memory recruits modality specific systems (a conclusion that is reinforced by verbal suppression not having an effect), this data supports the use of sensory capacities during comprehension.

Turning to studies which look at the availability of information in situation models, Kaup & Zwaan (2003) gave participants narratives in which a protagonist was thinking about particular target object (e.g. a dress), in the last sentence a colour term was mentioned (e.g. a *pink* dress). This last sentence mentioned the colour in an affirmative (1/2) or negative phrase (3/4); and the colour was (2/3) or was not (1/4) the colour of the target object:

- (1) Sam wished that Laura was wearing her pink dress
- (2) Sam was pleased that Laura was wearing her pink dress
- (3) Sam wished that Laura was not wearing her pink dress
- (4) Sam was pleased that Laura was not wearing her pink dress

Following the reading of the narratives, a colour term (e.g. pink) was presented as a probe and participants had to decide if the probe term had been in the story. When the probe was presented 500ms after the last sentence, participants were slower to respond following a negative sentence (3/4) (experiment 1). When the probe word was presented after 1500ms, participants were faster following sentences in which the colour term was the colour of the actual object (2/3) (experiment 2). The authors argued that after a short delay, a propositional representation was still active so negation made the colour term less accessible, but after a longer delay a situation model with the actual object has been created. The authors took this as support for a shift from a propositional to situation model where the objects and events are represented.

The effect of spatial distance in narratives has been nicely demonstrated. Participants first memorise a layout of rooms and then read sentences describing a protagonist

moving through those rooms, e.g. a janitor who has to clean particular rooms, or prepare a room for an event. Each text has a critical motion sentence that describes the protagonist moving from a source room (e.g. the hallway) to a location room (e.g. the office). This always requires the protagonist to move through another room, the path room, which is never explicitly mentioned. Following the motion sentence is the test sentence that refers to an object in one of the rooms. Reading times for the object sentence are faster when the object is in the location room (the protagonist's current location) than the path room (which the protagonist has moved through), similarly, reading times are faster when the object was in the path room as compared to the source room (where the protagonist started). (Rinck, Hähnel, Bower & Glowalla, 1997). This finding has been replicated and extended to distance in time as well as space. Rinck & Bower (2000) used the same method, but after the motion sentences (where the protagonist moves into a different room) there was one sentence describing how the protagonist had to do something in the location room (e.g. clean it up), and a second temporal sentence describing how long it took (e.g. minutes or hours). Following the temporal sentence participants were presented with two probe words that referred to an object and a room (e.g. table - office). Participants had to decide if the object was in the named room. Objects from the location room were responded to faster than objects in the path room, replicating the spatial distance effect found previously. In addition, participants took longer to verify the object location when the activity of the protagonist took hours rather than minutes. Thus, just as a greater distance reduced the accessibility of entities, so did the increased passage of time. These experiments support the idea that situation models have a format that mirrors real experience, reflecting real-world effects of distance and time. A good demonstration of how situation models reflect real world visual experience was provided by Horton & Rapp (2002). Participants read stories describing one or two protagonists and their environment (e.g. a man sitting on the steps of his house). The fourth sentence of the narrative described a target details that were visible to the protagonist (e.g. a post-box and park across the road). The two final sentences described an event that occluded a target detail (e.g. a lorry pulling up in front of his house) or did not (e.g. a bicycle pulling up). A probe question was then presented that referred to the target detail (e.g. "Was there a post-box in front of the house?"). Reaction times to this question were significantly longer following an event that occluded the target, than one that did not (experiment 1). Crucially, reaction times were not longer when the target detail was not occluded by the event (e.g. the park rather than the post-box). These results suggest that the situation model includes what

was visually perceived by the protagonist, supporting the idea that the representation is perceptually based. If an object is occluded it is less available, just as in the real world when objects are occluded and I cannot see them.

There are several other demonstrations of how situation models are constrained in a 'real world' manner (see Zwaan, 199). Nevertheless, there is only preliminary evidence that visuo-spatial processing interferes with narrative comprehension (Fincher-Keifer, 2001), so whilst it is clear that situation models can be seen as embodied simulations, it is still possible to explain the results with an associative network where nodes and links represent the relative distances (Rinck & Bower, 2000). Therefore, more evidence is needed for the engagement of sensory and motor capacities in situation models. Stronger evidence for the activation of perceptual information during comprehension comes from eye-movement studies (see section 2.1.3).

2.1.2.5. *Summary of behavioural evidence with sentences/narrative*

When taken together the evidence for the automatic activation of sensory and motor information during sentence comprehension is compelling. Sentences that refer to motor events consistently interact with motor actions; sentences that refer to visuo-spatial location affect the allocation of visuo-spatial attention; sentences that refer to particular object forms influence the processing of pictures depicted those objects; the perception of motion influences the comprehension of 'motion' sentences. This final demonstration is crucial since it is one of the few which shows how perceptual processing influences the comprehension of sentences that describe perceptual events. Whenever language is shown to influence sensory or motor processing, it is possible to explain the effects by more indirect means. One of the most immediate consequences of comprehension is that it changes attention and the contents of working memory, therefore it may be changes to these domain general processes that influence perception, rather than a direct effect influence of language on perception or action. When it is shown that perception or action affects comprehension, this is stronger evidence for a direct connection since perception (a bottom up process) is less likely to have such pervasive effects on domain general processes. For narrative comprehension, there are nice demonstrations that visuo-spatial information (i.e. objects and distances) are represented. However, some authors have accounted for these findings through amodal networks (Rinck & Bower, 2000) so they do not necessarily implicate embodied

simulation. At least when semantic information is specified (i.e. a token representation); the evidence points towards some form of embodiment and the direct activation of sensory-motor content.

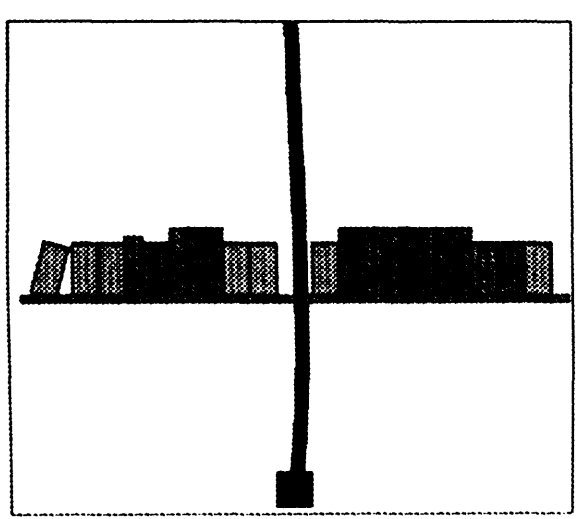
2.1.3. **Eye-Movements**

Eye movements have become an increasingly popular method for the study of language comprehension. The assumption is that eye-movements provide a measure of where attention is located and therefore what immediate object cognitive processes are focused on. You can move attention without moving the eyes, but you cannot move the eyes without moving attention. Eye movements have good temporal resolution, being able to measure the position of the eye to millisecond accuracy, and two dependent variables are typically extracted: the time spent fixating somewhere during a trial and the course of eye movements across a whole trial.

Laeng & Teodorescu (2002) demonstrated that eye-movements during visual imagery of a previously seen display were closely correlated to eye-movements during display observation. This was also supported by subjects who were told to fixate during observation keeping their eyes fixated during imagery. Eye-movements therefore reflect the on-line activity of memory and imagination, but they have also been shown to reflect the implied visual scene of narratives. Spivey & Geng (2001) corroborated the finding that saccades are made to the location of previously seen objects during recall (experiment 2); but they also presented participants with auditory scene descriptions that referred to objects and events extending upward, downward, leftward or rightward. Each story began with “Imagine..” and then described, for example, a downwards scene with the listener on top of a cliff watching some people rappel down the cliff face. There were significantly more eye-movements congruent with the implied direction of the scene as compared to that same direction for a control story with no directional bias (experiment 1). This result was later replicated when participants were not instructed to imagine the scene and had their eyes closed. Eye movements were inferred from the way the corneal bulge shaped the eye-lid and reflected a luminant spot. There were more saccades in the implied direction of a scene description than the un-implied directions (Spivey, Richardson, Tyler & Young, 2000). In concordance with these findings, eye-movements have also been shown to reflect

fictive motion as well as the literal location of objects and events. Matlock & Richardson (2004) presented participants with simple pictures that depicted a horizontally and a vertically extended object, e.g. a shelf and a chord (see Figure 2-1), whilst they looking at the pictures they listened to fictive or no-motion sentences, e.g. “The chord runs along the wall” versus “The cord is on the wall”. Participants spent significantly more time looking at the relevant object when the sentence contained fictive motion; the authors argued that this was because fictive motion is represented dynamically via simulation of real motion and eye-movements reflect this dynamic representation.

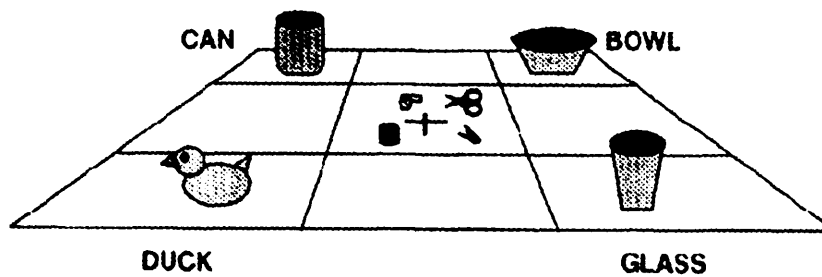
Figure 2-1: Example of visual stimuli used in Matlock & Richardson



Aside from eye-movements being used to track the motion and location of implied objects, they also provide evidence that participants map words to their referent objects quickly and in accordance with syntactic constraints. For example, Chambers, Tanenhaus, Eberhard, Filip & Carlson (2002) had participants manipulate real objects according to verbal instructions. Objects were spread out on a grid (experiment 1, Figure 2.2) or a circle (experiment 2, Figure 2.3) divided into sections; eye movements to different objects were recorded. A theme object to be moved (e.g. a whistle) was placed in the centre of the display and different goal objects (e.g. cans or bowls) were placed in the surrounding sections. In the first experiment instructions either contained the preposition ‘inside’ or ‘below’; e.g. “Put the whistle inside/below the can”. Spatial prepositions such as ‘inside’ restrict the goal objects that can be used, as they have to be able to contain things. This is in contrast to a preposition like ‘below’, where no specific properties of object are necessary. In the critical trials the target goal object

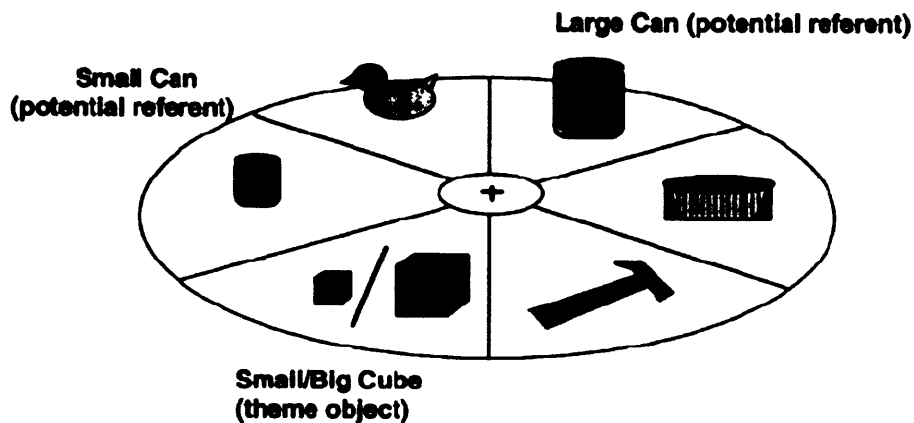
was container (e.g. a can) and 2 of the 3 distracter objects could be other containers (e.g. a bowl and glass, see Figure 2.1) or non-container objects.

Figure 2-2: Example lay out from Chambers et al (2002), Experiment 1.



In the condition when distracter objects were also containers, eye-movements to the goal object did not reliably diverge from the distracters until 300 - 400ms from the onset of the definite article for the goal object (i.e. "...inside *the* can") compared to 0-100ms when there was only one container present (the goal object). In contrast, when the preposition was "below" eye-movements for the goal object diverged at similar times in both conditions. This shows that the preposition was used to identify the target when no competitors were present. Supporting evidence of this fast indexing following comprehension was found in Experiment 2 where the size of the theme, goal and distracter objects was changed. Two possible goal objects were present (e.g. a large and a small can) and a unique distracter (e.g. bowl, see Figure 2.2). The theme object was also small or large so the instructed action was possible with one or both goal objects (e.g. a small cube can fit in both cans whereas a large cube can only fit in a large can). Sentences were of the form "Pick up the X, now put it in the/a Y" with a consistent manipulation of the definite ("*the*") or indefinite ("*a*") article for the goal object. When the sentence contained a definite article and there was one compatible goal, fixations reliably deviated from distracters at 300-400ms. However, when two compatible goals were present, fixations deviated at 400-500ms. When the sentence contained the indefinite article and there was one compatible goal, fixations deviated at 500 - 600ms, whereas with two compatible goals fixations deviated at 200-300ms. These results support the argument that that indefinite articles are compatible with multiple alternatives (therefore causing slower processing when only one option is present) whereas definite articles refer to a specific instance (therefore causing slower processing when two options are present).

Figure 2-3: Example layout for Chambers et al (2002), Experiment 2



Prepositions were used to decide which physical objects were task relevant, demonstrating that linguistic stimuli are quickly indexed to referent objects during comprehension (see also Chambers, Tanenhaus & Magnuson, 2004). These results are important as they show that eye-movements closely track the integration of linguistic stimuli and physical experience. Other studies have used this principle to explore the qualities that make distracters effectively compete with a target object.

In one such experiment, participants were presented with a display of four pictures and heard the name of a target object. They had to fixate on the picture of the target object as quickly as possible. On experimental trials, one picture shared visual features with the target object, e.g. if the target was 'snake' the distracter was a coiled rope (this was called the Strong distracter condition). The important control condition was when the original distracter and targets were reversed, e.g. participants heard 'rope' and the distracter was now the picture of a snake that had originally been the target (this was the Weak distracter condition). These conditions differentiate simple visual similarity, e.g. between pictures of a rope and a snake, and focus on how similar the distracter picture is to a prototypical category member for the target object. For example, the picture of a coiled rope is visually similar to a prototypical snake that is coiled. However, the picture of a snake with its raised head is not visually similar to a prototypical rope. Therefore there is an imbalance between the two pictures in how effectively they act as distracters. Results showed that participants fixated more on the visually related distracter than on the other pictures and that this effect was more pronounced in the strong distracter condition. In this experiment eye-movements showed that visual features of the named object were activated upon comprehension (Dahan & Tanenhaus, 2005), supporting the embodied idea that perceptual information is implicated in

semantic representation. However, given that the task was to fixate on the correct picture it is not surprising that visual similarity had such an effect. The participants had to use visual features to identify the picture, so visual features had to be extracted from the object name and therefore its referent. It is possible that perceptual or motor features are not accessed when the task does not require them.

Support for automatic activation, regardless of task, comes from a similar experiment that looked at motor rather than perceptual features. Myung, Blumstein & Sedivy (2005, Experiment 2) had presented participants with four pictures and then the name of a target object. Participants had to touch the screen location with the picture of the target as quickly as possible. Rather than sharing visual features, target and distracter items shared manipulation features. For example, a baseball can be manipulated in the same way as a grenade: both are grasped and thrown. They found that participants fixated on the distracter that shared manipulation features significantly more than control items (e.g. 'leaflet') between 500 - 800ms after word onset. Therefore, even though motor features were not necessary to successfully identify the target picture they were still active following comprehension, allowing objects with similar affordances to interfere with identification.

In summary, eye-movement data provides some support for sentences and narrative comprehension being a simulation of experience: when participants are asked to actively imagine or to simply listen to scene descriptions, their eye-movements reflect the implied location of events (Spivey & Geng, 2001; Matlock & Richardson, 2004). This is in line with an embodied interpretation where the eyes move as if those events were being observed. Eye-movements also provide evidence that perceptual and motor features of individual words are active during comprehension (Dahan & Tanenhaus, 2005; Myung, Blumstein & Sedivy, 2005). This provides support for perceptual and motor features being a part of semantic representation. Eye-movements are an increasingly useful tool to explore comprehension (see, for example, Henderson & Ferreira, 2004) and through the use of inventive methodologies, they are also supporting the role of perceptual and motor information in semantic representations. One small caveat is that eye-movements may not be a veridical mirror of the mind, directly reflecting the immediate contents of cognitive processing. The mechanisms which influence oculo-motor movements, such as attention, imagination and task-demands,

need to be very well understood before eye-movement evidence for embodiment is solid.

2.1.4. **Summary of behavioural evidence**

For the motor modality, the combined evidence supports the activation of manipulation features for objects (Myung et al, 2005; Richardson et al, 2001; Siakaluk et al, 2007) and motor procedures for actions. The evidence for actions is cogent as studies typically combine comprehension with a motor response (Glenberg & Kaschak, 2003; Borreggine & Kaschak, 2006; Boulanger et al, 2006; Buccino et al, 2005; Tseng & Bergen, 2005), thereby demonstrating an interaction between specific motor actions and comprehension. However, what applies in the motor domain may not apply to sensory modalities since efferent and afferent modes of experience may be distinct in how they respond to top-down influences (such as language). The mechanisms that control the motor system may be more open to influence since they are self-initiated and primarily controlled by top-down plans for action. In contrast, perceptual processing may be less open to manipulation and more dependent on bottom-up processes involved in stimulus perception. This is a speculative point, especially since perception and action representations appear to be ever more similar (e.g. Eckstein et al, 2007), however, this does not necessarily mean they respond to top-down influences in the same way. It is also true that the senses may respond more variably to semantic information since there is increasing evidence that the visual system routinely integrates irrelevant information in a task-dependent manner (e.g. Watanabe et al, 2001, Seitz & Watanabe, 2003); therefore, semantic information may also be integrated in a similar, variable, way.

For the visuo-spatial domain, there is evidence that semantic representations contain information about the location and motion of objects and events (Bergen et al, in press; Estes et al, in press; Kaschak et al, 2005, 2006; Matlock & Richardson, 2004; Richardson et al, 2003; Rinck & Bower, 2000; Spivey & Geng, 2001; Zwaan & Yaxley, 2003). There is also increasing evidence that perceptual information related to an object's visual form is activated during comprehension (Connell, in press; Pecher, Zeelenberg & Raaijmakers, 1998; Stanfield & Zwaan, 2001; Zwaan et al, 2002, 2004). The key problem for demonstrations involving the visual modality is that the tasks used

to assess the interaction between comprehension and perception are indirect. Categorisation at a particular location or picture judgements do not clearly define the visual processes that are involved. For example, does the visual information activated during comprehension effect early stages of visual processing, as we might expect from a strong formulation of embodiment, or are the effects the result of integration between visual and semantic information outside early visual processes. Studies that show the modulation of oculo-motor responses during comprehension (e.g. Matlock & Richardson, 2004; Spivey & Geng, 2001) are only supportive of embodiment if particular assumptions are made; i.e. that eye-movements are a necessary part of embodied representation and reflect the on-line processing of sensory-motor information present in cognitive states. So far, the clearest demonstration for this domain comes from one study which has shown that the perception of visual motion affects the comprehension of sentences referring to motion (Kaschak et al, 2005, see also 2006 for an extension). Overall, evidence for single words is weaker than that for sentences; this may be because of the nature of the representation that they activate. Single words provide only schematic, type representations of their referents so we might expect their effects to be weak or transient. In contrast, sentences (and particularly narratives) provide coherent, token representations, so we might expect a stronger influence of semantic representations (since they are more substantial and more defined) on sensory-motor processes. These differences may matter more for language that refers to a visual event, aggregating several perceptual features (e.g. colour, form, motion, brightness) in comparison to the motor domain which does not (i.e. it is primarily made up of action/motor content). Single words like 'kick' may be able to evoke sufficient motor content to influence motor action, whereas words like 'rise' may require more specification (i.e. a sentence context) to evoke sufficient visual information to influence perception.

A large variety of tasks show that sensory and motor information is routinely accessed by semantic representations. The evidence also indicates that this information is accessed automatically, supporting a direct connection between semantic and sensory-motor content. However, comprehension/production may influence many levels of cognitive processing, from domain general (e.g. attention and working memory) to low level, modality specific processes (e.g. the processing of visual signals or the execution of motor actions). Interactions may therefore result from the influence of comprehension on low level sensory-motor processing (as predicted by strong

embodiment) or they could be a more trivial result of comprehension influencing domain general processes which then influence sensory-motor systems (i.e. indirect effects). Tasks which tap specific perceptual processes will help to clarify where interactions take place. It is increasingly clear that methodological parameters, such as the timing between comprehension and sensory/motor tasks or the modality of language presentation, are critical in how interactions are manifest (e.g. Borreggine & Kaschak, 2006; Zwaan & Taylor, 2006), therefore it is no less important that sensory and motor tasks be clearly interpretable so that interactions can be nailed down to particular cognitive processes. Similarly, few behavioural studies use a control group (a notable exception being Zwaan, Stanfield & Yaxley, 2002, Experiment 2) and reaction times differences in experimental conditions cannot be compared to a baseline. It is therefore unclear whether results indicate facilitation in one condition or interference in another. This is equally important for interpreting the nature of interactions.

2.2. *Neuroscientific evidence*

There is a long standing idea in neuroscience that the structure of semantic memory for objects, and possibly actions, reflects the sensory and motor experience of those objects and events. In this view, concepts are defined by sensory and motor attributes that arise from experience, when we see, hear, touch and manipulate things in the environment. Distributed feature networks of sensory and motor attributes will be reflected in sensory and motor cortices of the brain; for example the ventral occipital cortex (fusiform gyrus) supports knowledge about object form and the lateral temporal cortex (MT) supports knowledge about object motion (Martin & Chao, 2001). Embodied theories strengthen this idea by proposing that the sensory and motor cortices utilised during real experience also support conceptual and semantic knowledge (e.g. Gallese & Lakoff, 2005), rather than, for example, knowledge being represented by areas linked to primary sensory and motor areas (Rogers et al, 2004). This section reviews evidence that language production and comprehension does activate primary sensory and motor areas, when the language refers to sensori-motor objects and events. To that end, when reviewing studies I will not report activations for ‘typical’ language areas such as the Left Inferior Frontal Gyrus which is active in naming (e.g. Grabowski, Damasio & Damasio, 1998), or Superior Temporal areas that are active during semantic processing (e.g. Damasio, Tranel, Grabowski, Adolphs, & Damasio, 2004). Instead, I will only report activity that reflects the modal distribution of information and therefore implicates sensory or motor areas. Different techniques measure neural activity in different ways, therefore the literature is sub-divided according to method: Electroencephalography (EEG), which measure electrical activity arising from neuronal firing; Transcranial Magnetic Stimulation (TMS), which directly influences cortical activity in a particular area; functional Magnetic Resonance Imaging (fMRI) and Positron Emission Tomography (PET), which measure the brain’s blood flow through magnetic resonance or radioactive markers respectively.

2.2.1. EEG

Two studies in German have used EEG to explore the representation of action verbs. Speeded lexical decisions were made to visually presented action words corresponding to arm, face and leg actions. In the first study, the C-line and FC-line electrodes were isolated as a Region of Interest (ROI) since they lie above the motor and premotor strip of the cortex respectively. Current source densities were calculated, which reflect the contribution of cortical activity close the electrode. Differences between verb categories at these sites emerged 240ms after word onset and were absent after 400ms, supporting the argument that the difference reflect early lexico-semantic processing. In comparison to Arm words, Leg words resulted in more activity (as measured by ingoing currents) at the electrode Cz (positioned on the midline), corresponding to the dorsal and medial leg area of the motor cortex. Face words produced more activity at C5, corresponding to dorso-lateral frontal sites previously identified with the face area of the motor and somatosensory cortices. The activity for Arm words may have been conflated with the manual lexical decision response that participants had to make, but increased activity was seen at premotor sites, as recorded by FC electrodes on the left hemisphere (Pulvermüller, Härle & Hummel, 2000). A following study confirmed these results. Again, action verbs related to the face/articulators (face words), hand/arms (arm words) and feet/legs (leg words) were used. Similar to the first study, CSD estimates showed topographical differences around 250ms at electrode Cz and C5. Face words produced an increased negativity at C5, and Leg words produced an increased negativity at Cz. Cortical difference maps were also constructed for each word class, by subtracting the CSD at time-window 240-260ms from each other. This confirmed the differing topography for leg (dorsal), arm (superior) and face (inferior lateral) words. The three electrodes positioned at relevant points over the motor strip were compared; Cz = leg, C5 = face (left hemisphere) and C6 = arm (right hemisphere). At Cz there was a greater negativity for leg words compared to arm words and at C5 there was a greater negativity for face as compared to arm words. The results show somatotopic differentiation of the different word types. For Current Source Densities these differences appeared early, at around 200-300ms (Pulvermüller, Hummel & Härle 2001)

Using a different approach and a different language, Dutch, Kellenbach, Wijers, Mulder (2000) presented concrete nouns referring to inanimate objects. A lexical decision task was used but target items were paired with a visual-perceptual prime, i.e. a word whose

referent shared visual-form with the target referent, or an unrelated prime. No effect on reaction times were found for visual-perceptual primes, but ERPs showed that the N400 was attenuated when targets were preceded by visual-perceptual primes as compared to unrelated primes. The N400 is an ERP component, a negative going peak approximately 400msec after stimulus onset, that is increased when two items are unrelated, and reduced (attenuated) when two items are related (e.g. Holcomb, 1993). It is thought to reflect the ease of semantic integration between items. When two semantic representations are similar they are more easily integrated, thereby reducing the amplitude of the N400 response. This data therefore supports the activation of visual-form information during semantic access, as this was the only relation between the words. The authors explained the presence of a difference in ERPs without a difference in reaction times by the sensitivity of the two measures. The N400 showed that there was an overlap between semantic representations of the items related in visual-form, whereas response times are only sensitive to the degree of overlap between *active* features. This conclusion is supported by studies in which semantic priming between perceptually related pairs was seen only after tasks which made visual form salient during comprehension (Pecher, Zeelenberg & Raaijmakers, 1998; see Section 2.1.1.2).

2.2.2. TMS

Transcranial Magnetic Stimulation studies have focused on words that refer to motor actions for two reasons. First, the somatotopic organisation of the motor strip means that selective stimulation can be applied to cortical areas that correspond to different effectors. Second, the consequences of TMS at effector specific cortical sites can be measured using Motor Evoked Potentials (MEP) at the muscles of specific effectors. For example, TMS applied to the hand area produces an MEP (an increase in electrical activity) in muscles of the hand. Modulations of the MEP response provide information about the state of the cortex when TMS was applied, tapping directly into local cortical activity. Increased MEPs indicate greater activity at the cortical site, whereas lower MEPs indicate reduced activity at the cortical site.

Pulvermüller, Hauk, Nikulin & Ilmoniemi (2005) applied TMS to effector specific sites in the primary motor cortex. If motor words activate the motor cortex then activity in the motor cortex should influence comprehension of single words referring to motor actions. Participants performed a speeded lexical decision by making brief lip

movements, so that motor activity related to manual responses did not confound results. The experimental words related to leg and arm actions, e.g. fold, beat, kick, hike. Stimuli were presented for 100ms and single pulse was applied 150ms after stimulus onset. TMS was applied to two different sites, an arm site and a leg site. There was an interaction between leg and arm words and the leg or arm TMS site. When TMS was applied to the congruent site, lexical decisions were faster. However, this was only true for the left hemisphere sites, not for right hemisphere or sham TMS sites; this supporting the lateralisation of cortical areas involved in language, even when those areas are modality specific. This study provides some of the most direct evidence that modality specific cortical areas play a role in single word processing, rather than these areas being activated as a secondary consequence of semantic access.

Further evidence that the motor cortex plays a role in semantic processing comes from production, rather than comprehension. Paired pulse TMS was used, in which a sub-threshold conditioning stimulus (CS) that does not elicit an MEP is followed by a supra-threshold test stimulus (TS) that does elicit an MEP. When the ISI between the CS and TS is low (e.g. 1 ms), this inhibits the cortical response, producing lower MEPs. A longer ISI (e.g. 10ms) facilitates the cortical response, producing higher MEPs. TMS was applied to the left motor cortex, on a site that produced the optimal MEP response in a particular muscle of the right hand. Participants were presented with nouns and verbs associated with motor actions, e.g. the axe, the pen, to kick, to rub, and control nouns and verbs without such associations, e.g. the planet, to hate. Nouns and verbs were presented in their citation form (*the X / to Y*), and 250ms after word onset a cue was presented that instructed participants to produce the item with a particular morphological inflection. For the nouns, this was the singular or plural inflection; for the verbs, this was the third person singular or plural. Paired pulse TMS was applied 500ms after stimulus onset (250ms after the production cue), and when the ISI of the pulses was long (10 ms) MEPs were greater for action associated nouns and verbs, but not the control items. No effect of grammatical class was found, supporting semantic over syntactic distinctions for representation at the motor cortex (Oliveri, Finocchiaro, Shapiro, Gangitano, Caramazza, Pascual-Leone, 2004). The same technique has been applied to sentence comprehension. Buccino, Riggio, Melli, Binkofski, Gallese & Rizzolatti (2005) asked Italian participants to listen to hand action sentences, e.g. "He turned the key", foot action sentences, e.g. "He kicked the chair" and sentences with abstract content as a control, e.g. "He liked the apple". A single TMS pulse was

presented during the verb. In two sessions, in which participants heard the same sentences, TMS was applied to the hand or foot area of the motor cortex and MEPs were recorded from the hand or foot muscles respectively. MEP amplitude was significantly smaller for the hand muscles when listening to hand action sentences, and smaller for the foot muscles when listening to foot action sentences. These results contrast to those of Oliveri et al (2004) who found increased MEPs for the production of single words with motor associations; however in that study TMS was applied after stimulus onset. When TMS is applied after comprehension, MEPs are facilitated and one can infer that motor activity caused by the sentence facilitates subsequent activations of the motor cortex. In contrast, when TMS is applied during comprehension (presumably at the same time as motor activity resulting from the sentence) the two interfere with each other, reducing MEPs.

These studies provide converging evidence that lexical items with motor associations activate motor areas of the cortex (Oliveri et al, 2004) and localised motor cortical areas corresponding to the specific effector of an action (Pulvermüller et al, 2005; Buccino et al, 2005). The timing of the TMS, early in the time-course of comprehension and production, supports the argument that modality specific activations are part of the early lexico-semantic processes. However, there is one caveat. TMS activates the site of immediate application, but it also effects well connected but cortically distant areas by the passage of activity along white-matter connections. Therefore, it is possible that the effect of TMS at these distant but connected areas is the causal factor, especially given that semantic representations may be built by associative connections to sensory and motor sites rather than directly engaging them. Semantic access may 'prime' these connections and produce the effects of TMS that we see here. For this reason, stimulation at these distant sites is a necessary control, although identification of these sites is difficult and labour intensive (requiring TMS and fMRI). Despite this concern, TMS studies provide good evidence for the direct involvement of motor areas in the representation of motor words.

2.2.3. PET/fMRI

By far the most abundant neuroscientific evidence comes from PET and fMRI studies with single words and a handful of sentence comprehension studies. Table 2-1 summarises the findings for production and Table 2-2 summarises the findings for comprehension. The relatively large number of studies is a result of comprehension and production being consistently investigated since techniques with fine spatial resolution became available. Brain regions associated to different categories of words, usually concrete objects or actions, are explored. If modality specific areas, for example sensory or motor cortices, are found to be active during semantic processing, this is support for embodied views of semantic representation. The majority of studies look at the motor cortex and the representation of motor actions or manipulable objects. However, there is some evidence for other modality specific areas, those involved in visual and sensory processing, being active for objects and actions with salient sensory qualities.

Vigliocco, Warren, Arcuili, Siri, Scott & Wise (2005) used PET to compare sensory and motor nouns and verbs, e.g. *darken*, *darkness*, *depart* and *descend*. Italian was used as this allows single word presentation for nouns and verbs, morphological marking distinguishes the two grammatical classes. Words were presented aurally in a passive comprehension task. Words with motor associations activated the premotor cortex (specifically the Left Precentral Gyrus and Central Sulcus) and sensory words activated anterior temporal areas, close to the fusiform area. There were no differences for grammatical class, supporting semantic (rather than grammatical) distinctions as the organisational principle for the lexicon. The key finding is even in passive comprehension areas are activated that correspond to sensory or motor representation. Further support for the motor cortex in motor action semantics was found when actions differing in the implicated effector were compared (Hauk, Johnsrude & Pulvermüller, 2004). Words referring to actions with the leg, face and arm were presented in a passive reading task during fMRI, the baseline was strings of hash marks matched to the words for length. After the reading task, participants were asked to move their foot, index finger and tongue so that the areas active during movement were localised. When all words were compared to the baseline, activity in the left Fusiform, Inferior Frontal Gyrus, the bidirectional Precentral Gyrus (motor cortex) and the Right Superior Posterior Middle Frontal Gyrus (premotor cortex) was found. Leg words specifically

activated the Dorsal Pre and Postcentral Gyri and the Dorsal Premotor cortex on the midline; this overlapped with the localised foot movement area in the Dorsal frontal gyrus and left Pre and Postcentral gyri. Arm words activated the Middle Frontal Gyrus bidirectionally and the left Precentral Gyrus; this overlapped with the localised finger movement area in the right Middle Frontal Gyrus and the left Precentral Gyrus. Face words activated the bidirectional Inferior-Frontal Premotor cortex; the tongue movement area was posterior to this location in the Premotor cortex. This provides good evidence for somatotopically organised semantic representation for words that refer to motor actions. These results are complemented by two studies, both in Italian, that have looked at the comprehension of action sentences. In one study, participants heard sentences from four conditions, mouth actions (“I bite an apple”), hand actions (“I grasp the knife”), leg actions (“I kick the ball”) and abstract events/states (“I appreciate sincerity”). They found activation foci in the Premotor cortex related to each of three motor sentences. These foci corresponded to areas identified with those specific effectors, i.e. mouth, hand and leg. As in other studies (e.g. Pulvermüller and colleagues, 2000, 2001, 2004) leg actions were represented dorsally, followed by hand actions and then mouth actions laterally. The authors argued that a left lateralised fronto-parieto-temporal network, similar to that for action execution and observation, was involved in the semantic representation of body action sentences (Tettamanti, Buccino, Saccuman, Gallese et al, 2005). In a study which localised effector specific areas more directly, Aziz-Zadeh, Wilson, Rizzolatti & Iacoboni (2006) had participants read sentences that described literal or metaphorical mouth, hand and foot actions. Participants then viewed actions performed with the mouth, hand and foot. For example, a literal sentence took the form “biting the peach” or “grasping the pen”. Metaphorical sentences took the form “biting off more than you can chew” or “kicking off the year”. Individual analysis for each participant identified the strongest responding voxels for action observation with each effector; these voxels were then examined for responses to the linguistic stimuli. This Voxel Of Interest (VOI) analysis showed that left hemisphere VOIs responded maximally to the congruent literal effector specific phrases. Metaphorical sentences did not show such effector specific interactions. Thus, studies with sentence stimuli reinforce the evidence with single words that show general and effector specific motor activation during the comprehension of language that refers to motor actions.

Activity in the motor cortex has also been observed during conceptual and semantic access for manipulable objects as well as motor actions. Grabowski, Damasio & Damasio (1998) contrasted picture naming for tools, animals and famous people (experiment 1) with a baseline task of verbally reporting the orientation of unknown faces. Premotor and Frontal regions were active across all three categories, but the activations for tools extended to the Inferior Frontal Sulcus and Premotor Sulcus. In another experiment participants named tool pictures and generated verbs to tool pictures. A region of interest (ROI) at the superior end of a large activation area along the Precentral sulcus was taken from the first experiment. Activity in this ROI was higher for verb generation than tool naming, but was still present for both (experiment 2). Looking only at tool objects, rather than objects and actions, Chao & Martin (2000) presented photographs for viewing (experiment 1) and silent naming (experiment 2) during fMRI. Activations were compared for tools relative to animals, faces and houses. A baseline task presented scrambled versions of the same photographs. Voxels responding to objects compared to scrambled images were analysed further for tools. For both viewing and silent naming, activity in the left Ventral Premotor and Posterior Parietal cortices were observed for tools compared to all other objects and for tools versus animals. Gerlach, Law & Paulson (2002) used PET to look at areas associated to manipulable objects per se, rather than tools. Pictures were presented and participants had to decide if the objects were natural or man made; conjunction analyses were performed for areas common to manipulable objects (vegetables, fruit and clothes) compared to non-manipulable objects (animals and buildings) and Premotor activation was found. The authors discussed these results in relation to a visuo-motor network, including the Medial Temporal area (MT), which is involved in representing goal directed actions. As we will see in other studies, MT, an area implicated in motion processing (e.g. Britten, Shadlen, Newsome & Movshon, 1993; Silvanto, Lavie & Walsh, 2005), is often observed during the semantic processing of actions and tools.

Starting with the assumption that differentiating animals relies on subtle visual features (e.g. size, form) whereas tools rely on function (i.e. how they are used); Martin, Wiggs, Ungerleider & Haxby (1996) had participants identify line drawings whilst undergoing PET. Separate scans looked at silent versus audible naming, in order to collect naming latency and error data. Tool and animal names were equated for frequency and category typicality. In comparison to nonsense objects, tools and animals activated the bidirectional Fusiform Gyri, bidirectional Medial Cerebellum, Left Anterior

insula/Inferior Frontal region (Broca's area) and the Ventral Temporal lobes. When compared to nonsense objects, animals activated the Calcarine Sulcus (Primary Visual Cortex) and tools activated the left Middle Temporal Gyrus (MT) and the left Premotor area. These activations were also present when animals and tools were directly contrasted with each other. A further study presented the objects in silhouette to rule out stimulus complexity in the visual activation for animals: activity in the Occipital cortex and MT/Premotor areas were still seen for animals and tools respectively. The authors hypothesised that the Occipital activation corresponds to making distinctions between animals based on subtle visual stimuli, the MT activation for tools is associated to patterns of visual motion associated to tool use and the Premotor activity reflects knowledge about how tools are used. Semantic "representations [are] stored close to the tissue that is active when perceiving motion and using objects" (p.651, *ibid*) and they depend on reactivation of previously acquired knowledge about physical and functional attributes that is dependent on the "intrinsic properties of the object to be identified" (p.652, *ibid*). These conclusions are embodied in flavour, although they stop short of simulation (and therefore strong embodiment). Instead, areas proximal to those used during actual experience are implicated in representation; this is in line with weak embodied theories⁷.

Turning to a multi-modal dimension of experience, space, one study has looked at naming spatial relations, as well as tools, tool actions (e.g. stir, draw) and body actions (e.g. run, crawl). In a between subjects design using PET, one group named pictures of tool actions, body actions and tools and the other group named spatial relations between concrete entities, abstract shapes and tools. The control task for the action naming group was reporting the orientation of a face whilst for the spatial relation group the same task was responded to positively or negatively ('yes' for upright faces, 'no' for upside down faces). Bidirectional activity in the MT was seen for tool actions compared to tools and body actions, as well as activity in the left MT for all actions compared to spatial relations. In comparison to the control task, Motor and Premotor activations were found for tool and body actions; however, these were removed for tool actions when contrasted to naming concrete objects. Activity in the left Supramarginal Gyrus (Parietal cortex) was found for spatial relations compared to tool naming and in the right Supramarginal Gyrus for abstract shape spatial relations compared to concrete

⁷ It is interesting to note that the neuroscientific literature has long held views that are in line with embodiment, following a somewhat separate tradition to canonical symbolic cognitive theories.

object spatial relations; this was explained as a purer reflection of spatial relation processing (Damasio, Grabowski, Tranel, Ponto, Hichwa, Damasio; 2001). Thus, there is more evidence for Motor and MT activity in tools and action naming, as well as the area most implicated in spatial processing (the Parietal lobe) being active during the naming of spatial relations. However, this task does not separate the semantic representation of spatial prepositions and phrases (e.g. 'besides', 'in front of', 'to the left of') from the perception of the spatial relations themselves, making this Parietal activation slightly ambiguous. Comprehension of such language would be a clearer demonstration of the semantic representation of space.

Returning to tools in comparison to other objects, Phillips, Noppeney, Humphreys, Price (2002) looked at the retrieval of action versus perceptual (size) knowledge for tools, fruit and vegetables. Participants were presented with words and pictures and had to make constrained decisions about actions and perceptual qualities, for example, "Can you peel this by hand?" or "Is this tool longer than a paintbrush?". The baseline task was a decision about the relative size of the picture/word to a line presented underneath. Both tasks (action and size knowledge) produced fusiform activation, reinforcing the role of this area in accessing object knowledge. Retrieving action knowledge produced activity in the left Posterior MT (LPMT) and the Cerebellum. When tools were compared to fruit, the Supplementary Motor Area (SMA) and LPMT were active. Therefore, activations for action knowledge overlapped with that for tools in the LPMT. There was a trend for action knowledge to activate the left Premotor area, but only with a low threshold, making it unreliable. Retrieval of size knowledge compared to action knowledge and the baseline produced activations in the PreSMA. There were no reliable activations for fruit alone, with activity in the left Anterior MT only when thresholds were uncorrected. These results are in contrast to Gerlach et al (2002) who found Premotor activity for fruits and vegetables, but it is possible that the contrast with tools in Phillips et al (2002) masked any motor activity for fruit. Whilst supporting Motor and MT activity for tools and tool actions, these results do not support the modality specific semantic representation of fruit. The authors hypothesised that the representation of fruit requires the conjunction and integration of various semantic features; therefore no modality specific areas were seen⁸. Martin et al (1996) provided evidence that a category whose salient properties are visual (animals) correlates with

⁸ The activity in the PreSMA and Cerebellum were explained as task effects, resulting from the retrieval and online maintenance of information

activity in the visual cortices, so it is possible that fruits present a category where no one modality of information is dominant (for a similar argument, see Rogers et al, 2004). It is self-evident that modality specific areas, e.g. the motor or visual cortex, are not the same as areas implicated in the conjunction of different sensory features, e.g. the fusiform gyrus or temporal lobes. Strong embodiment predicts the concurrent activity of separate modality specific systems (like the Motor cortex and MT for tools), rather than areas that represent the conjunction of modality specific features. These areas can be nominally considered amodal or supramodal but the bottom line is that they are *not modality specific* as embodied, multi-modal systems require (e.g. Gallese & Lakoff, 2005). It is clear that the dominance of one modal class for a given category (e.g. motor information for tool actions) plays an important role in the observed cortical topography, but this does not necessarily support embodiment or simulation. More evidence is needed for entities associated with the sensory, rather than motor, modalities.

Words referring to colour have been used to investigate sensory information and modality specific regions. Martin, Haxby, Lalonde, Wiggs, & Ungerleider (1995) presented participants with achromatic line drawings (experiment 1) or the names (experiment 2) of objects during PET. Participants produced the name of a colour word for the object, an action/verb associated with the object or the object name itself, e.g. for 'pencil' they would say 'yellow', 'write' or 'pencil'. For colour word generation bidirectional activity was found in the Fusiform area (with stronger activity in the left hemisphere), an area implicated in colour processing. The specific site of activity was slightly anterior to previous PET studies reporting on colour perception. Verb generation produced activity in left Posterior MT and the Cerebellum; MT activity was slightly anterior to previous PET studies looking at motion processing⁹. In a study which directly compared colour words to abstract form words (e.g. blonde versus square), participants passively read words from these categories, as well as a large number of fillers, during fMRI (Pulvermüller & Hauk, 2006). The baseline was a string of hash marks matched to the words in length. On the basis of previous studies, six ROIs were identified: the Fusiform Gyrus, Parahippocampal Gyrus, MT, Premotor cortex, Inferior-Frontal and Dorsolateral Prefrontal areas; these were then compared for

⁹ The lack of motor activity for actions/verbs was explained by the items not being heavily motor in nature, e.g. sleep, see, and fly. The authors hypothesized that if they had used objects whose actions involved reaching, grasping and manipulating, they would expect Motor and Parietal activation (see footnote 26, p.105).

the colour and form words. Colour words activated the Parahippocampal Gyrus; the authors interpreted this as corresponding to an area involved in visual feature conjunction as it is near to the Fusiform and higher visual association areas. They extended this argument further, hypothesising that this area binds different tones and luminances for an 'abstract' representation of colour. Form words activated the Inferior Frontal and Fusiform Gyri as well as MT. This was explained as areas that calculate the conjunction and disjunction of concrete actions related to acting on objects. Further support for the Fusiform Gyrus in the processing of object knowledge has been found when a property verification task was administered during fMRI. Kan, Barsalou, Solomon, Minor & Thompson-Schill (2003) had participants perform a concept-property verification task with pairs of words, e.g. Rose – Thorn. Participants had to respond if the property was true or false for the concept. In a between subjects manipulation, filler trials consisted of unassociated pairs (e.g. Stapler – Vegetable) making the experimental trials easy in comparison or associated pairs (e.g. Stapler-Paper) making all trials more demanding. The baseline was a letter verification task in which participants decided if a single letter was part of string. A ROI analysis was used to isolate the Fusiform, as it is implicated in object processing. Left Fusiform activity was found, but only when fillers were associated. A whole brain analysis confirmed this finding, with an area of activation slightly superior to the ROI co-ordinates. The authors concluded that when the task was easy it could be performed using only word-word associations, whereas when it was hard participants had to access conceptual-semantic information (activating the fusiform); the increase in task difficulty was confirmed by longer reaction times when fillers were associated. In sum, there is mixed evidence for Fusiform activity in the semantic representation of colour, but it does appear to be consistently active for object processing generally. Whether or not this corresponds to a modality specific or embodied representation is equivocal.

In one of the rare studies to move outside of object or action processing, Wallentin, Lund, Ostergaard, Ostergaard & Roepstorff (2005) presented Danish participants with auditory sentences that referred to fictive and factive motion. Factive motion sentences used animate subjects and motion verbs, e.g. “The man goes into the house” whereas fictive motion sentences used inanimate subjects, e.g. “The pipe goes into the house”. These were compared to sentences with static animate and inanimate subjects; for example, “The man lies in the house” and “The pipe lies in the house”. Activity for both factive and fictive motion was seen in the left Posterior MT, anterior to MT/V5

which is implicated in motion processing. This study provides the only neuroscientific test, to my knowledge, of language that specifically refers to motion.

Finally, there is limited evidence that words with emotional content are associated to the emotional circuits of the brain. Using PET, participants were presented with threat words (e.g. evil, stab, assault) and neutral words (e.g. list, candle, wash). Participants had to name the colour in which the word was presented, a modified Stroop task. Bidirectional Amygdala, left Parahippocampal and left Premotor activation was found for threat versus neutral words. The authors interpreted this as the Amygdala modulating the visual stream for threat words, resulting in enhanced semantic encoding (Isenberg, Silbersweig, Engelien, Emmerich, Malavade, Beattie, Leon & Stern, 1999). The embodied interpretation is that the Amygdala is essential for fear responses; therefore it also plays a role in the semantic representation of words that are associated with fearful situations.

2.2.4. Summary of Neuroscientific Evidence

The strength of the neuroscientific evidence for embodiment depends on modality. The premotor and motor cortices are consistently activated across studies and methods. These cortical areas are not only seen for language referring to body actions (Vigliocco et al, 2005; Pulvermüller and colleagues, 2000, 2001, 2004, 2005; Tettamanti et al, 2005; Aziz-Zadeh et al, 2006), but also for tool actions and tools/manipulable objects (Chao & Martin, 2000; Gerlach et al, 2002; Grabowski et al, 1998). For most studies, the motor activation is left lateralised, although there is some evidence that the right hemisphere is implicated for tool action generation (e.g. Damasio et al, 2001). This strongly suggests that the motor cortex plays a role in the semantic representation of objects and actions with salient motor associations. This is in line with strong and weak embodied theories where motor features could be represented via associations with the motor cortex. Simulation during comprehension is supported by effector specific manipulations (e.g. Aziz-Zadeh et al, 2006; Pulvermüller et al, 2005) which suggest that the motor cortex is selectively recruited depending on the content of the language. Alongside the motor cortex, MT activity is also repeatedly seen for body and tool actions as well as tool objects (Damasio et al, 2001; Martin et al, 1995, 1996; Phillips et al, 2002; Tettamanti et al, 2005); there is also some evidence that it is active during

comprehension of words referring to fruit or an object's form (Pulvermüller & Hauk, 2006; Phillips et al, 2002).

Table 2-1: Sensory and Motor areas found in studies using fMRI or PET and language production.

Condition	Stimulus Material	Area	Study
Action/Verb generation	Pictures & Words	L MT L Sup Temporal Gyrus	Martin et al (1995)
		R Lat cerebellum	Martin et al (1995)
Body Action	Pictures	R Motor & Premotor	Damasio et al (2001)
Tool action generation	Pictures	L Premotor	Grabowski et al (1998)
		R Motor & Premotor	Damasio et al (2001)
		L & R MT	Damasio et al (2001)
Tool Naming	Pictures	L Insula L Post Parietal	Chao & Martin (2000)
		L & R Fusiform L Inf Temporal Gyrus	Martin et al (1996)
		L Premotor	Martin et al (1996) Grabowski et al (1998) Chao & Martin (2000)
		R Med Cerebellum	Martin et al (1996)
Animal Naming	Pictures	L & R Fusiform Med Occipital Calcarine Sulcus	Martin et al (1996)
Spatial Relations	Pictures	L Supramarginal Gyrus (parietal)	Damasio et al (2001)
Colour	Pictures and Words	L & R Fusiform R Pulvinar	Martin et al (1995)

L = Left R = Right
 Post = Posterior Sup = Superior
 Lat = Lateral Med = Medial
 Inf = Inferior

Table 2-2: Sensory and Motor areas found in studies using fMRI or PET and language comprehension

Condition	Stimulus Material	Area	Study
General Sensory	Spoken words	L Ant Inf Temporal Gyrus	Vigliocco et al (2005)
Colour	Written Words	L Cerebellum L Fusiform L Parahippocampal gyrus	Pulvermüller & Hauk (2006)
Motion	Spoken Sentences	L Post MT	Wallentin et al (2005)
Size knowledge	Written Words & Pictures	Pre Supplementary Motor Area (SMA)	Phillips et al (2002)
Shape/Form	Written Words	L Precentral Gyrus L Fusiform L MT	Pulvermüller & Hauk (2006)
Object property verification	Written words	L Fusiform	Kan et al (2003)
Fruit	Written Words & Pictures	L MT	Phillips et al (2002)
Tools	Pictures	L Premotor L Post Parietal	Chao & Martin (2000)
	Written Words & Pictures	L MT R SMA	Phillips et al (2002)
Manipulable Objects	Pictures	L Premotor	Gerlach et al (2002)
Threat	Written words	L & R Amygdala	Isenberg et al (1999)

L = Left R = Right
 Post = Posterior Ant = Anterior
 Inf = Inferior

Table 2-3 continued: Sensory and Motor areas found in studies using fMRI or PET and language comprehension

Condition	Stimulus Material	Area	Study
Action knowledge	Written words & Pictures	R Cerebellum L MT	Phillips et al (2002)
Motor Action	Written words	L & R Precentral Gyrus	Hauk, Johnsrude & Pulvermüller (2004)
	Spoken words	L Central Sulcus L Precentral Gyrus	Vigliocco et al (2005)
Face	Written words	L & R Inf Frontal Gyrus	Hauk, Johnsrude & Pulvermüller (2004)
Mouth	Spoken sentences	L Inf Frontal Gyrus L Med Temporal Gyrus (MTG)	Tettamanti et al (2005)
Arm	Written words	L Middle Frontal Gyrus L Precentral Gyrus	Hauk, Johnsrude & Pulvermüller (2004)
Hand	Spoken sentences	L Motor L & R MTG L Posterior Intraparietal Sulcus (IPS) L Ant IPS	Tettamanti et al (2005)
Leg	Written sentences	L Motor	Aziz-Zadeh et al (2006)
Foot	Written words	L Pre & Post Central Gyrus R Sup Frontal Gyrus L Dorso-Med Frontal Region	Hauk, Johnsrude & Pulvermüller (2004)
	Spoken sentences	L Sup Frontal Sulcus L Ant & Posterior IPS L MTG	Tettamanti et al (2005)

L = Left **R** = Right
Med = Medial **Ant** = Anterior
Inf = Inferior **Sup** = Superior

When activity in this area is observed for tools and tool actions, it is usually explained as a reflection of knowledge about the movement of objects during their use (e.g. Phillips et al, 2002). This is also in line with accounts that propose modality specific areas (in this case, those processing visual motion) are implicated in the representation of knowledge from that modality.

Despite the fact that MT is typically understood as a motion processing area, there is only one study that has used motion sentences (both literal and fictive). However, the active area in this case was proximal, but not isomorphic, with MT (Wallentin et al, 2005). It is therefore unclear whether the MT itself is implicated in the representation of semantic motion. It is of course crucial whether the cortical areas implicated in semantic representation are isomorphic with the cortical areas involved in experience, since this is the strong version of embodied simulation when it is applied to neural structures. Aside from the motor cortices, the Amygdala is the only other brain area that supports 'modality' specific representation, being active for threat words (Isenberg et al, 1999).

Beyond body actions, tool actions and tools, the evidence is considerably less coherent. The Fusiform Gyrus is documented as playing in role in the representation of object form (e.g. Chao, Martin & Haxby, 1999; Vuilleumier, Henson, Driver & Dolan, 2002) and different areas of the fusiform have been implicated for different categories, i.e. lateral fusiform for animals and medial fusiform for tools (Martin & Chao, 2001). In the current review, fusiform activity was observed for tool and animal names relative to a nonsense object baseline (Martin et al, 1996), for conceptual access during property verification for objects (Kan et al, 2003) and for words related to form and colour (Pulvermüller & Hauk, 2006; Martin et al, 1995). For sensory words in general (e.g. *darken*, *darkness*), an area proximal to the fusiform was observed (Vigliocco et al, 2005). These results support the role of the fusiform in representing the visual attributes of known objects, and more generally this area of the cortex as involved in higher order visual association; combining features from different modalities (Vigliocco et al, 2005). As regards embodiment, fusiform activation is not that informative. It can be taken as a predominantly visual area, therefore supporting modality specificity, but its role as an area that represents objects regardless of idiosyncratic variations in appearance (e.g. Vuilleumier et al, 2002) suggests that it responds to combinations of features or attributes to provide a more abstract representation of objects. As mentioned above, higher order association areas are slightly problematic for embodiment, which predicts the concurrent activation of different modality specific areas rather than concentrated activity in one area that is connected to these modal systems. It is an open question whether heteromodal areas that combine information across modalities still constitute embodied representations, or whether they indicate a progression from modality specific to modality invariant (and ultimately modality independent) representations.

It is clear that language referring to objects and actions with a salient modality (e.g. tools or body actions) activate cortical areas involved in the experience of that modality. However, this can be taken as support for weaker versions of embodied theories that do not necessitate simulation, or full embodiment. It is always possible that sensory and motor cortices become active in a secondary manner, incidental to necessary processing in semantics; however, the evidence from TMS and EEG speaks against this conclusion (e.g. Pulvermüller et al, 2001; 2005). The isomorphism between the cortical areas used during real-world experience and semantic representation is supported for the motor cortex, but it is less clear what the literature shows for non-motor information. Again, it is worth noting that the motor cortex has a special status as an efferent area, therefore what applies there may not apply for afferent sensory areas.

2.3. *Conclusions*

As stated in the introductory paragraph, the strong prediction from embodied theories of semantic representation is the *direct engagement* hypothesis: to achieve representation, semantic content *necessarily* and *directly* recruits the sensory and motor systems used during experience. The necessity condition states that without the support of sensory and motor systems, semantic representation for concrete objects and events is impaired. The directness condition states that sensory and motor systems are engaged during semantic access without being mediated by other cognitive processes. So, what can be concluded about the necessity and direct engagement of sensory and motor systems in semantic representation?

Neuropsychological evidence is the most intuitive test of the necessity constraint, since impairments in sensory and motor systems should result in impaired semantic and conceptual knowledge (e.g. Mahon & Caramazza, 2005). Thus, most empirical work with healthy adults informs only the directness condition. However, one possible correlate of necessity is speed: the faster the access to sensory and motor information, the more likely it is to be a typical and elemental part of semantic processing (Pulvermüller, 2001). There is evidence of fast access to motor information during comprehension (i.e. around 200ms following word presentation, Boulenger et al, 2006;

Pulvermüller et al, 2000, 2001) and behavioural studies for the motor domain do support timing as a crucial element (Borreggine & Kaschak, 2006; Zwaan & Taylor, 2006). TMS studies also speak to the necessity constraint, providing evidence that direct activation of the motor system affects the comprehension of motor words (Pulvermüller et al, 2005). Outside the motor domain, the only clear behavioural evidence comes from one study which shows that the perception of motion affects the comprehension of motion sentences (Kaschak et al, 2005). As noted above, effects of perception on comprehension are less easily explained by domain general processes. Neuroscientific evidence reliably shows motor cortex activation for tools, tool actions and body actions (e.g. Tettamanti et al, 2005; Gerlach et al, 2002), but the evidence for other domains is less consistent (e.g. Vigliocco et al, 2005; Pulvermüller & Hauk, 2005). However, brain activity (particularly in fMRI/PET) is always correlational rather than causal. Sensory and motor activity could be result of the high association between particular semantic domains and particular modalities, rather than the result of direct engagement in representation.

Turning to the directness constraint, there are an increasing number of studies that demonstrate the automatic activation of sensory-motor information during semantic access; evidence comes from both behavioural and neuroscientific evidence.

Automaticity suggests directness, but unless low level processes are directly tapped (as they are in the motor domain through the manipulation of motor responses) it could still reflect domain general mediation. Sensory and motor behaviour could be modulated by these mediating mechanisms, not because of direct engagement. However, if sensory or motor activity is shown to affect comprehension, it is harder to explain away these effects by mediating processes. Such evidence is available for the motor domain (e.g. Pulvermüller et al, 2005; Glenberg & Kaschak, 2003; Zwaan & Taylor, 2006) but evidence is limited for the senses (Kaschak et al, 2005, 2006). It is also unclear how different visual processes (e.g. visuo-spatial attention or motion perception) would be influenced by the same semantic content. Non-embodied theories would predict more variability since the connections between semantic and visual domains are indirect and therefore entirely task dependent; in contrast, strong embodiment would predict some continuity since the simulation of visual information in the visual system should have consistent consequences for perception.

In sum, the motor domain provides the most coherent support for embodied cognition. As mentioned above, the motor and sensory domains are not directly comparable, so embodiment in the motor system may be the exception rather than the rule. Outside this domain, questions remain about exactly what sensory processes are being engaged by particular tasks (e.g. visuo-spatial attention or picture judgements). In order to clarify where and how interactions arise between comprehension and perception, we need to know what perceptual processes are implicated. In a similar vein, these interactions cannot be well characterised unless a control condition is present as a baseline. Neuroscientific evidence usually includes a control, due to the use of contrasts, but behavioural evidence is typically without them.

If embodied theories of semantics are to advance, the extent and nature of embodied effects has to be systematically addressed. This requires more evidence from the sensory, rather than the motor, domain, and well characterised tasks with appropriate baselines so that the location and direction of interactions can be observed. The plan for the current investigation is motivated by these conclusions, and it is outlined in the next chapter.

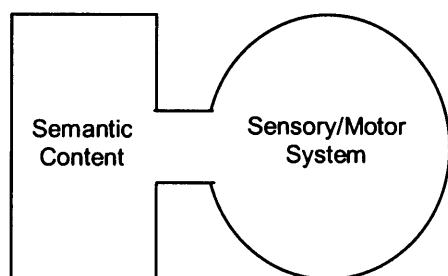
3. Structure of the Investigation

In this chapter a plan of the investigation will be provided. I will give a summary of the theoretical background and the key questions we want to address, establishing vision as the sensory domain to be investigated. General predictions from strong and weak versions of embodied theories are given and then specific questions for the domain we will be testing (visual perception). In line with the majority of behavioural work in embodiment the key manipulation is the congruence between semantic and sensory events. However, we will use a control condition so that congruence effects are always evaluated against a baseline.

3.1. *Strong versus Weak embodiment*

The strong prediction from embodied theories of semantic representation is what I will refer to as the *direct engagement* hypothesis: to achieve representation, semantic content *necessarily* and *directly* recruits the sensory and motor systems used during experience. The necessity condition states that sensory and motor systems are essential for the semantic representation for concrete objects and events. The directness condition states that sensory and motor systems are engaged during semantic access without being mediated by other cognitive processes. One important idea here is modulation; semantic representation modulates activity in sensory or motor areas because those areas simulate the experience of the referent. Since the two share a common substrate, effects should be observed bidirectionally, from language to perception/action and vice-versa. See Figure 3-1.

Figure 3-1: A schematic of direct engagement



Of the embodied theories summarised in Chapter 1, there are five which subscribe to strong embodiment (see Figure 3-3). The most extreme of these is Gallese & Lakoff (2005) for whom most (if not all) cognitive functions are carried out *within* modal

systems. Everything that is needed for representation (e.g. decomposition or abstraction) is already present in sensory-motor systems and simulations within these modal systems underpin semantic representation. Pulvermüller (1999; 2000) also begins with the neural mechanisms that might underlie cognition and proposes that Hebbian learning produces embodied content: activity related to a word form occurs alongside sensory-motor activity corresponding to the word's referent, therefore the two become associated. In this way, these sensory-motor activations can become the semantic representation for a particular word. Barsalou (1999) presents a comprehensive theory of representation-as-simulation. Here, a more traditional cognitive model is presented where representations are schematic re-enactments of sensory and motor experience. However, the central tenet is the same with simulations taking place within the sensory and motor systems themselves. Finally, Glenberg and colleagues (Glenberg & Robertson, 2000; Glenberg & Kaschak, 2003) and Zwaan (2004) refer to the theories of Barsalou (1999) and Pulvermüller (1999) respectively when fleshing out their own theories of sentence/narrative comprehension, therefore taking on the same strong assumptions about embodiment that are present in these theories. Since these theories also take in sentence level representation, simulation at all levels is proposed (single word, sentence and narrative) and details of the integration of individual words, syntactic structures and the existing context are provided. All of these theories subscribe to the following two predictions:

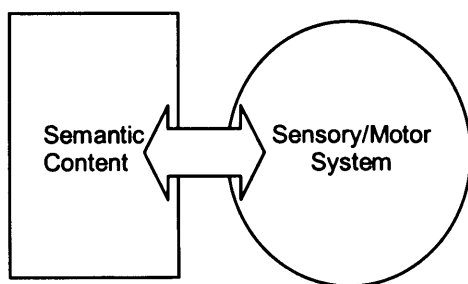
Strong Predictions:

- (i) Semantic representation engages areas in sensory and motor systems used in direct experience so low level sensory and motor systems are recruited.
- (ii) Semantic representation necessarily recruits these low level processes (modulation) so effects should be consistent across tasks.

A weaker version of embodied theories is what I will call the *indirect engagement* hypothesis. There are several possible formulations of this hypothesis, but in terms of necessity and directness it can be summarised as follows: to achieve representation, semantic content requires close contact to sensory and motor systems but it is *not dependent* on those systems. The nonessential condition states that sensory and motor systems are implicated in semantic processing because of stable associative relationships between the semantic representation for concrete objects and events and the experience of those events. However, sensory and motor content is not necessary

for semantic representation (at least once semantic representations are stable). The indirect condition states that sensory and motor systems are engaged during semantic access in a task dependent manner, being mediated by cognitive processes such as attention or perceptual learning. An important idea here is mediation: the impact of semantic representation is equivalent to an external system influencing activity in sensory or motor areas. Mediation means that bidirectional effects will not always be present since the connection between semantic and sensory-motor systems is variable. See Figure 3-2.

Figure 3-2: A schematic of indirect engagement



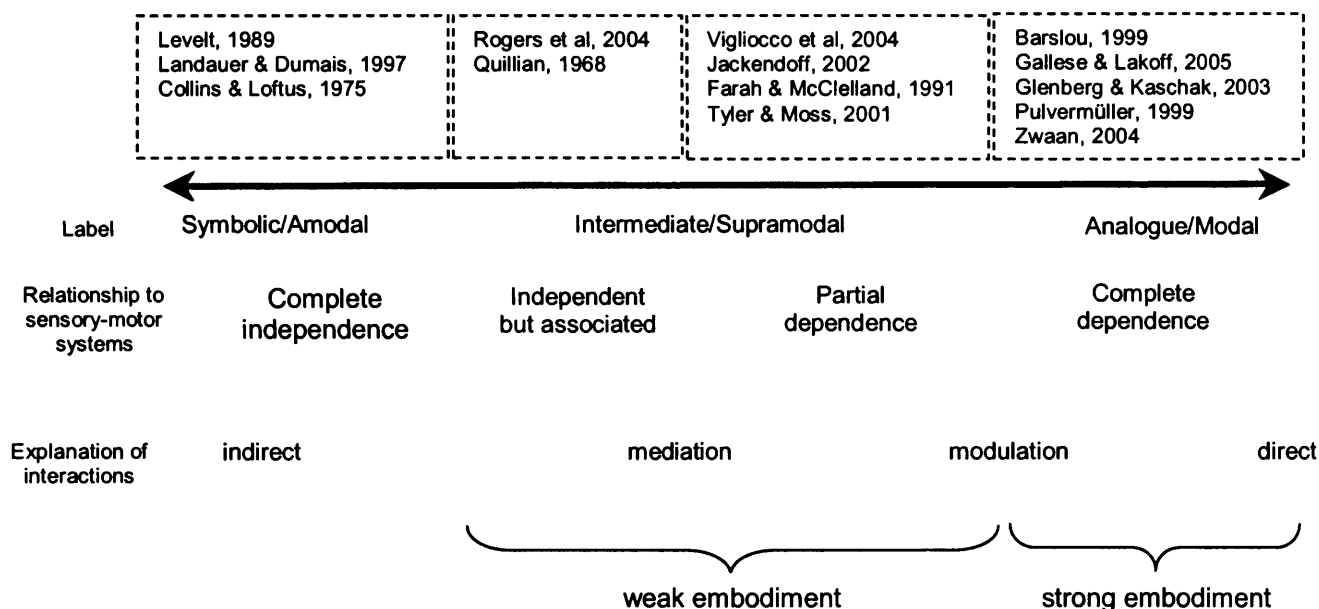
There are a number of theories that adopt (or could adopt) some weak version of embodiment. Vigliocco et al (2004) make a clear statement about the modal content of conceptual and therefore semantic representation; semantic features which correspond to modal information are linked to modality specific systems. Jackendoff (2002) also proposes that modality specific features are grounded in their respective modal systems whilst maintaining that much of language processing is based in an abstract, amodal conceptual structure. Other featural theories could subscribe to embodiment if their 'visual' and 'functional' features were grounded in the visual and motor system respectively (Farah & McClelland, 1991; Tyler & Moss, 2001). These theories propose partial dependence; although the precise mechanism is ambivalent between modulation and a strong form of mediation (see Figure 3-3). One further step away from embodiment are theories which propose an amodal, abstract semantics with associations to sensory-motor content (Rogers et al, 2004; Quillian, 1968). Rogers et al (2004) are explicit that semantic representations do not carry any content at all, but act as links to the relevant conceptual information. Quillian (1968) is included here as he does make a brief reference to a common representational level between semantic and perceptual information, but his theory is also one the paradigmatic amodal semantic networks. These theories propose an independent but associative relationship where mediation is the only mechanism by which semantic and sensory-motor content can interact. For

example, modality specific information could be recruited by association areas that integrate the modal information and therefore have access to it. All of these theories subscribe to the weak predictions:

Weak Predictions:

- (i) Semantic representation engages areas in sensory and motor systems but low level sensory and motor systems are not necessarily recruited.
- (ii) The effects of semantic representation are mediated so effects will vary across tasks.

Figure 3-3: Schematic of where theories lie along the continuum from amodal to modal
 * *McRae, de Sa & Seidenberg and Smith, Shoben & Rips not included as no clear assertions are made*



Finally there are theories which propose a completely independent semantic store (Levelt, 1989; Landauer & Dumais, 1997; Collins & Loftus, 1975). The link between sensory-motor and semantic content is outside the semantic system, produced by designation processes (see section 1.2.1). Here, interactions would be explained via indirect mechanisms (coming via other cognitive processes such as working memory or attention) or produced by the connection between semantic representations and the level at which designation occurs (i.e. a theoretically opaque process). Figure 3-3 summarises where all these theories lie on the continuum from modal to amodal and which fall under weak or strong embodiment. it is an open question whether sensory

and motor information is implicated in semantics because of strong embodiment (i.e. direct engagement that modulates activity in sensory-motor systems) or weak embodiment (i.e. indirect engagement that mediates activity in sensory-motor systems).

3.2. *Questions to be addressed*

The investigation will focus on the recruitment of visual perception during language comprehension; this was chosen for a number of reasons. The evidence for embodied effects on visual processing is less coherent than that for the motor domain even though embodied simulation implicates all modalities. As briefly mentioned in Chapter 2 (section 2.1.4), we control our actions through the motor system so it may be more open to influence from language or other cognitive processes that we use to guide action; it is an efferent system. When comprehension is combined with the motor system, this places one system which typically responds to the environment (motor actions) with some information from that environment (language). In contrast, perceptual processing is afferent and tuned to the reception of information, the effect of language may therefore be more indirect (i.e. typically exploited to guide attention and action rather than directly modulate what is perceived). When comprehension is combined with perception, this places two afferent systems together. However, recent evidence suggests a close relationship between perception and action (e.g. Fischer & Zwaan, in press); a fuzzy boundary between these two systems means that both may be equally malleable since they are already integrated. Turning to the available literature, a variety of visual processes are implicated in experiments that use, for example, picture judgements (e.g. low level signal processing, visual object recognition, concept retrieval) so the main aim of this investigation is to isolate specific, low level visual processes. By localising a particular level of visual processing we can make clearer inferences about where semantics and vision interact. There is a large literature on visual perception, particularly as it relates to motion processing (e.g. Blake, Sekuler, & Grossman, 2004). Therefore when specific visual processes are identified through established perceptual tasks (e.g. motion detection with random dot kinematograms, *ibid*) interactions can be interpreted within the context of what is known about those processes. Finally, the visual domain allows us to see if interactions are similar for different processes when the same linguistic stimuli are presented. This will inform both where interactions take place and hopefully refine the quality of those interactions.

For example, if we find similar effects across different processes, the influence of language will appear pervasive and more global in character. This supports a more general influence of language on perception and goes against strong embodiment. On the other hand, if different processes in the same domain show different effects, the influence of language will appear localised and more specific. This supports a subtle but close integration between semantic content and particular visual processes. We will look at visuo-spatial attention (which we can motivate from Richardson et al, 2003) and motion perception (for which there is limited, but suggestive, evidence, Kaschak et al, 2005).

3.2.1. **Question 1: Are embodied effects present for single words and sentences?**

The literature on visual processing focuses on sentence comprehension, it is not clear whether single words can produce similar effects to those seen for sentences. Single words without a sentence context present schematic information (a type representation) whereas sentences specify a full event, giving more constrained and specific information (a token representation). Any differences between the two might be due to this difference in the quality of semantic content that is accessed. Embodied theories predict that semantic representation is based in simulation; therefore both single words and sentences have embodied content. The difference between type and token representation is articulated in Zwaan (2004), where the embodied content of single words is more general (a number of different simulated webs are activated) and is constrained only when a sentence context allows (one web is selected). Since the visual domain combines several perceptual features (e.g. colour, form, motion, brightness) in comparison to the motor domain which does not (i.e. it is primarily made up of action/motor content), single words like 'kick' may be able to evoke sufficient motor content to influence motor action, whereas words like 'rise' may require more specification (i.e. a sentence context) to evoke sufficient visual information to influence perception. An alternative possibility is that the specific information in sentences constrains the interpretation to such a degree that broader influences are lost: a sentence may shift the focus onto a particular dimension and away from the perceptual property of interest. For example, the sentence "The man climbed the tower" may focus more on

the tower, and less on the upwards motion involved in climbing. In this case, single words would be expected to show more consistent effects than sentences, particularly if they are presented in semantically coherent blocks (e.g. 20 'up' words presented sequentially). We will use verbs that refer to upwards and downwards motion as the single word stimuli and sentences derived from those verbs. Therefore the single words and sentence refer to the same target event (vertical motion) and effects across them can be compared.

3.2.2 Question 2: Are low level visual processes implicated in interactions between comprehension and perception?

As we saw in the previous chapter, the motor domain produces the strongest converging evidence for embodiment in language comprehension. The comprehension of sentences referring to visual percepts has been shown to affect visuo-spatial attention (e.g. Richardson et al, 2003) and picture judgements (e.g. Zwaan, Madden, Yaxley & Aveyard, 2004). The perception of motion has been shown to affect the comprehension of sentences referring to motion events (Kaschak et al, 2005). These studies implicate some level of visual processing but it is not clear whether the effects are due to low level visual processes (i.e. visual signal processing) or effects at later stages (i.e. decision and response processing). By localising visual processes involved in motion perception we can isolate low level visual perception (i.e. visual signal processing).

3.2.3 Question 3: Are visuo-spatial attention and motion perception affected in the same way by comprehension?

When different visual processes (i.e. visuo-spatial attention and motion perception) are combined with comprehension, they appear to be affected in different ways. For example, for visuo-spatial attention, detection of a target object is slower when the semantic and perceptual stimuli are congruent (e.g. a sentence referring to upwards motion and an object at the top of the visual field). In contrast, pictures that approximate the motion of an object are judged more quickly under congruent conditions (Zwaan et al, 2004). This suggests motion processing and visuo-spatial

attention are affected in different ways. However, this difference is evident across different visual manipulations (top and bottom screen location, apparent motion towards or away from the observer) and different linguistic stimuli. A better test of this difference between visual processes is to use the same linguistic stimuli in combination with comparable visual manipulations (e.g. top and bottom screen location, upwards and downwards motion).

3.2.4. **Question 4: For visual motion, are comparable effects found for comprehension on perception and perception on comprehension?**

As summarised above, there is evidence that the perception of motion interferes with the comprehension of motion sentences when the two are in the same direction (Kaschak et al, 2005). There is evidence that the comprehension of motion sentences facilitates judgements of pictured motion when the two are in the same direction (Zwaan et al, 2004). However, dynamic motion stimuli have only been used to explore the effect of perception on comprehension (Kaschak et al, 2005) so it is an open question whether comprehension will impede or facilitate the perception of dynamic motion stimuli.

In the final chapter these four questions will be restated and discussed, alongside the general predictions from strong and weak embodiment. Table 3-1 details the structure of the experiments exploring comprehension on visuo-spatial attention and motion perception and Table 3-2 details the experiments exploring the bidirectional influence of comprehension on perception and perception on comprehension. Within the cells of each table the number of the experiments addressing this comparison is given.

Table 3-1: The experiments which address comprehension on different visual processes

Linguistic Stimuli	Visual Process	
	<i>Visuo-spatial attention</i>	<i>Motion Perception</i>
<i>Single Words</i>	5.1 & 5.2	6.1a, 6.2, 6.3, 6.5 & 7.1
<i>Sentences</i>	5.3 & 5.4	6.1b & 6.4

Table 3-2: The experiments which address comprehension on motion perception and motion perception on comprehension

Linguistic Stimuli	Direction of Influence	
	<i>Comprehension on Motion Perception</i>	<i>Motion Perception on Comprehension</i>
<i>Single Words</i>	6.1a, 6.2, 6.3, 6.5 & 7.1	8.1a & 8.3
<i>Sentences</i>	6.1b & 6.4	8.1b & 8.2

3.3. *Visual processes to be manipulated*

As noted above, by exploring different processes we can refine what we know about how language and visual perception interact.

3.3.1. **Visuo-spatial Attention**

There is some consistent evidence that visuo-spatial attention is influenced by the comprehension of complete motion events. Categorisation of a shape in the top or bottom of the visual field is slower when participants comprehend a sentence that refers to motion in the same direction as the target location. In contrast with initial results for abstract sentences (Richardson et al, 2003) other experiments have shown that motion events have to be concrete in order to influence visuo-spatial attention (Bergen et al, in press). The first set of experiments will explore how strong a linguistic context is needed to affect visuo-spatial attention. The current evidence suggests that only sentences referring to concrete events influence the allocation of attention, this suggests that a substantial linguistic context is needed. We will explore the amount of linguistic context by looking at both single word and sentence comprehension and manipulating the presentation of linguistic stimuli in relation to the presentation of target visual stimuli. Specific hypotheses are given in Chapter 5.

3.3.2. Motion Perception

For words and sentences that refer to motion, dynamic motion stimuli provide a better match between semantic and perceptual content. There is evidence that motion perception influences the comprehension of motion sentences (Kaschak et al, 2005) and that the comprehension of motion sentences influences the perception of pictures depicting motion (Zwaan et al, 2004). However, low level motion perception has not been manipulated (for example, in motion detection tasks). The visual perception of motion has been extensively investigated (e.g. Blake, Sekuler & Grossman, 2004) therefore we can carefully manipulate visual processing as well as comprehension. For that reason interactions between comprehension and perception will be explored from both directions. Experiments are designed to look at the influence of comprehension and perception and vice-versa. As for visuo-spatial attention, both single words and sentences will be used. Specific hypotheses are given in Chapters 6, 7 and 8.

3.4. Experimental Design

The bulk of evidence for the embodied content of language exploits the congruence between linguistic and perceptual/motor events. All experiments in this investigation will follow the same design (see Table 3-3 for a schematic of the design). However, unlike the majority of studies we will also use a set of control words and sentences to act as a baseline. In this way, effects of congruence can be interpreted as facilitation or interference (or no change) relative to the control condition. This will further clarify the direction of effects and whether or not these effects are similar across experiments.

Table 3-3: Summary of Experimental Design

Visual Stimuli	Linguistic Stimuli		
	<i>Up</i>	<i>Down</i>	<i>Control</i>
<i>Top/Upwards</i>	Match	Mismatch	Control
<i>Bottom/Downwards</i>	Mismatch	Match	Control

Experimental items are selected that refer to upwards or downwards motion. The visual stimuli will be manipulated along the same axis; presenting stimuli in the top or bottom

of the visual field (visuo-spatial attention) or upwards/downwards motion (motion perception). When combined with the linguistic stimuli this produces Match (congruent), Mismatch (incongruent) and Control (control) conditions. The Match condition occurs when Up linguistic stimuli are presented with visual stimuli at the top of the visual field or stimuli moving upwards and when Down linguistic stimuli are presented with visual stimuli at the bottom of the visual field or stimuli moving downwards. The Mismatch condition occurs when Up linguistic stimuli are presented with visual stimuli at the bottom of the visual field or stimuli moving downwards and when Down linguistic stimuli are presented with visual stimuli at the top of the visual field of stimuli moving upwards (see Table 3-3). The Control condition occurs when control linguistic stimuli are presented with any visual stimuli. Chapter 4 details the creation of linguistic stimuli to fit these conditions.

3.4.1. **Statistical Conventions**

Repeated measures ANOVAs are used to compare results by Subjects (F1) and mixed effects ANOVAs are used to compare results by Items (F2). In the by Items analysis, Sentence/Word Category (Match, Mismatch, Control) is manipulated as a between subjects variable (i.e. between items). All effects are significant at the $p < 0.05$ level, with exact p values reported for marginal effects. The critical test is whether congruency effects will appear in comparisons between levels in the factor of Word/Sentence Category, therefore, planned comparisons between each level (Match, Mismatch and Control) will be performed. These planned comparisons will use the $p < 0.05$ significance value. As an additional measure of effect size, partial eta-squared ($p-\eta^2$) values will be provided for each F-test. This details the percentage of variance accounted for by a particular factor, independently of other factors (i.e. as if it was the only variable) in the comparison. Raw error scores will be analysed using appropriate non-parametric tests (i.e. repeated measures by Subjects and 'between subjects' by Items).

4. Item Norming, Selection and Generation

4.1. *Rebus Norming*

The aim of this initial large scale study was to establish sets of verbs with consistent semantic properties that could then be used in on-line behavioural experiments. The conceptual properties we chose to focus on were movement, specifically the axis (e.g. vertical or horizontal) and direction of motion (e.g. upwards or downwards) in the event. The items reported in this section are those selected for use in the experiments, rather than arbitrary examples.

4.1.1. **Participants**

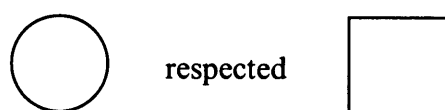
A total of 96 native English speakers participated, 32 rated each list. 2 participants had to be excluded for mistakenly filling out the booklets.

4.1.2. **Method**

4.1.2.1. *Experimental materials*

The method is adapted from Richardson, Spivey, Edelman & Naples (2001), in which norming was completed using a forced choice task, with verbs placed within ‘rebus’ sentences. Rebus sentences contain shapes as the subject and object of the sentence, with a written verb between the two (see Figure 4-1).

Figure 4-1: Example of a Rebus Sentence



We adapted this forced choice task so that the sentences presented for judgement took the form “ ‘circle’ *verbs* ‘square’ ”; it was decided to use words and not rebus sentences

as in Richardson et al. (2001) in case this produced a bias in picture selection. In rebus sentences the circle is always left most as it is in the subject position of the sentence, and the square right most as it is in the object position; the pictures that were presented for selection balanced left and right positioning of the circle (subject) and square (object) so could have been susceptible to a bias produced by presenting rebus sentences. Present tense was chosen to make the activity involved in the verb more salient. Verbs that would require a preposition to accommodate an object were presented in intransitive form (e.g. “ ‘circle’ *runs* ” rather than “ ‘circle’ *runs to* ‘square’ ”). This was done as prepositions such as ‘to’ and ‘at’ communicate a lot of spatial information which would be confounded with the information given in the verb alone.

Participants had to choose between several options but they were allowed to express a preference. Thus, if no picture was preferred, participants would select more pictures with a random order of preference, whereas if one picture was strongly preferred, only it would be selected with the highest order of preference. Sixteen pictures were generated, eight of which corresponded to transitive events (a circle and a square) and eight to intransitive events (just a circle). The pictures incorporated two axes, vertical and horizontal, and two directions for the event to occur, towards or away from the subject/circle. The pictures also balanced the positioning of the circle and square (left, right, top and bottom). For each sentence, eight pictures were presented for the participant to choose from. Transitive sentences were presented with ‘transitive’ pictures and intransitive sentences were presented with ‘intransitive’ pictures. The pictures were presented in two rows of four, each row had pictures with either a vertical or horizontal axis. There were two possible orders of presentation for the pictures within each row, to control for order effects. In the top left hand corner of each picture was a small box which the participants were instructed to use when marking their choice (see Appendix 1a).

A list of 281 verbs was compiled, mostly taken from Levin (1993). Verbs were selected for their probable spatial qualities (e.g. verbs of inherent direction such as *to leave*, *to fall* and calibratable change of state verbs such as *to decline*, *to surge*) or as controls with no probable spatial qualities (e.g. *to finish*, *to know* and *to laugh*).

It was decided that 14 of the verbs would be judged in both their transitive and intransitive form, this raised the total number of verbs to be normed to 295. These 14 verbs were all equally usable in a transitive or intransitive form (e.g. *to decrease*, *to*

drop), and it was therefore necessary to have both possibilities realised in the norming so that future use would not be constrained (see Appendix 1b for a complete list).

The verbs were then quasi-randomly divided up into 3 lists (two lists of 98 and one of 99) such that each verb only appeared once in a list. It was felt that these smaller lists would be more manageable for participants to norm.

4.1.2.2. Apparatus

A5 booklets were created for each participant. Each page had a verb placed in a sentence (outlined above) below which were eight pictures. Participants filled in their choices of pictures using pencils/pens (see Appendix 1a).

4.1.2.3. Design

A quasi within subjects design was used as each participant normed a third of the total set of verbs. The order of verbs in the booklet was randomised for each participant. The independent variable was the verb being judged and the dependent variables were the consistent elements across the choice of pictures (axis of motion, direction of motion, motion towards or away from the circle/square).

4.1.2.4. Procedure

Participants were given an instruction sheet that described the stimuli they were going to see. They were instructed to choose a minimum of one picture for each sentence and a maximum of however many they felt were applicable for the event described in the sentence. Participants were also instructed to give their choices an order of preference, with 1 being the most preferred, 2 the next and so on. It was decided that order of preference would be enforced so that agreement across participants could be better assessed. Participants were also told that it was a subjective judgement task and that responses should be based on whatever they felt was correct. Once it was clear that the participant understood the nature of the task, they were given the booklets to fill in (see Appendix 1a for complete instructions).

4.1.2.5.

Analysis

The choices were scored by weighting the first choices (as established by order of preferences) heavier than the last. Points were allocated to each picture on the basis of how many pictures each participant had chosen over all, and the order of preference given to each picture choice. Points were allocated in descending magnitude with the order of preference. For the scoring system it was only necessary to stay faithful to the order of preference, so the number of points allocated is essentially arbitrary. We used the following logic: For each sentence, there was a minimum choice of one picture and a maximum choice of eight pictures. Initial surveys of the data showed that most participants chose somewhere between one and four pictures, for this reason the scoring system was based around four picture choices and modified accordingly. Thus, for one to four picture choices, the first choice is given 4 points, and the rest are scored in descending magnitude such that the fourth choice receives one point. For more than four picture choices, the initially preferred pictures are scored more equally. See Table 1 for the scoring system.

The advantage of using this scoring system is that the more pictures are chosen for each sentence, the more scores are distributed equally across pictures; for example when eight pictures are chosen, each pair of preferences (1 & 2, 3 & 4, etc.) are weighted equally. This reduces the variability inherent in a forced order of preference, whilst still remaining faithful to the chosen order. It is logical to assume that the fewer pictures that are chosen, the more relevant and important are the attributes of the chosen pictures and the higher weighting they should receive. In contrast, the more pictures that are chosen the less relevant are the attributes of any one picture and so the weighting should be more equally distributed. This is what our scoring system achieved in a simple manner.

Table 4-1: Scoring System used for picture choices*

No. of picture choices	Order of preference as chosen by participant							
	1	2	3	4	5	6	7	8
1 to 4	4	3	2	1				
5	4	4	3	2	1			
6	4	4	3	3	2	1		
7	4	4	3	3	2	2	1	
8	4	4	3	3	2	2	1	1

*the cell values in italics are the points allocated to each picture choice in the scoring procedure.

For each verb, the final scores for each picture were averaged across participants (the sum of all the points for that picture, divided by the total number of participants). This allowed us to see which pictures were most preferred for each verb. For each attribute, scores were summed across pictures with that attribute, e.g. horizontal or vertical axis, upwards or downwards arrow. For example, for the vertical axis there were four pictures (two up and two down), whereas for upwards motion there were two pictures. The maximum score for one picture was 4, if every participant selected it as their first preference. This meant that summing across pictures produced scores above 4 for strongly selected attributes, and below 1 or 2 for weakly selected attributes. The main aspects of interest were verbs that had strong agreement on the vertical axis and directionality of motion (upwards or downwards).

4.2. *Item Selection*

For each table, scores along the relevant attribute are provided as well as scores for the opposite spatial attribute; i.e. scores along the upward and downward dimension, and scores along the horizontal and vertical dimension. This is to illustrate the variability and range of the scores.

The results are split into two sections, section 4.2.1. summarises the initial set of items selected. These were chosen on the basis of the normed scores and previous experimental literature. After the completion of preliminary experiments, it was apparent that the initial selection may have been problematic. A second set of items

was then compiled, to give a larger item set and concentrate on more concrete verbs. These are summarised in section 4.2.2. In both cases, three categories of verbs had to be created: Down verbs, Up verbs and Control verbs. Therefore, they are items selected as having strong downwards motion, strong upwards motion and no particular direction of motion respectively.

4.2.1. **Item Set One**

4.2.1.1. *Single Words*

During the initial selection process, the focus was on a set of strongly directional verbs. The directional items (Up and Down categories) were selected with three constraints. First, and most importantly for the Up and Down categories, items had normed scores that had a relevant dominant direction. The experimenter's intuition was also used to narrow selection to verbs whose referents were felt to be clearly restricted to one direction, this excluded verbs such as "wither" and "support". Second, to select items that were unambiguously Up or Down, the dominant direction had to be complemented by a low score for the opposite dimension (e.g. a high 'Up' score and a low 'Down' score). A high score was defined as above 3, and a low score as below 1.5. One item ("hang") had a low score of 1.7 in the non-dominant direction, but as it was only 0.2 higher than the cut-off, it was decided that this exception would not be problematic. This criteria excluded items such as 'bounce' which were strongly vertical (Vertical score = 6.02), but were equally distributed for upwards (2.75) and downwards (3.27) direction. Third, the Up and Down groups had to be matched for frequency and length.

As can be seen in Table 3.2, the Up category has an average Up score of 4.5, and an average down score of 0.47. The group is also dominantly vertical, with an average Horizontal score of 0.92, and an average Vertical score of 4.98. The Down category has an average Up score of 0.33 and an average Down score of 4.69. This group is also dominantly vertical, with an average Horizontal score of 0.66 and an average Vertical score of 5.03.

Table 4-2: Item Set One - Up Category Verbs

Verb	Up Score	Down Score	Horizontal	<i>Vertical</i>
Ascend	5.17	0.18	0.13	5.35
Climb	5.05	0.10	0.20	5.15
Delight	3.23	0.80	2.03	4.03
Fly	4.47	0.47	1.23	4.93
Grow	5.27	0.17	0.70	5.43
Hoist	3.72	1.13	0.53	4.84
Honour	3.31	0.66	2.25	3.97
Hop	4.23	0.90	1.37	5.13
Hope	3.96	0.57	1.71	4.53
Increase	4.15	0.42	1.61	4.57
Jump	5.00	1.27	0.40	6.27
Leap	4.70	0.43	1.33	5.13
Lift	4.27	0.50	0.10	4.77
Raise	4.84	0.22	0.09	5.06
Rejoice	5.13	0.13	0.57	5.27
Rise	6.34	0.00	0.00	6.34
Sprout	4.57	0.27	1.00	4.83
Worship	3.67	0.33	1.37	4.00
average	4.50	0.47	0.92	4.98

Table 4-3: Item Set One - Down Category Verbs

Verb	Up Score	Down Score	Horizontal	<i>Vertical</i>
Crush	0.30	3.93	1.57	4.23
Decay	0.10	5.33	0.90	5.43
Decline	0.20	4.63	1.17	4.83
Decrease	0.15	4.67	1.42	4.82
Depress	0.30	3.70	1.33	4.00
Descend	0.28	4.95	0.25	5.23
Dig	0.43	3.80	0.50	4.23
Drain	0.80	3.63	1.03	4.43
Drip	0.22	4.03	0.50	4.25
Drop	0.00	5.15	0.05	5.15
Fall	0.00	6.10	0.00	6.10
Hang	1.70	3.22	0.05	4.92
Lower	0.63	4.28	0.28	4.91
Plummet	0.31	6.28	0.09	6.59
Plunge	0.19	6.13	0.28	6.31
Pour	0.10	3.77	1.33	3.87
Sink	0.24	5.21	0.24	5.45
Tumble	0.06	5.63	0.88	5.69
average	0.33	4.69	0.66	5.03

Table 4-4: Item Set One - Control Category

Verb	Up Score	Down Score	Horizontal	<i>Vertical</i>
Burn	2.60	1.80	1.07	4.40
Deny	0.67	1.77	4.07	2.43
Exhale	1.23	2.03	3.53	3.27
Exist	2.19	1.41	3.13	3.59
Fear	1.10	2.47	3.33	3.57
Hear				
Inherit	1.06	2.44	3.94	3.50
Inject	1.07	2.60	3.57	3.67
Kick	0.84	1.97	4.69	2.81
Promise	1.67	1.23	4.00	2.90
Provoke	1.40	1.83	4.77	3.23
Regret	1.37	1.93	2.43	3.30
Sell	0.43	1.30	4.30	1.73
Speak	2.09	0.94	3.88	3.03
Thank	0.93	0.83	4.83	1.77
Tug	1.47	1.59	4.88	3.06
Vault	2.10	1.17	1.77	3.27
Warn	0.93	1.80	4.17	2.73
average	1.36	1.71	3.67	3.07

The control category is equally distributed for Up (1.36) and Down (1.71) scores. The group has slightly higher Horizontal (3.67) as compared to Vertical scores (3.07), but this difference was not significant ($t(16)=1.539, p>0.1$).

Three measures of frequency were used. As the planned experiments were all to use auditory presentation, two of these frequency measures were for spoken items only. All three groups were matched for Brown Verbal frequency (Coltheart, 1981), Celex Spoken frequency per million and Combined Written and Spoken frequency per million (Baayen, Piepenbrock, & Gulikers, 1995) (all comparisons $t<1.7, p>0.1$). Two measures of length were used, length in letters and in syllables. All three groups were

also matched for these variables (all comparisons $t < 1.0$, $p > 0.1$); see Table 3.5 for the means.

Table 4-5: Item Set One Verbs - Mean values for each matched variable*

Lexical Variable		Category		
		Control	Down	<i>Up</i>
Frequency	Brown Verbal	10.39 (4.66)	2.50 (0.72)	3.78 (1.91)
	Celex Spoken	15.17 (7.43)	10.00 (4.48)	6.06 (2.89)
	Celex Combined	9.44 (3.07)	13.61 (5.71)	7.11 (2.07)
Length	Letter	5.00 (0.29)	5.28 (0.34)	5.11 (0.35)
	Syllables	1.50 (0.15)	1.44 (0.12)	1.33 (0.11)

* Standard deviation in brackets

4.2.1.2. Sentences

Sentences were also generated from the normed verbs, as far as possible the same verbs selected for Verb Item Set One were used in the sentences. For the Control Category, 3 items were shared; for the Down Category, 13 items were shared and for the Up Category, 15 items were shared. Sentences were designed so that each one had three parts, Subject, Verb and Object/Adverbial phrases. The active/simple past tense was used because it allowed a homogeneous and easily comprehensible set of sentences. The progressive tense necessitates the use of auxiliaries and the ‘ing’ form (‘The man *is digging*’) and sentences in the simple present imply habitual action for most of the verbs in the set (‘She digs (every day)’). Sentences were created so that the subject and object were not repeated between sentences, and the sentence was as well-formed as possible. For this reason, several of the sentences expressed abstract events (e.g. ‘The situation deteriorated quickly’). An initial set of 19 Up and 19 Down sentences were generated, along with 38 Control sentences. A pilot experiment was then used to establish the sentences’ acceptability. Sentences were played to participants over headphones and participants were asked to judge the sensibility of the sentences. Nonsense sentences were also created and included (see below). The mean accuracy of judging sentences correctly (as sensible) was then used as a measure of validity. From this pilot experiment, 3 Down sentences were rejected as their accuracy was below 70%, this necessitated the removal of 3 Up sentences, in order to balance the item numbers across groups; 3 Control sentences were also rejected. This left 16 Up and 16 Down sentences, for which 16 Control sentences were selected from the remaining 35.

The groups were matched for length in letters and syllables (all comparisons, $t < 1.7$, $p > 0.1$), see Table 3.6.

As can be seen, the sentences cover a large variety of event types, from abstract events such as emotional and financial state changes, to concrete events, primarily consisting of bodies in motion.

4.2.2. **Item Set Two**

Following the completion of initial experiments at the Science Museum (e.g. experiment 6.1) it was decided that the item set was problematic for two reasons. First, there were both concrete and abstract items in the set, which may have reduced effects of the concrete words by when analysed as a group. Although the abstract words had also been normed and selected on the basis of dominant vertical motion, without a concrete referent, we could not be sure what motion representation was being accessed during comprehension. In addition, new data was available that showed null effects for abstract sentences (Bergen et al, in press) confirming the decision to focus on concrete items. Second, the item set was still fairly limited in number and the verbs used for the sentences were not identical with the single verb set. For these reasons, it was decided that only verbs which could refer to concrete events would be used. In addition, the item set would be expanded to include as many verbs as possible, this led to the inclusion of verbs such as “wither” and “inflate”. Complimentary sentences, emphasising concrete events, would then be generated from the verb set.

4.2.2.1. *Single Words*

The constraints for selecting directional verbs remained the same as for Item Set One, except that all verbs had to refer to concrete events and the emphasis was on selecting the greatest number of items to increase power. This broadened the scope of which verbs were included, since if they were normed with a dominant direction and referred to a concrete event, they were included. This led to the inclusion of verbs such as “wither”, “crumble” “support” and “inflate”. In addition, Control verbs were now also selected for their concrete referents. Verbs referring to mental or emotional states were

excluded as far as possible, with the exception of “distract” in the Control group. For each group, 30 verbs were selected.

Table 4-6: Item Set One Sentences

Category	<i>Sentence</i>
Control	The woman accepted the gift
	The couple adopted a child
	The noise alarmed the dog
	The girl craved chocolate cake
	The ship endured the storm
	The student feared failure
	The box-office held the tickets
	The doctor injected the insulin
	The hypnotist mesmerised his audience
	The sportsmen obeyed the rules
	The ghost scared the children
	The mother spoke to her child
	The cat stayed indoors in winter
	The siren warned us of danger
	The baby grabbed the toy
The criminal robbed the bank	
Down	The cook crushed the fruit
	The crime rate decreased last year
	The house depreciated in value
	The book depressed everybody
	The lady descended the staircase
	The situation deteriorated quickly
	The farmer drained the pond
	The paint dripped onto the floor
	The postman dropped the parcels
	The villagers dug a well
	The coconuts fell from the palm-trees
	The stagehand lowered the curtain
	The temperature plummeted to sub-zero
	The lifeguard plunged into the water
	The pebbles sunk into the stream
The child tumbled out of bed	
Up	The walkers ascended the mountain
	The cat climbed up the tree
	The film delighted the critics
	The rocket flew to the moon
	The corn grew tall in the field
	The sailor hoisted the anchor
	The crowd honoured the athlete
	The rabbit hopped really high
	The jackpot increased each week
	The woman jumped for joy
	The horse leapt over the fence
	The builder lifted the bricks
	The manager raised wages
	The student rejoiced at her results
	The balloon rose into the air
The bird soared high in the sky	

Table 4-7: Item Set Two - Up Category Verbs

Verb	Up Score	Down Score	Horizontal	<i>Vertical</i>
arise	6.33	0.42	1.00	6.75
ascend	5.17	0.18	0.13	5.35
boost	5.67	0.83	1.58	6.50
climb	5.05	0.10	0.20	5.15
elevate	3.38	0.72	1.28	4.10
emerge	4.50	0.30	2.07	4.80
erect	3.92	0.50	3.00	4.42
escalate	5.75	0.92	1.08	6.67
expand	2.83	0.50	4.58	3.33
float	3.87	0.43	2.83	4.30
fly	4.47	0.47	1.23	4.93
grow	5.27	0.17	0.70	5.43
heighten	4.08	0.75	1.42	4.83
hoist	3.72	1.13	0.53	4.84
hop	4.23	0.90	1.37	5.13
increase	4.15	0.42	1.61	4.57
inflate	2.83	0.58	4.17	3.42
jump	5.00	1.27	0.40	6.27
launch	0.17	3.92	4.08	4.08
leap	4.70	0.43	1.33	5.13
lift	4.27	0.50	0.10	4.77
raise	4.84	0.22	0.09	5.06
rise	6.34	0.00	0.00	6.34
soar	5.56	0.31	1.25	5.88
spring	4.50	0.75	2.17	5.25
sprout	4.57	0.27	1.00	4.83
stack	2.72	1.84	1.63	4.56
support	3.38	1.28	1.78	4.66
surge	3.53	0.81	2.91	4.34
tower	0.93	0.83	4.83	1.77
average	4.19	0.73	1.68	4.92

The Up category has a mean Up score of 4.19, and an average down score of 0.73. The group is also dominantly vertical, with an average Horizontal score of 1.68, and an average Vertical score of 4.92. There are two anomalous items that need some comment. “Launch” has a normed Up score of 0.17, a Down score of 3.92 and a Horizontal score of 4.08. The nature of the norming task meant that the majority of participants selected a picture with a downwards facing arrow, as this represented the downwards thrust of the object as it launched. The variance of this items norming is also illustrated by the high horizontal score, since objects can also launch horizontally. However, we assumed that the central referent of the verb was an object taking off,

therefore referring to upwards motion despite the normed score. “Tower” has a normed Up score of 0.93, a Down score of 0.83 and a Horizontal score of 4.83. Although the verb does not refer to upwards motion, it does refer to a tall object in relation to other smaller objects, implying height and an ‘upward’ central reference point. In addition, “expand” and “inflate” also had Horizontal scores above 4.0; this is clearly due to their referent events occurring along all axes. However, these items were included as it was felt that they would not distort the group statistics, and were still appropriately ‘upward’ in nature. See Table 3.7 for the verbs and their normed scores.

Table 4-8: Item Set Two - Down Category Verbs

Verb	Up Score	Down Score	Horizontal	<i>Vertical</i>
bomb	0.21	3.69	1.14	3.90
cascade	1.08	5.25	1.50	6.33
collapse	0.33	6.00	2.08	6.33
crumble	0.33	5.25	2.33	5.58
crush	0.30	3.93	1.57	4.23
decline	0.20	4.63	1.17	4.83
decrease	0.15	4.67	1.42	4.82
deflate	0.83	5.58	2.50	6.42
demolish	0.41	3.69	1.88	4.09
depress	0.30	3.70	1.33	4.00
descend	0.28	4.95	0.25	5.23
deteriorate	0.00	5.27	1.40	5.27
dig	0.43	3.80	0.50	4.23
dive	0.33	6.17	2.33	6.50
drain	0.80	3.63	1.03	4.43
drip	0.22	4.03	0.50	4.25
drop	0.00	5.15	0.05	5.15
dump	0.17	3.75	3.67	3.92
fall	0.00	6.10	0.00	6.10
hang	1.70	3.22	0.05	4.92
lapse	0.17	3.92	4.08	4.08
lower	0.63	4.28	0.28	4.91
plummet	0.31	6.28	0.09	6.59
plunge	0.19	6.13	0.28	6.31
pour	0.10	3.77	1.33	3.87
rot	0.33	4.40	1.40	4.73
sink	0.24	5.21	0.24	5.45
slump	0.17	5.83	1.92	6.00
tumble	0.06	5.63	0.88	5.69
wither	0.40	4.50	1.27	4.90
average	0.36	4.75	1.28	5.10

The Down category has an average Up score of 0.36 and an average Down score of 4.75. This group is also dominantly vertical, with an average Horizontal score of 1.28 and an average Vertical score of 5.10. As can be seen in Table 3.8, the Down verbs

make up a more homogeneous set than the Up verbs, with no anomalous items (i.e. all have a Down score above 3.0 and an Up score of 1.7 or below). “Lapse” and “dump” have Horizontal scores above 3.5, due to the fact that their referent events are not strictly limited to the vertical dimension.

Control verbs were selected so that they matched the Down and Up Categories for frequency and length. As stated above, they also had to refer to concrete events, so that the contrast between them and the directional verbs was based on motion rather than abstraction. The same measures of length were used here (letters and syllables), with Celex combined and spoken frequency per million used as the frequency measures. It was decided that it was unnecessary to have two measures of verbal frequency, so Brown Verbal frequency was not used (it being based on older corpora than Celex).

Table 4-9: Item Set Two - Control Category Verbs

Verb	Up Score	Down Score	Horizontal	<i>Vertical</i>
attack	1.66	2.81	3.72	4.47
burn	2.60	1.80	1.07	4.40
catch	1.47	2.37	3.17	3.83
chase	1.40	0.87	5.60	2.27
cross	0.63	0.80	5.60	1.43
depart	1.57	1.42	5.03	2.98
distract	1.09	1.34	5.50	2.44
drag	0.90	0.97	6.37	1.87
eat	1.25	2.00	3.78	3.25
eject	2.10	2.17	3.73	4.27
enter	2.50	2.25	3.44	4.75
exchange	0.78	0.91	5.50	1.69
exhale	1.23	2.03	3.53	3.27
glide	0.73	0.63	6.07	1.37
jog	0.97	0.23	5.77	1.20
kick	0.84	1.97	4.69	2.81
pull	1.91	1.66	4.84	3.56
quit	1.20	2.00	4.50	3.20
race	0.67	0.30	5.10	0.97
retrieve	1.03	1.77	4.03	2.80
reward	1.97	2.37	3.63	4.33
rust	0.43	2.80	2.47	3.23
snatch	1.23	1.03	4.57	2.27
sneeze	0.90	0.83	4.90	1.73
stagger	0.50	1.27	5.03	1.77
sweep	0.59	1.41	5.22	2.00
swim	1.10	0.93	6.23	2.03
tremble	0.70	1.93	3.47	2.63
tunnel	1.10	3.50	3.10	4.60
wash	0.30	2.23	4.80	2.53
average	1.18	1.62	4.48	2.80

The Control group had an Up score of 1.18 and a Down score of 1.62, this difference was significant ($t(29)=2.917$, $p<0.01$) but small given the size of the differences present for the Up and Down groups (Control mean difference = 0.44; Up mean difference = 3.46; Down mean difference = 4.39). The mean horizontal score was high (4.48), bolstered by several verbs which refer to concrete motion events that typically involve horizontal motion, e.g. “chase”, “swim” and “drag”.

The Control group was matched to the Up and Down groups for both frequency measures and all length measures (all comparisons $t<1.2$). In addition, because the items were to be presented as auditory stimuli, the groups were matched for number of phonemes (all comparisons $t < 1$) and phonological neighbourhood size (all comparisons $t < 1.2$) (Pisoni, D; Hernandez, L. et al, 2007). The Up and Down groups were matched for all measures except Celex Combined Frequency (Baayen, Piepenbrock & Gulikers, 1995), for which Up verbs were slightly more frequent ($t(58)=2.081$, $p<0.05$); see Table 4.10.

Table 4-10: Item Set Two Verbs - Mean values for each matched variable*

Lexical Variable		Category		
		Control	Down	<i>Up</i>
Frequency	Celex Spoken	4.07 (8.62)	3.17 (7.69)	5.60 (9.68)
	Celex Combined	6.27 (10.49)	3.63 (6.58)	10.17 (15.88)
Phonological Neighbourhood	Luce's N	2.63 (5.14)	1.77 (6.72)	2.24 (4.40)
Length	Letters	5.23 (1.41)	5.63 (1.85)	5.37 (1.38)
	Syllables	1.37 (0.49)	1.60 (0.86)	1.43 (0.63)
	Phonemes	4.37 (1.38)	4.47 (2.03)	4.03 (1.10)

*Standard deviation in brackets

4.2.2.2. Sentences

The major constraints were to have one sentence created for each of the verbs and for the sentences to refer to concrete events. The sentences were created with the same structural constraints as Item Set One, being three part constructions with Subject, Verb and Object/Adverbial phrases. Sentences were also created to maximise the concrete aspects of any event by having a concrete subject and, where possible, a concrete direct

object. Where this was not possible, the sentence was created ambiguously so that it could refer to a concrete event (e.g. “The quantity decreased over time”).

Table 4-11: Item Set Two – Control Sentences

Verb	<i>Sentence</i>
attack	The army attacked before sunrise
burn	The bonfire burned brightly
catch	The boy caught the base-ball
chase	The cat chased the mouse
cross	The woman crossed the road
depart	The train departed at 2 o'clock
distract	The noise distracted the dog
drag	The dogs dragged the sled
eat	The woman ate her breakfast
eject	The hi-fi ejected the CD
enter	The woman entered the room
exchange	The family exchanged Christmas presents
exhale	The man exhaled cigarette smoke
glide	The skater glided on the ice
jog	The man jogged for 20 minutes
kick	The girl kicked the football
pull	The horses pulled the carriage
quit	The man quit his job yesterday
race	The cars raced around the track
retrieve	The dog retrieved the stick
reward	The lecturer rewarded his students
rust	The car rusted in the garage
snatch	The child snatched the lolly-pop
sneeze	The man sneezed loudly
stagger	The man staggered towards home
sweep	The caretaker swept the floor
swim	The girl swam for an hour
tremble	The ground trembled under foot
tunnel	The mole tunnelled in the garden
wash	The woman washed her clothes

To provide variety, specific subjects and objects were not repeated across sentences; some repetition of the subjects was necessary, but these were only for generic terms such as ‘boy’, ‘woman’ and so on. Table 4.11 lists the Control sentences, Table 4.12 lists the Up sentences and Table 4.13 lists the Down sentences. The sentences were matched for length in number of words and syllables (all comparisons, $t < 1.3$).

Table 4-12: Item Set Two - Up Sentences

Verb	<i>Sentence</i>
arise	The woman arose from bed
ascend	The walkers ascended the mountain
boost	The jet boosted for take-off
climb	The boy climbed the rope
elevate	The nurse elevated the man's leg
emerge	The shoots emerged from the soil
erect	The men erected the platform
escalate	The amount escalated rapidly
expand	His chest expanded with pride
float	The kite floated on the wind
fly	The rocket flew to the moon
grow	The corn grew quickly
heighten	The builders heightened the ceiling
hoist	The sailor hoisted the anchor
hop	The frog hopped really high
increase	The ladder increased in height
inflate	The gas inflated the balloon
jump	The girl jumped onto the wall
launch	The helicopter launched from it's base
leap	The horse leapt over the fence
lift	The builder lifted the bricks
raise	The woman raised the curtain
rise	The smoke rose into the air
soar	The bird soared in the sky
spring	The boy sprang to his feet
sprout	The plants sprouted in the garden
stack	The man stacked the boxes
support	The walls supported the roof
surge	The waves surged against the cliffs
tower	The church towered over the town

The constraint that sentences should refer to concrete events resulted in some slightly atypical uses for verbs whose canonical context was with abstract subjects or objects, even though they do not necessarily refer to an abstract event; specifically “lapse” and “heighten”. Following a pilot experiment in which 29 subjects made sensibility judgements about the sentences, three sentences (one from each Item Group) were found to have acceptability ratings substantially lower than the group mean. Acceptability is measured as the percentage of time each sentence was judged as ‘sensible’ rather than ‘nonsense’. Unacceptable sentences were as follows: Control “The army attacked before sunrise” (64%), Down “The bridge lapsed into disrepair” (40%) and Up “The builders heightened the ceiling” (62%). The overall group mean was 86% (standard deviation = 9.4%). Following this, these sentences were not included in experimental analyses, leaving 29 items in each group.

Table 4-13: Item Set Two - Down Sentences

Verb	<i>Sentence</i>
bomb	The planes bombed the city
cascade	The water cascaded onto the rocks
collapse	The wall collapsed completely
crumble	The castle crumbled to dust
crush	The hammer crushed the glass
decline	The path declined steeply
decrease	The quantity decreased over time
deflate	The dingy deflated slowly
demolish	The work-men demolished the chimney
depress	His finger depressed the button
descend	The woman descended the stairs
deteriorate	The level deteriorated quickly
dig	The villagers dug a well
dive	The eagle dived into the valley
drain	The farmer drained the pond
drip	The paint dripped onto the floor
drop	The postman dropped the parcels
dump	The lorry dumped the rubbish
fall	The fruit fell from the tree
hang	The clothes hung from the line
lapse	The bridge lapsed into disrepair
lower	The men lowered the bridge
plummet	The meteor plummeted to earth
plunge	The lifeguard plunged into the water
pour	The water poured from the tap
rot	The leaves rotted into the soil
sink	The pebbles sunk into the stream
slump	The boy slumped in his seat
tumble	The child tumbled out of bed
wither	The plant withered in the heat

4.3. *Single Word Norming*

After Item Set Two had been generated and established as the final item set, these items were normed in a Single Word context, rather than the Rebus Sentence context used initially. During experiments, items were going to be presented both within and without a sentence, and it was possible that the Rebus norming procedure biased the scores towards a sentence context so that they would not reflect the interpretation of the individually presented item. In the Rebus context all items are unambiguously verbs, whereas individual presentation allows wider interpretation, e.g. as a noun which does not retain the verb meaning (e.g. 'to fly' versus 'the fly').

The single word norming was an adaptation of the original procedure, with verbs presented in their uninflected bare stem forms, with four arrows (up, down, left and right) alongside the verb. 24 Native English speakers were asked to select a minimum of one and a maximum of four arrows that best represent the event described by the verb. The ratings were scored using the same scale as for the Rebus Norming, with higher preference given larger weighting than lower preference selections. The scores for each arrow were then averaged across subjects to provide a normed score for each direction, the higher the normed score, the stronger the rating for that direction (with a maximum value of 4). For vertical and horizontal directions, the scores for the two vertical and horizontal arrows were summed, respectively. See Appendix 1c for the norming materials.

Statistical comparisons were used to compare the Rebus and Single Word Normed scores for four dimensions: upwards, downwards, horizontal (summed across left and right arrows) and vertical (summed across upwards and downwards arrows), for each Item Group (Up, Down and Control). A 2 (Rebus vs. Single Word Norming procedures) by 4 (Upwards, Downwards, Horizontal and Vertical Direction) by 3 (Control, Up and Down Item Groups) mixed effects ANOVA, with Item Group between subjects, was used to compare the Single Word and Rebus normed scores. The interaction between Norming procedure and Direction was significant ($F(3,261)=5.287$) as was the interaction between Norming and Item Group ($F(2,87)=16.087$). Both of these interactions were driven by lower scores overall for the Single Word norming ($m = 2.316$, $SE = 0.021$) as compared the Rebus norming ($m = 2.775$, $SE = 0.043$), reflecting the difference between the number of pictures in the two procedures.

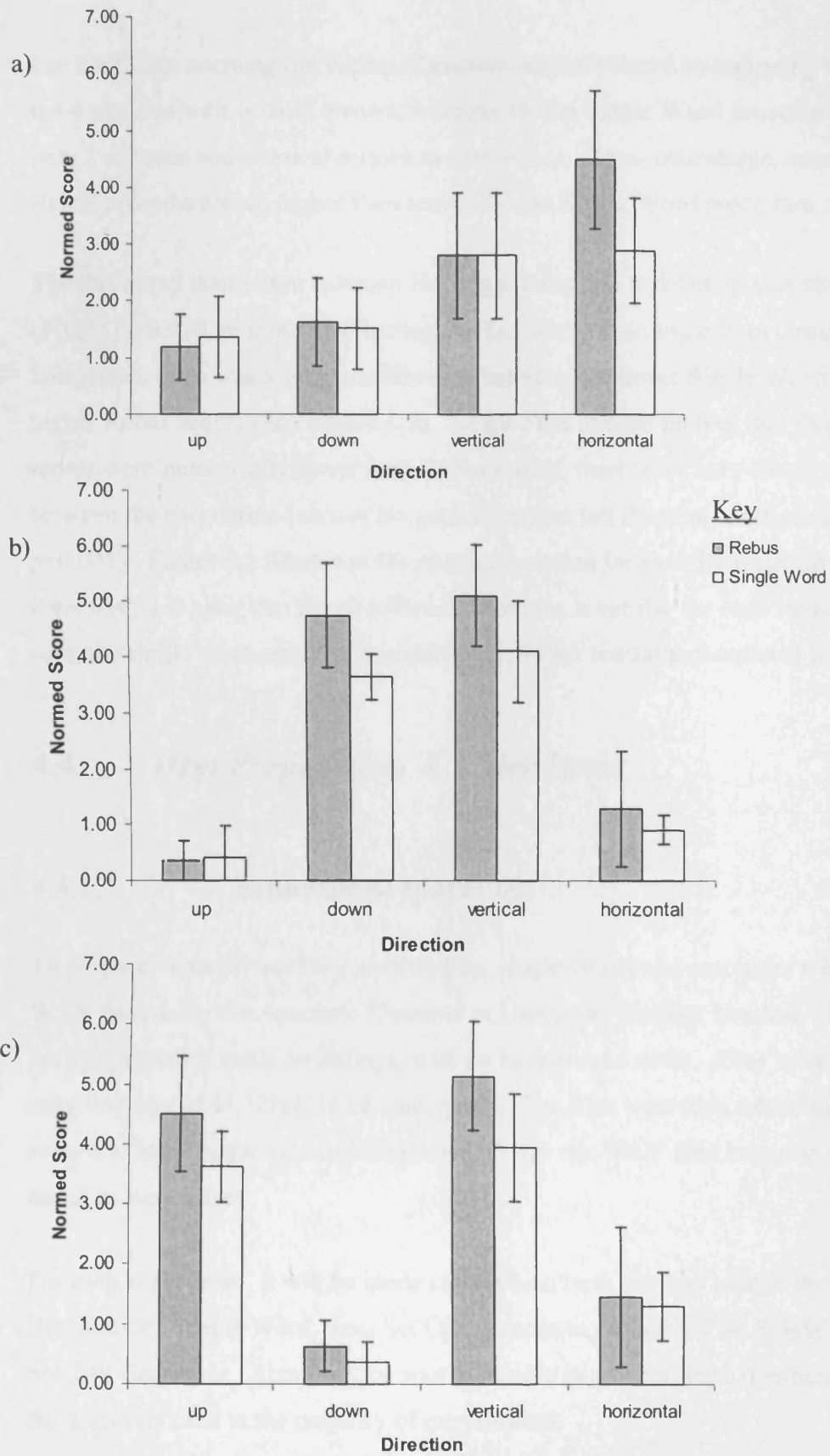


Figure 4-2: Rebus and Single Word Normed scores for each Direction by Item Group (a) Control (b) Down and (c) Up

For the Rebus norming the vertical direction was calculated by summing the scores for the 4 pictures with vertical arrows, whereas for the Single Word procedure, there were only 2 pictures with vertical arrows to sum across. Thus, on average, scores for the Rebus procedure were higher than scores for the Single Word procedure.

The three way interaction between Norming, Direction and Group was also significant ($F(6,261)=9.938, p<0.001$), reflecting the fact that within some Item Groups, for some Directions, there was a greater difference between the lower Single Word scores and higher Rebus scores (See Figure 4.2). Despite the overall finding that Single Word scores were numerically lower than Rebus scores, there were very strong correlations between the two normed scores for each Direction (all Pearson coefficients $> 0.74, p<0.001$). Figure 3.2 illustrates the close correlation for each Item Group across each Direction; it is clear that the distribution of scores is similar for each Item Group. In sum, the single word norming backed up the Rebus norming completed previously.

4.4. *Item Preparation & Filler Items*

4.4.1. *Soundfile preparation*

To prepare items for auditory presentation, single words and sentences were recorded as WAV files using the Anechoic Chamber at University College London. This produces very high quality voice recordings, with no background noise. Files were recorded at a sampling rate of 44.1kHz, 16 bit rate, mono. The files were then edited using Audacity software (<http://audacity.sourceforge.net>) so that the WAV files began at item onset and ended as item offset.

For each experiment, it will be made clear which Item Set was used in the experiment: Item Set One Single Word, Item Set One Sentences, Item Set Two Single Word or Item Set Two Sentences. Item Set One was used only in preliminary experiments, and Item Set Two was used in the majority of experiments.

4.4.2. Filler Words and Sentences

4.4.2.1. Non-words for Lexical Decision

103 nonsense words were created for lexical decision experiments. Words were generated from the ARC Nonword Database (Rastle, Harrington & Coltheart, 2002). Words were generated that had 3-8 letters, with only orthographically existing onsets and bodies. In addition, they had to be pronounceable but not a homophone of any existing English word. A full list can be found in Appendix 1d.

4.4.2.2. Nonsense Sentences for sensibility judgements

60 nonsense sentences were constructed for experiments that used sentence sensibility judgements, for example “The woman decorated in mice” or “the laughter infected the jam-jar”. A pilot experiment showed that the nonsense sentences were judged as nonsensical over 90% of the time. A complete list can be found in Appendix 1e.

4.4.2.3. Filler words and sentences for active comprehension

For experiments that used active comprehension, filler words and sentences had to be created that could be followed by comprehension questions.

60 filler words were selected that could be followed by comprehension questions of the form “Was the word related to X?”. Probe items (X) were chosen to be associated to the target filler for ‘yes’ responses. Several methods were used to select associated words. Some were selected from the University of South Florida association norms (Nelson, McEvoy, & Schreiber, 1998) e.g. “blush” for “embarrass”. Synonyms were used, e.g. “compatible” for “agree” and some were judged intuitively by the experimenter to be related to the target, e.g. “clay” for “sculpt”. For the ‘no’ responses, these same words were randomly re-paired with unassociated targets. 18 fillers that could not be paired with a strongly associated target were left as additional items to be used when needed (e.g. to produce blocks with the same number of trials so the

computer programmes could run smoothly). See Appendix 1f for a complete list of these words and their associated probes.

45 filler sentences were created that were followed by comprehension questions of the form “Did the X verb the Y?”. Questions requiring a positive response were identical to the previously presented sentence. Questions requiring a negative response had a different subject noun (X), verb or object noun (Y). For example, the filler sentence “The man regretted his behaviour” had the question “Did the man regret his haircut?”. In all cases the altered lexical item was chosen to be a plausible alternative. See Appendix 1g for a complete list of the sentences and their comprehension questions that required a negative response (i.e. that differ from the original sentence).

5. Comprehension on Visuo-Spatial Attention

In this section four experiments are reported which investigate the effect of comprehension on visuo-spatial attention. All experiments are adaptations and extensions of the method used by Richardson et al (2004). Participants comprehend sentences or single words that refer to motion events and then categorise a shape (a circle or a square) that appears in the top or bottom of the visual field. In the following experiments, we presented participants with single words (experiments 1 & 2) and sentences (experiments 3 & 4), varying the synchronisation between the presentation of the linguistic and visual stimuli. Comprehension was passive in Experiment 1, and active in Experiments 2, 3 & 4, when participants had to answer questions about the linguistic stimuli. We varied the linguistic context to address how much linguistic information is necessary to effect visuo-spatial attention. In Experiment 1 participants heard blocks of single words unsynchronised with categorisation trials. In Experiment 2 single words were presented in a random order with categorisation synchronised to word offset. In Experiment 3 sentences were presented with categorisation at sentence offset and in Experiment 4 with categorisation at verb offset.

5.1. *Blocked Word Comprehension & Shape Categorisation*

Whilst passively comprehending the words, participants performed a categorisation task for shapes that appeared in the top or bottom parts of the visual field. This experiment was run at the Science Museum as part of their Live Science initiative. The time taken to complete the experiments had to be brief so that volunteers from the general public did not have to commit too much of their museum visiting time. This experiment presented blocks of single words from each category (Control, Up, Down) as well as a Mixed block (half Up and half Down items). We hypothesised that blocked presentation would create a consistent linguistic context, in addition it allowed passive comprehension of the stimuli (as they are irrelevant to the task). The hypothesis is that if directional words influence visuo-spatial attention there should be effects of congruency on reaction times for detecting stimuli in the top or bottom of the visual field. In line with previous experiments (Richardson et al, 2003; Bergen et al, in press), we expect congruent interference, such that reaction times are longer when the words and location are matched (e.g. up words and items at the top of the visual field).

5.1.1. **Participants**

52 native English Speakers (25 female) took part in the experiment. The mean age was 29.12 years (SD = 9.54). All had normal or corrected to normal vision.

5.1.1. **Method**

5.1.1.1. *Design*

A 2x4 repeated measures design was used. The independent variables were the screen Position of the object (top or bottom) and the Word Category by block (Control, Mixed, Match and Mismatch). The dependent variables were the reaction time (ms) and errors for categorising the shape. There were a total of 4 blocks, each of which contained 20 object categorisation trials. Each block contained an equal number of shape (square or circle) and shape location (top or bottom) trials. The presentation order of blocks and trials was fully randomised.

5.1.1.2. *Materials*

Item Set One Single Words was used. For each Word Category (Control, Up, Down and Mixed) four different random orders of items were created. The Mixed block was created by combing 9 Up items with 9 Down items; 2 random orders contained one half of the Up and Down items and the other 2 random orders contained the other half of the Up and Down items. Individual audio files for each random order were concatenated into a single continuous WAV sound file. The concatenated sound files were 20 seconds long, with 2 seconds of silence followed by one word presented every second. During each Word Category block two continuous files were consecutively presented, each with a different ordering of items. Files for the Mixed blocks were selected so that all the Up and Down items were heard once. Therefore, participants heard each directional item 5 times across the whole experiment.

5.1.1.3. Graphic Display

The circle and square objects subtended 0.66° of visual angle. The upper object was presented at 6.5° above the location of the central fixation cross, and the lower object was presented at 6.5° below fixation.

5.1.1.4. Apparatus

The experiment was run using PC computers (Intel Pentium III) using E-Prime 1.0 Software (Schneider, Eschman, & Zuccolotto, 2002). A standard CRT monitor was used with a refresh rate of 60 Hz. Participants responded using standard 'qwertyuiop' keyboards. Chin rests were also used to minimise head movements, these were placed at a distance of 57cm from the computer screen.

5.1.1.5. Procedure

Participants were instructed to respond to the shape as quickly and accurately as possible by pressing the "square" key if it was a square, and the "circle" key if it was a circle (square or circle shaped stickers were placed over the A and L keys); allocation of response keys was balanced across participants. Participants wore stereo closed headphones and placed their chin on the rest. Participants were given 10 practice trials without words, they then completed 4 experimental blocks. The first WAV file began playing at the block onset, therefore the first word was heard 2000ms after the start of the block. The second WAV file began playing after the 10th categorisation trial (half way through the block). A fixation cross was presented for 1500ms at the start of each WAV file so that categorisation trials were not completed when no words were heard. Each categorisation trial began with a fixation cross for a random duration between 1000-1200ms in 50ms steps, a shape presented in the top or bottom of the visual field for 200ms and then a fixation cross until the response or a timeout of 2500ms.

5.1.2. Results

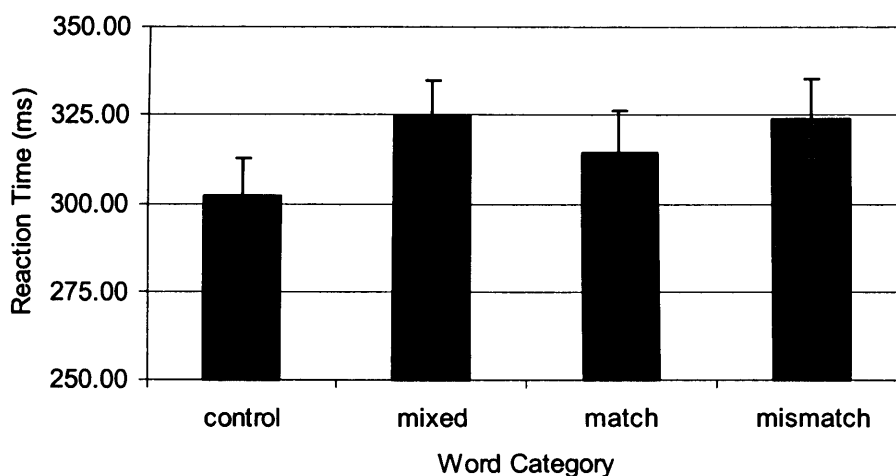
5.1.2.1. *Analysis*

6 participants were removed as their categorisation accuracy was under 80%. Reaction times below 100 and above 1000 milliseconds for correct trials were excluded, this removed 3.3% of the data. Errors were analysed separately.

5.1.2.2. *Reaction Time Data*

A 2x4 repeated measures ANOVA showed a significant main effect of Word Category ($F(3,45) = 5.771$, $\text{partial-}\eta^2 = 0.114$), no other effects reached significance. Planned comparisons showed that the Control category had significantly shorter reaction times than the Mixed ($F(1,45) = 15.478$, $\text{p-}\eta^2 = 0.256$) and Mismatch categories ($F(1,45) = 11.176$, $\text{p-}\eta^2 = 0.199$). The difference between the Match and Mismatch categories was marginal ($F(1,45) = 3.303$, $\text{p-}\eta^2 = 0.068$), with shorter reaction times for Match. As there was no significant main effect or interaction for Position, Figure 5-1 presents the reaction times by Word Category collapsed across Position.

Figure 5-1: Reaction times for shape categorisation by Word Category collapsed across Position*



*Error bars are one standard error of the mean

A follow up analysis compared reaction times for the first half of each block (the first 10 trials) with reaction times for the second half of each block (the second 10 trials). This comparison showed a main effect of Word Category that was weaker than the effect when whole blocks were analysed ($F(3,45) = 3.910$, $p-\eta^2 = 0.08$). There was also a significant main effect of Block Half, with reaction times in the first half an average of 13.57 milliseconds ($SE = 4.337$) faster than reaction times for the second half ($F(1,45) = 9.796$, $p-\eta^2 = 0.179$). Although there was no interaction between Word Category and Block Half, Table 5-1 illustrates that the increase in reaction times was predominantly for the Mixed, Match and Mismatch conditions. See Table 5-2 for mean reaction times for each condition.

5.1.2.3. Errors

The four conditions were summed across Position and a Friedmann test was used to compare the four conditions for number of errors; there was no significant difference across Word Category ($X^2 = 6.92$, $p = 0.075$). The Wilcoxon Signed Rank Test was then used to compare errors in each condition (Control, Mixed, Match and Mismatch) for the Top and Bottom screen position. The difference for these locations approached significance for the Mixed block, with more errors when shapes were presented in the bottom than the top of the visual field ($Z = -1.909$, $p = 0.056$).

Table 5-1: Mean Reaction Times (ms) for Word Category by Block Half*

Word Category	Block Half		Difference
	First	Second	
<i>Control</i>	304.41 (96.72)	304.48 (69.08)	0.06
<i>Mixed</i>	313.37 (84.96)	334.30 (79.80)	20.92
<i>Match</i>	305.55 (90.40)	321.75 (83.38)	16.194
<i>Mismatch</i>	314.81 (89.56)	331.92 (82.11)	17.11

*Standard deviation in brackets

5.1.3. Discussion

The results show a main effect of Word Category, with Mismatch and Mixed blocks producing longer reaction times than Control blocks. Therefore, when the direction of the words is incongruent, or inconsistent, with the location of the shape, people take longer to categorise the shape compared to when no directional words are presented. However, when the direction of the words is congruent with the location of the shape (Match) participants are no faster than when no direction (Control) or inconsistent direction (the Mixed block) words are presented. There was a trend for participants to be faster when the words were congruent rather than incongruent with the location. These results are not in line with the hypothesis, which predicted congruent interference. Instead, the data show an incongruent interference effect. When the word direction is congruent with location this is not significantly different from the Control condition, so congruence does not slow categorisation relative to the baseline. The data is in contrast to previous findings which show that congruence produces interference relative to incongruence (Richardson et al, 2003; Bergen et al, in press), but both these experiments used sentences rather than single words. In addition, the blocked presentation means that participants were aware of the directional content of the words; a few participants reported this to the experimenter following the experiment. This could have influenced the direction of the observed effects, since strategies could change task performance. The comparison across block halves showed that reaction times were significantly longer in the second block half than that first, and that this increase in reaction times was predominantly for conditions where direction words were presented. It appears that the interference from the direction words builds up over the course of the block, supporting the inference that it is the context created by blocked presentation that is effective, rather than the individual words. To assess whether the directional content of single words is enough to influence visuo-spatial attention, the next experiment presented individual words in a randomised order

5.2. *Single Word Comprehension & Shape Categorisation*

Participants were presented with single words over headphones; after the presentation of a word they performed a categorisation task for shapes appearing in the top or bottom of the visual field. On filler trials, participants answered comprehension questions about the words. Blocked single word presentation influenced categorisation but it is not clear whether single words will produce a similar effect. The hypothesis is that if single words are sufficient to effect categorisation, similar effects of congruency will be found to Experiment 5.1; when words are incongruent to the shape location, reaction times will be longer.

5.2.1. **Participants**

20 undergraduate students (14 female) volunteered as part of an opportunity sample. The mean age was 21.4 years (SD = 1.90). All subjects were Native English speakers and had normal or corrected vision.

5.2.2. **Method**

5.2.2.1. *Design*

A 3x2 repeated measures design was used. The independent variables were the screen Position of the object (top or bottom) and the Word Category (Control, Match and Mismatch). The dependent variables were the reaction time and errors for categorising the shape (ms). There were two halves to the experiment, with each word presented once in each half. In each half, 6 blocks of 28 trials quasi-randomly sorted to include 15 experimental and 13 filler items. In the first half each item was followed by a shape in one location (Top or Bottom), in the second half each item was followed by a shape in the other location. Comprehension questions followed filler items, each filler item was followed by one question during the experiment; 5 comprehension questions were asked per block. Each block contained an equal number of shape (square or circle) and shape location (top or bottom) trials; shape was kept constant across both presentations of the item.

5.2.2.2. Materials

Item Set Two Single Words was used.

78 filler words were selected from the set created for this purpose; 60 were followed by comprehension questions of the form “Was the word related to X?”, requiring a yes/no answer. Two lists were created. In list 1, all the experimental words appeared once with the circle or square appearing at the top or bottom of the screen. In list 2, the words appeared once with the shape appearing at the opposite location. The shape for each sentence was kept constant across lists, as was the randomised interval before shape presentation. For the filler words, the shape was kept constant across lists, but the interval was not. This was to create extra variation between verb offset and shape onset, preventing expectations of shape onset. In addition, in list 1, half the filler words were followed by a comprehension question; in list 2, the other half of the filler words were followed by a comprehension question. Across each half, there were an equal number of ‘Yes’ and ‘No’ responses to the comprehension questions. Thus, comprehension questions occurred in 18% of trials.

5.2.2.3. Graphic Display & Apparatus

These were the same as Experiment 1, with the exception that participants did not place their chin on a rest and the monitor refresh rate was set to 75Hz.

5.2.2.4. Procedure

Participants were instructed to respond to the shape as quickly and accurately as possible by pressing the X key if it was a square and the Z key if it was a circle. Participants were also told that on some trials, following the shape categorisation, they would be asked a question about the word that they had just heard. The response options were displayed underneath the question with NO on the left and YES on the right. Participants were instructed to respond using the same keys as in the experimental trials, pressing the key that corresponded to the side their chosen answer was displayed i.e. Z for yes and X for no. Participants were given 18 practice trials, 8 of which were followed by comprehension questions. They were given feedback on their

performance for the questions. They then completed 12 experimental blocks. During a trial, a fixation cross appeared for 1000ms and stayed on screen whilst the word was presented aurally. There was then an ISI of between 27 – 95 ms (in 13ms steps, chosen to match the refresh rate of the monitor). The shape then appeared for 100ms followed by a fixation cross until the response, or to a timeout of 3000ms. On filler trials, comprehension questions were presented until a response or to a timeout of 5000ms.

5.2.3. Results

5.2.3.1. *Analysis*

One participant was removed as their accuracy for the comprehension questions was less than 75%. Correct trials with reaction times between 100ms and 1000ms were analysed, this removed 3.7% of the data. Errors were analysed separately.

5.2.3.2. *Reaction Times*

3x2 repeated measures ANOVA showed no significant differences in the by Subjects analysis. The by Items analysis showed a significant main effect of Position ($F(1,87) = 9.733$, $p-\eta^2 = 0.101$), with mean reaction times to shapes in the bottom of the visual field 14.39ms (SE = 4.61) faster than reaction times to the top of the visual field. See Table 5-2 for mean reaction times for each condition.

Table 5-2: Mean Reaction Times (ms) by Condition for Experiments 5.1-5.4†

Experiment	Condition							
	Control		Match		Mismatch		Mixed	
	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom
5.1 Blocked Single Word	299.99 (72.04)	303.99 (75.14)	311.76 (79.55)	317.20 (78.32)	318.78 (78.68)	328.96 (79.41)	324.98 (64.33)	324.11 (77.05)
<i>Category Mean</i>	301.99 (73.23)		314.47 (78.55)		323.87 (78.76)*		324.55 (70.59)*	
5.2 Single Word	497.82 (113.13)	490.65 (128.31)	504.82 (117.65)	488.08 (126.05)	500.54 (112.63)	494.72 (114.24)		
<i>Category Mean</i>	494.23 (119.37)		496.45 (120.57)		497.63 (111.93)			
5.3 Full Sentence	436.50 (65.69)	422.34 (81.82)	416.36 (80.43)	420.85 (65.40)	436.30 (72.98)	451.88 (80.19)		
<i>Category Mean</i>	429.42 (73.63)		418.60 (72.44)		444.09 (76.14)*			
5.4 Partial Sentence	894.37 (96.81)	899.47 (88.83)	862.44 (78.96)	836.19 (93.15)	828.62 (85.89)	860.31 (86.31)		
<i>Category Mean</i>	896.92 (91.37)		849.31 (84.97)*		844.46 (83.77)*			

†Standard deviation in brackets

*Significantly different to Control condition, $p < 0.05$

5.2.3.3. Errors

The Friedman test showed no significant difference for errors between Word Categories by Subjects ($X^2_1 < 1$) or by Items ($X^2_2 < 2.1$). The Wilcoxon Signed Rank Test showed no significant differences for any of the conditions (Control, Match or Mismatch) for the two screen locations by Subjects. By Items, the Mismatch condition had more errors for the Bottom screen location ($Z_2 = -2.120$).

5.2.4. Discussion

There was no significant effect of Word Category. When words were synchronised to the presentation of the shape, but presented in a randomised order, there was no effect of congruence. There is some evidence that nouns activate the canonical spatial location of their referent (e.g. attic vs. cellar) (Estes et al, submitted). However, our stimuli are verbs and they refer to a direction of motion rather than a specific location. Experiment 1 demonstrated that blocked presentation of single words provides enough context to affect visuo-spatial attention, blocked presentation develops a context over-time that can bias attention, producing incongruent interference but not congruent facilitation. In the case of Experiment 2, single words that refer to a general direction do not influence visuo-spatial attention. For the sake of argument we can assume that comprehension of single concrete nouns does activate a spatial representation of a particular location, or bias attention towards that location. Single verbs that contain directional information may not be concrete enough to actively effect visuo-spatial attention, as they specify general motion rather than a specific location.

There was a trend, by items, for faster responses to the bottom of the visual field. There is evidence that the lower visual field (equivalent to the Bottom location for the shape) is more sensitive to fine discrimination for objects that are close to the observer (Previc, 1990). Therefore, this trend that we see in the data for faster categorisation in the lower hemi-field may be a reflection of this specialisation.

There was a trend, significant by Items but not by Subjects, for more errors in the bottom screen location for the Mismatch condition. This means that Up words produced more errors for shapes in the Bottom screen location. This provides limited support for incongruence producing interference, as seen by errors, but it not a reliable finding. The next set of experiments present full sentences, therefore the location (e.g. the endpoint of the object's motion) is specified. If a concrete location is essential in influencing visuo-spatial attention, we should replicate the finding from Richardson et al (2003) and Bergen et al (in press).

5.3. *Full Sentence Comprehension & Shape Categorisation*

The experiment was the same as Experiment 2 except that participants were presented with sentences rather than single words. The shape for categorisation was presented at sentence offset. In line with previous findings (Richardson et al, 2003; Bergen et al, in press), the hypothesis is that when sentences describe motion that congruent with the target location, reaction times to categorise the shape will be longer.

5.3.1. **Participants**

21 undergraduate students (17 female) took part in the experiment as part of a course requirement. The mean age was 18.9 (SD = 0.58). All subjects were monolingual Native English speakers and had normal or corrected to normal vision.

5.3.2. **Method**

5.3.2.1. *Design*

The design was the same as Experiment 5-2. There were a total of 36 blocks (18 in each half) of 6,7 or 8 trials. Each block contained an equal number of shape (square or circle) and shape location (top or bottom) trials. Within each block there were 5 sentences from the experimental conditions and 1, 2 or 3 filler sentences, of which 0,1 or 2 were followed by comprehension questions; there were questions on 14% of trials. Presentation of blocks was fully randomised within each half of the experiment. There were two halves to the experiment, with each word presented once in each half. In the first half each item was followed by a shape in one location (Top or Bottom), in the second half each item was followed by a shape in the other location; the shape was constant for both presentations of the same item. See Materials (section 5.2.2.2) for details of how items were ordered in each block.

5.3.2.2.

Materials

Item Set Two Sentences was used. 36 filler sentences were selected from a set created for that purpose, each was paired with a comprehension question of the form “Did the X verb the Y?” and required a yes/no answer. Filler sentences that were followed by a comprehension question in one presentation were not followed by a question in the other presentation. Two lists were created, once for each half of the experiment. Within each list, items were quasi-randomly sorted into 18 block lists, so that each contained 5 sentences made up of 1 or 2 from each experimental category and 1, 2 or 3 filler sentences. Within each block, the first item was always a filler sentence, sentences were otherwise randomly ordered. These two lists were then quasi-randomly resorted into two new lists (1b and 2b) so that the items appeared with different accompanying items in each block, whilst maintaining the structure described above. By combining these lists, four grand lists were created for the experiment as follows: 1a/2a, 1b/2b, 2a/1b, 2b/1a. These were then used for ordering items in the experiment.

5.3.2.3.

Graphic Display & Apparatus

These were the same as for Experiment 5-2, except that the programme was run using DMDX Software (Forster & Forster, 2003) and the monitor refresh rate was set to 60Hz.

5.3.2.4.

Procedure

The instructions were the same as for Experiment 5-2 but were adapted for sentence presentation. Participants completed 9 practice trials, 4 of which were followed by comprehension questions. They then completed the experimental blocks. Each trial began with a fixation cross for 1000ms and remained on screen whilst the sentence was presented aurally. At sentence offset the fixation cross remained for an ISI between 50 – 250ms (in 50ms steps), and the shape was then presented for 100ms. The fixation cross was presented until a response, or to a time out of 3000ms. Comprehension questions were presented until the participant responded, or to a time out of 6000ms. Feedback for the answer given to the question was then presented for 1500ms.

5.3.3. Results

5.3.3.1. Analysis

All participants scored above 80% accuracy for correctly categorising the object, and for answering the comprehension questions. Correct trials with reaction times between 100ms and 1000ms were analysed, this removed 4% of the data. Errors were analysed separately.

5.3.3.2. Reaction Times

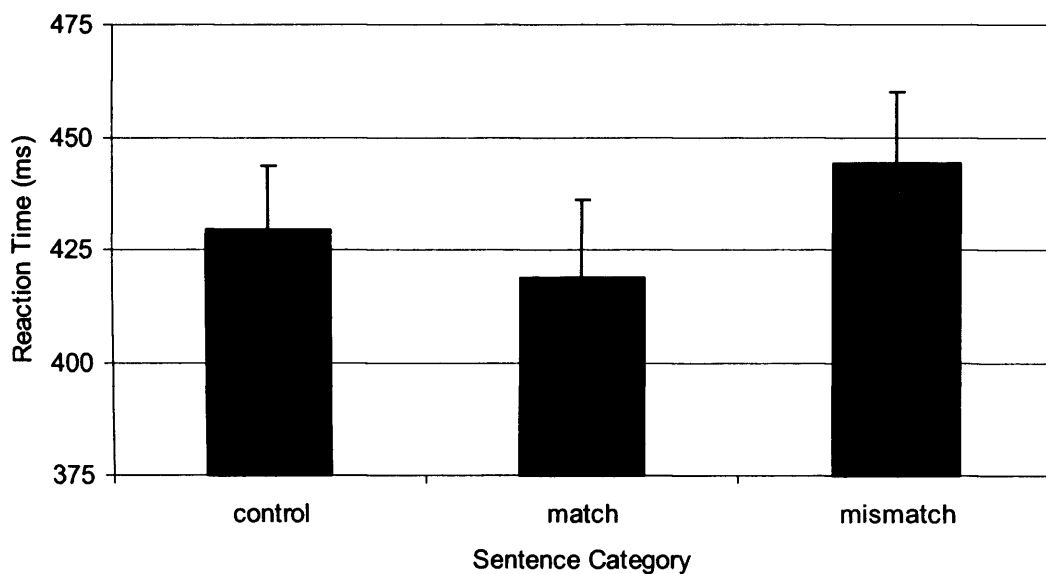
A 3x2 repeated measures ANOVA showed a significant main effect of Sentence Category both by Subjects ($F(1,20) = 11.930$, $p-\eta^2 = 0.374$) and by Items ($F(2,84) = 3.858$, $p-\eta^2 = 0.084$). 2x2 ANOVAs were then used to compare each Sentence Category to each other. The Mismatch category had significantly longer reaction times than Control ($F(1,20) = 7.287$, $p-\eta^2 = 0.267$; $F(1,56) = 4.394$, $p-\eta^2 = 0.073$) and Match categories ($F(1,20) = 31.841$, $p-\eta^2 = 0.614$; $F(1,56) = 6.174$, $p-\eta^2 = 0.099$). Mismatch RTs were on average 14.67ms (SE = 5.43) longer than those for Control and 25.48ms (SE = 4.52) longer than those for Match. The main effect between Control and Match conditions approached significance by Subjects ($F(1,20) = 3.696$, $p = 0.069$, $p-\eta^2 = 0.156$), but not by Items ($F(2) < 2$, $p > 0.2$); the trend was for Match sentences to produce shorter reaction times than Control sentences. The interaction between Sentence Category and Position for Control and Mismatch conditions approached significance by Subjects ($F(1,20) = 3.616$, $p = 0.072$, $p-\eta^2 = 0.153$) but not by Items; the trend was for the Control sentences to produce longer reaction times for shapes appearing in the top of the visual field, and for Mismatch sentences to produce longer reaction times for shapes in the bottom of the visual field. Figure 5-2 displays the mean RTs for each Sentence Category, see Table 5-2 for mean reaction times for each condition.

5.3.3.3.

Errors

The Friedman test showed no significant differences for Sentence Category By Subjects ($X^2_1 < 2$) or By Items ($X^2_2 < 1$). The Wilcoxon Signed Rank Test showed more errors for the Bottom shape location for the Match condition, both By Subjects ($Z_1 = -2.250$) and By Items ($Z_2 = -2.304$).

Figure 5-2: Mean reaction times (ms) for shape categorisation by Sentence Category and Position*



*Error bars are one standard error of the mean

5.3.4.

Discussion

The results showed a clear interference effect for shape categorisation, when the sentences described motion incongruent with the location of the shape, participants were slower to make the categorisation judgement. When the sentence described motion that was congruent with the location of the shape, there was trend for participants to be faster at categorisation than in the Control condition, where no directional information was present in the sentences. This goes against the hypothesis, which was motivated by previous findings, but corroborates what was found in Experiment 5-1, where incongruence between the direction of words, presented in blocks, and the shape location also resulted in slower categorisation. The results are therefore also in contrast to the findings that motivated the hypotheses (Richardson et al, 2003; Bergen et al, in

press), where congruence between the direction/location of objects in sentences produced interference. It is unclear why the results should be in the opposite direction but note that it does agree with studies which show that the eyes go in the same direction as events described by linguistic stimuli (e.g. Spivey et al, 2000). Therefore the results are in line with a conflict when the eyes have to attend to a location that is in conflict with described events. I will return to this important matter in the discussion. These results show that when a concrete location is specified by the sentence, a congruent location speeds categorisation and an incongruent location slows categorisation. This is in support of the idea that attention is influenced by location information, rather than just motion information. The following experiment tests this further by presenting categorisation during the sentence, here, motion information is available but location information is not. Therefore, if location is the important element we should not see the same pattern in reaction times.

The error data showed no differences between the Sentence Categories, but the Match condition showed more errors for the bottom shape location than the top. In other words, Down sentences produced more errors for shapes presented in the lower visual field. This was a reliable difference and is apparently in conflict with the incongruent interference, which would predict more errors at the bottom screen location following Up sentences. However, across the three experiments the only significant differences have been for the Bottom screen location, for the Mixed block in Experiment 1, for Up words in Experiment 2 (although this was not seen by Subjects) and for Down sentences in Experiment 3. Given that there are hypothesised differences between visual processes in the upper and lower visual fields, with the lower visual field more specialised for fine-grained spatial distinctions (Previc, 1990), it is possible that it is more sensitive to errors overall.

5.4. *Partial Sentence Comprehension & Shape Categorisation*

This experiment was the same as Experiment 3 except that the shape categorisation task was performed during sentence comprehension, with the shape appearing at verb offset. This was to test whether the entire sentence, and all the information it conveys, has to be comprehended to produce the interference effect demonstrated in Experiment 3. It is possible that after comprehending the subject noun and the verb this would be enough to interact with visuo-spatial processing. In contrast, the whole event (and therefore the location of the described object) may be necessary to affect visuo-spatial processing. The hypothesis is that if the verb drives the incongruent interference effect, longer reaction times should be produced when the sentences and the shape location are incongruent. If the whole sentence is necessary to produce interference, there should be no effects of congruence.

5.4.1. **Participants**

20 undergraduate students (14 female) volunteered as part of an opportunity sample. The mean age was 20.9 years ($SD = 1.12$). All subjects were Native English speakers and had normal or corrected vision.

5.4.2. **Method**

5.4.2.1. *Design*

The design was the same as for Experiment 2. There were a total of 18 blocks (9 in each half), each of which contained 15 trials. Across the experiment there were an equal number of circles and squares and an equal number of shape appearances in the top or bottom of the screen. Within each block, sentences were selected randomly from 90 experimental and 45 filler sentences. Each filler item was followed by a comprehension question once during the experiment, thus there were questions on 16% of trials. For experimental sentences, the shape was presented at verb offset. For the filler sentences,

the shape was presented at various points throughout the sentence from the subject noun phrase offset up to sentence offset.

5.4.2.2. Materials

Item Set Two Sentences was used. 45 filler sentences were selected from the set created for that purpose, paired with comprehension questions of the same form as Experiment 3. There were an equal number of questions requiring 'yes' and 'no' responses. Two lists were created. In list 1, all the experimental sentences appeared once with the circle or square appearing at the top or bottom of the screen. In list 2, the sentences appeared once with the shape appearing at the opposite location. The shape for each sentence was kept constant across lists, as was the presentation time for the shape (verb offset). For the filler sentences, the shape was kept constant across lists, but the presentation time was not. In list 1, half the filler sentences were followed by a comprehension question; in list 2, the other half of the sentences were followed by a comprehension question. Filler sentences never had shapes appearing at verb offset, instead, shapes were presented at the offset of the subject, preposition/article offset following the verb, direct object/adverb offset or sentence offset. The offset time was altered between the two presentations of each sentence. This was to stop participants recognising that the shape presentation always followed the verb. Offset times were established by editing the sentences at verb offset (or other location) and then measuring the length of the created sentence fragment using a bespoke E-Prime programme that recorded onset and offset times for the fragments. Onset times were subtracted from offset times to establish clip length.

5.4.2.3. Graphic Display & Apparatus

These were the same as for Experiment 2.

5.4.2.4. Procedure

Instructions were the same as for Experiment 2, adapted for sentences. Participants completed 8 practice trials 4 of which were followed by comprehension questions with

feedback for the response. They then completed the experimental blocks. Each trial began with a central fixation cross for 500ms, the sentence was then presented binaurally over headphones. The fixation cross stayed on screen until a specified time, i.e. verb offset for experimental items, various times for filler sentences. The shape was then presented for 100ms at the top or bottom of the screen followed by a central fixation cross until a response or to a time out of 2000ms. On 50% of filler trials the comprehension question followed the categorisation response and was displayed until the participant responded or to a time out of 4000ms.

5.4.3. Results

5.4.3.1. *Analysis*

One participant was removed for having low accuracy on the comprehension questions (<70%). Correct trials with reaction times between 100ms and 1250ms were analysed, this removed 6.7% of the data¹⁰. Errors were analysed separately.

5.4.3.2. *Reaction Times*

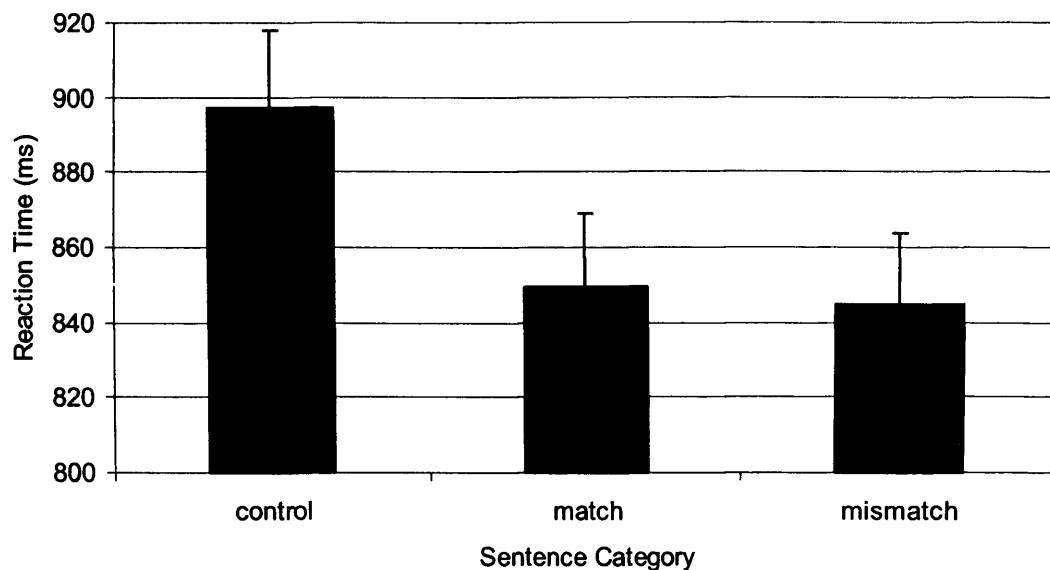
A 3x2 repeated measures ANOVA showed a significant main effect of Sentence Category by Subjects ($F(2,36) = 54.844$, $p-\eta^2 = 0.753$) but not by Items ($F(2) < 1.2$). There was also a significant interaction between Sentence Category and Shape Position by Subjects ($F(2,36) = 13.008$, $p-\eta^2 = 0.419$) but not by Items ($F(2) < 1$). This interaction was driven by Down sentences producing mean reaction times that were 28.97 ms faster than the Up sentences. An interaction resulted when these were analysed as Match and Mismatch categories¹¹ so it will not be analysed further. Planned comparisons showed that the Control category had mean reaction times that were 47.61ms longer than the Match category ($F(1,18) = 55.287$, $p-\eta^2 = 0.754$) and 52.46ms longer than the Mismatch Category ($F(1,18) = 89.335$, $p-\eta^2 = 0.832$). As the interaction between Sentence Category and shape Position was driven by differences for the Down and Up

¹⁰ The 1000ms upper cut off was not used in this experiment as this removed 26% of the data, indicating that participants were taking longer to categorise the shape than in previous experiments.

¹¹ The data was re-analysed with the 3 level factor Sentence Category set as Control, Down and Up. A significant main effect of Sentence Category was found by Subjects ($F(2,36) = 54.916$, $p-\eta^2 = 0.753$) but not by Items ($F(2) < 1$); follow up comparisons showed all Sentence Categories were significantly different from each other (all $F(1) > 23$). Mean reaction times for Control sentences was 896.92ms (SE = 21.03), for Down sentences was 832.40ms (SE = 20.30) and for Up sentences was 861.37ms (SE = 18.72).

sentences, and therefore not informative for the critical congruence comparison, Figure 5-3 presents the mean reaction times collapsed across Position. See Table 5-2 for the mean reaction times for individual conditions.

Figure 5-3: Mean reaction times for Sentence Category



5.4.3.3. Errors

The Friedman test showed no significant differences between Sentence Categories by Subjects ($X^2 = 1.3$) or by Items ($X^2 < 1.5$). The Wilcoxon Signed Ranks test showed significant differences for all Sentence Categories by Subjects and by Items, with more errors for shapes appearing at the Bottom screen location for Control ($Z1 = -1.991$; $Z2 = -2.249$), Match ($Z1 = -2.746$; $Z2 = -3.492$) and Mismatch ($Z1 = -2.315$; $Z2 = -3.082$) conditions.

5.4.4. Discussion

The results showed facilitation for shape categorisation following directional verbs as compared to non-directional verbs in the Control group. When contrasted with the results from 5.3, where incongruent interference was found, this result suggests that the whole sentence has to be comprehended in order for it to influence visuo-spatial attention and show effects of congruence. However, there were also significant

differences between the Down and Up sentences, when not collapsed into Match and Mismatch conditions. The critical difference between the method used here and that in Experiment 3 was that the shape appeared at verb offset, immediately following the comprehension of the verb and *during* the comprehension of the remainder of the sentence. This made the task harder, as evidenced by reaction times being 400ms longer for this task when compared to making the categorisation after sentence offset (see Table 5-2). Therefore, the reaction time differences may be the result of particular sentence properties rather than the spatial and directional information to which the sentences refer. For example, control sentences may have been harder to comprehend and therefore interfered more with the categorisation task than directional sentences. Comparing the directional sentences, Up sentences may have been easier to comprehend than Down sentences producing the same result. Another alternative explanation is that directional verbs provided an implicit cue for the shape appearing at verb offset, priming responses to the categorisation task. Filler sentences and Control sentences had no consistent verb qualities which could be used as cues, and the Fillers ensured that target appearance was not guaranteed after verb offset. Only for the directional sentences was there a stable relationship between semantic content (upwards or downwards 'semantic' motion) and target appearance (at verb offset). It is possible that strategies sensitive to this relationship were established during the experiment, speeding categorisation of the target for directional sentences. These two alternative explanations make it unlikely that sentences with directional content provide some general facilitation for the categorisation task, regardless of congruence. The (embodied) explanation of general facilitation relies on the presence of a directional verb giving a preparatory benefit to the allocation of visuo-spatial attention along the axis of motion implicated by the verb (in this case, vertical). Without the comparison with the horizontal axis (as in Richardson et al, 2003), we cannot establish whether this explanation is correct.

There were significantly more errors for shapes presented in the bottom of the screen for all Sentence Categories. This is in line with the mixed results from the other experiments which show more errors for the bottom location in varying conditions. It supports the inference that the bottom of the visual field is more prone to errors in this task; perhaps as a result of specialisation in the lower visual field that makes it more open to disruption (Previc, 1990).

5.5. *General Discussion*

The experiments demonstrated that both single word and sentence comprehension can influence the allocation of visuo-spatial attention. The first two experiments used single words, presented randomly or in blocks. The third and fourth experiments presented sentences with the categorisation task at sentence offset or verb offset respectively. I will first discuss the single word experiments, then the sentence experiments and finally the conclusions that can be drawn from all four. Table 5-3 summarises the results.

Table 5-3: Descriptive summary of results

Linguistic Stimuli	Results
Blocked single word presentation	Interference from Mismatch and Mixed conditions
Single word presentation in random order	No significant effects
Full sentence comprehension	Interference from Mismatch conditions and a trend for facilitation in Match conditions
Partial sentence comprehension	Facilitation from both Match and Mismatch conditions

The single word experiments showed that single word comprehension does influence visuo-spatial attention when a strong context is created, through presentation of the words in semantically coherent blocks. Participants were slower to categorise a briefly presented shape when they heard a block of single words all referring to a motion direction incongruent with the location of the shape; e.g. when they heard ‘down’ words and categorised a shape in the upper visual field. The interference was similar in magnitude to that produced by blocks of directionally inconsistent items (the Mixed block). When words were presented in a random order with categorisation after verb offset, the directional content of the words made no difference to the speed of categorisation for different visual field locations. This shows that the presentation of single words referring to a direction of motion is not enough to influence visuo-spatial attention, in comparison to the literature which has used nouns referring to objects with a canonical location (Zwaan & Taylor, 2003; Estes et al, in press). Zwaan & Yaxley (2003) found that pairs of words (e.g. attic – basement) were judged more quickly when their spatial relation was congruent with their referents typical location, e.g. attic displayed above basement, than when the spatial relation was incongruent, e.g. basement displayed above attic. Therefore, there was a facilitation when the spatial relation of two concrete referents matched the presentation of their labels; this can also

be interpreted as interference when the labels are presented in the 'wrong' relation. The authors concluded that the words activate an embodied representation of their referents that is facilitated by the congruent presentation of the words. More closely related to the method used in these experiments, the comprehension of single words which refer to objects in a particular location impairs the identification of targets in the congruent location (Estes et al, submitted); for example, reading 'foot' subsequently slowed letter identification for targets at the bottom of the screen. These stimuli refer to concrete objects with locations associated to those objects that are also concrete. In contrast, words referring to motion require an object to produce a concrete referent, the motion itself is abstract without a moving subject. As demonstrated by Bergen et al, intransitive motion sentences which contain only a motion verb and a moving subject do affect visuo-spatial attention, producing congruent interference; e.g. when the sentence refers to the upwards motion of an object, participants are slower to categorise an object in the upper visual field. The fact that a concrete event is necessary to affect visuo-spatial attention also supports Bergen et al, as they found no effect of congruence for abstract and metaphorical motion events (e.g. "The prices rose" or "Her temper rose"). Therefore, when the semantic representation refers to concrete events that have real-world physical correlates, it can influence the allocation of attention to a visuo-spatial location.

In contrast to the single word presentation, blocked presentation produced incongruent interference. The blocked presentation makes it probable that participants were aware of their directional content, and this in turn would affect conscious and unconscious strategic processes. Therefore, it is not possible to say that the effects are an automatic consequence of single word comprehension. However, it is still interesting to discuss the results as the effect was one of incongruent interference, contrasting with the existing literature which shows congruent interference (Estes et al, submitted; Bergen et al, in press). For blocked word presentation, the incongruent interference was equivalent to the interference when the directional information was inconsistent, during mixed blocks. Thus, having directional information opposed to the visuo-spatial location is no more problematic than when that directional information is unreliable/inconsistent. When the directional information was congruent to the visuo-spatial location, this did not interfere with categorisation more than words with no directional content, but it did not interfere less than incongruent or inconsistent directional content. These effects built up over the course of the block so attention was

biased over time, rather than by individual words. So, when directional content is consistently in opposition to the location of a target object, or inconsistent and therefore either unreliable or confusing, the allocation of visuo-spatial attention is impaired.

The third experiment showed that incongruent interference was also present when categorisation followed sentence comprehension. This is in line with the above conclusion, that a concrete referent is necessary for comprehension to affect visuo-spatial attention. It also reinforces the finding of incongruent interference for Experiment 1, and is in contrast to the existing literature. Both Richardson et al (2003) and Bergen et al (in press) found that the comprehension of sentences referring to a direction of object motion, or to objects with a canonical location, interfered with the categorisation of a target in a congruent visuo-spatial location. This effect had been explained by the 'image schema' or visual imagery produced by comprehension, which interferes with target detection in the same way as conscious visual imagery (Perky, 1910), through the direct engagement of visual processing systems. It has also been argued that the similarity between the detected object and the referent object makes a difference (Estes et al, submitted). If the objects are not similar, e.g. the word 'head' and then detection of a letter X in the upper visual field, interference will result. However, if the word 'head' was followed by detection of a head-like object in the upper visual field, facilitation would result. This has yet to be tested. However, our data show that it is the incongruent condition that produces interference and thus these findings do not fit with explanations from visual imagery or object similarity. Our data is more easily accounted for by comprehension influencing the allocation of attention to a location in space, for example through influencing eye-movements, rather than affecting the perception of the visual scene. In line with this, there is some evidence that narratives describing events in a particular dimension (up, down, left and right) elicit eye-movements in the congruent direction (Spivey & Geng, 2001; Spivey et al, 2000). If language directs attention to particular locations (e.g. Richardson & Spivey, 2000), then the intuitive prediction is that performing a task at a different location should be impaired; this is what we find. It is unclear why our results differ from the previous literature, however, these studies did not use a control condition, so it is unclear what the baseline performance is during the task. A good starting point would be a within subjects experiment that combines the sentence stimuli from all available studies that have used the shape categorisation task. It is still possible that the

construction of sentences contributes to the different results, although it is not clear what variables would produce opposing results, since all were created to be directional.

The fourth experiment presented the object for categorisation at verb offset; both congruent and incongruent directional sentences were faster than the control condition. It is possible that differences in the ease of comprehending Down and Up sentences as compared to Control sentences produced these changes. It is notable that reaction times for Experiment 4 were almost twice as long as those for the other experiments, supporting the increased difficulty of active comprehension concurrently with target detection. It is also possible that directional verbs facilitated categorisation, due to the consistent relationship between a directional verb and the appearance of a target at verb offset. Further experiments would need to be carried out to establish whether directional verbs provide an advantage over control verbs because of implicit cuing, or because of directional content priming a general axis of motion. The same experiment could be completed with filler items that have target appearance at verb offset, meaning there is no ambiguity in the time of target appearance. Any differences that emerge between Control and directional sentences could then be more securely attributed to influences of directional verbs on the allocation of visuo-spatial attention. In order to establish whether a general direction of motion can be primed, there would need to be a comparison between targets presented on the vertical and horizontal axis (Richardson et al, 2003). If we account for this result through effects on the allocation of attention or eye-movements, we can hypothesise that facilitation for a broad dimension may be present before specific information about a complete concrete motion event can interfere with attention. One clear prediction would be for an eye-movement study: we would expect to find differing effects on the latency of saccades to a target following the comprehension of partial or complete sentences.

Taken together, the results show that influences of comprehension on visuo-spatial attention are subtle and variable. The tentative conclusion is that concrete spatial information is needed before language can affect visuo-spatial attention, in line with previous findings (Bergen et al, in press). When language has a visual-physical referent, semantic representations that are activated upon comprehension engage with the visuo-motor system sufficiently to influence the allocation of attention in visual space. Crucially, the allocation of attention is modulated according to the location and motion of the physical referent. In contrast to previous findings, but in line with

predictions from eye-movement studies, our data show that target categorisation at incongruent locations is slowed. This supports the inference that directional language directs attention to congruent locations. However, our data also shows that processing at congruent locations is not facilitated; at least as measured by reaction times. Experiments which make the central task harder (e.g. presenting object detection rather than categorisation) and allow other measures to be extracted (e.g. d' and c , see Appendix 2) would make the level of influence clearer.

These results are in line with both strong and weak versions of embodiment, although since the results can be explained through mechanisms which direct attention they can be partially accounted for by strategic or higher order influences. The experiments tapped visuo-spatial attention implicitly (the task was categorisation of the target rather than judgement about where it appeared) so any influence of language was automatic. Blocked single words were able to interfere with attention over time but sentences produced immediate influences, specifying a concrete location/axis that impacted upon visuo-spatial attention. Thus, at the very least, when semantic information is available and informative it is automatically recruited during the direction of attention. Since the data is in conflict with previous results, it does not support the interpretation that the object was visually simulated, resulting in congruent interference. Therefore, they go against this particular explanation from strong embodied theories. Instead, the results support the subtle integration of semantic information into visuo-spatial attention. This can be explained by indirect top-down influences. For example, language could bias covert expectations towards the motion of an object in a particular direction. In order for this to be the case, the expectations used by attention are biased by *particular* qualities of semantic content (e.g. a concrete location) as well as being more globally influenced by general semantic content (e.g. a general direction). In essence, this still means that visuo-spatial attention is implicitly and automatically manipulated by the specific contents of semantic representations. The results are in line with weak embodiment where connections between semantic and perceptual processes are direct but mediated, in this case by whether or not semantic content specifies information that is potentially (but not always) useful for attention.

6. Comprehension on Visual Motion

In this chapter several experiments are presented which look at the comprehension of sentences and single words on the perception of visual motion. The previous experiments (see Chapter 5) used the upper or lower visual field as the perceptual independent variable and the visual stimuli were static. The linguistic stimuli describe *motion* so visual field location approximates the destination of moving objects, but does not tap motion processing. As already discussed (see Section 5.5) these experiments allow inferences about the effect of comprehension on the allocation of visuo-spatial attention; it is not entirely clear what perceptual processes are affected by the linguistic stimuli. By using dynamic stimuli we can look at how comprehension of linguistic stimuli describing motion affects the perception of visual motion. Aside from the better match between the linguistic and perceptual stimuli, motion processing has been extensively studied (e.g. Raymond, 2000; Blake, Sekuler & Grossman, 2004), therefore we can use established perceptual stimuli and make inferences about specific perceptual processes. Random Dot Kinematograms (RDKs) were chosen as the dynamic stimuli; they present first-order motion by animating white/grey dots on a black background. First order motion results from a change of energy in the visual field, such as that produced by a moving object. A percentage of the dots move randomly and some move coherently (i.e. in one direction). The percentage of dots moving coherently varies the strength of perceived coherent motion. They are a well established tool used widely in visual perception research to access low level motion perception (e.g. Braddick, 1973, 1980; Britten, Shadlen, Newsome & Movshon, 1993; Raymond, 2000); as such they have a sound base from which to interpret the effects that are found with their use.

In Pilot Experiments (section 6.1), blocks of words or sentences were presented to participants making speeded judgments about the direction of a dynamic pattern of dots. However, as participants were making a directional judgment, the perception of motion was conflated with the preparation of a response. Following this, experiments used motion detection, a standard psychophysical task in which participants judge whether an RDK has random or coherent motion. Critically, the coherent motion is at the participant's threshold of conscious perception so the task is difficult. This allows us to use Signal Detection Theory to analyse the results; extracting measures of perceptual sensitivity (d'), internal decision criterion (c) and response bias (β) given a particular criterion. These measures separate the low level perception of motion (d') from the

higher level decision and response processes (c , β), allowing a clearer interpretation of where and how comprehension of language referring to motion affects motion perception. For a brief review of the principles of Signal Detection Theory see Appendix 2.

6.1. *Pilot Experiments*

In these experiments, participants were presented with blocks of single words (Pilot a) or sentences (Pilot b) for passive comprehension whilst they judged the direction of a moving dot pattern. Blocked presentation of single words and sentences that refer to the same direction of motion should create a powerful directional context. These studies were completed at the Science Museum, therefore they had to be kept short (approximately 5 minutes). The hypothesis is that if the meaning of the words and sentences affects the perception of visual motion, we should see effects of congruency on reaction times and error rates. The hypothesis is non-directional since the literature shows mixed results. For example, studies using picture judgments have found faster reaction times when linguistic and perceptual information is congruent (e.g. see Section 2.1.1.3 or 2.1.2.2). However, when motion is presented simultaneously with sentences, faster reaction times to judge sentences were found for incongruent rather than congruent conditions (Kaschak et al, 2005). Therefore it is unclear whether results will be in line with the majority of effects found for comprehension on perception (congruent facilitation) or the limited evidence found for perception on comprehension (incongruent facilitation).

6.1.1. **Participants**

Pilot a: 56 native English speakers took part in the experiment voluntarily. Participants run at University College London were paid for their time. There were 23 females. The mean (standard deviation) age was 27.6 years (10.7).

Pilot b: 37 participants took part, 16 female and 21 male. Mean age was 32.4 years (SD = 11.34).

6.1.2. **Method**

6.1.2.1. *Design*

A 2x4 repeated measures design was used. The independent variables were the direction of motion of the random dot kinematograms (upwards, downwards), hereafter RDKs; and the Word (Pilot a) or Sentence (Pilot b) Category by block (Up, Down, Mixed and Control). The dependent variables were the reaction time and errors for categorising the direction of motion of the RDKs (ms). There were a total of 4 blocks, each of which contained 24 RDK motion categorisation trials for Experiment 1 and 52 RDK trials for Experiment 2. Within each block there were an equal number of up and down motion RDKs. The presentation of blocks and of up or down RDKs within the blocks was fully randomised.

6.1.1.1. *Materials*

Pilot a: Item Set One Single Words. The same soundfiles as used in Experiment 5.1.

Pilot b: Item Set One Sentences. To avoid order effects, four orders for each group (Up, Down and Control) were produced and concatenated into a single WAV sound file (16 bitrate, mono, 44khz sampling rate). Four mixed groups were created by randomly selecting half the items from the up group and half from the down group and then concatenating them into a single sound file, such that across all the mixed groups each item appeared twice. The concatenated sound files were 44 seconds long, with two second of silence at the beginning of the soundfile followed by one sentence every 2625 ms.

During each experimental block 2 continuous files were consecutively presented, each with a different ordering of items. Files for the Mixed blocks were selected so that all the Up and Down items were heard once. Therefore, participants heard each directional item 5 times across the whole experiment.

6.1.1.2. Graphic Display

Each RDK consisted of 150 white dots (width = 0.067° visual angle) against a black background covering a square area (width = 5°) centred at fixation point. During each trial, a randomly chosen subset of 100 dots were vertically displaced up or down by 0.67° steps in 5 consecutive frames (total motion time = 85ms; speed = $4^\circ/\text{s}$). A small number of frames and short duration has been used successfully elsewhere in the motion detection literature (e.g. Silvanto, Lavie & Walsh, 2005). The rest of the dots were repositioned randomly from one frame to the next. To prevent participants becoming familiar with the sequences, 15 up and 15 down sequences of 5 frames were generated. They were selected and presented randomly within each block.

6.1.1.3. Apparatus

The experiment was run using PC computers (Intel Pentium II) using E-Prime 1.0 Software (Schneider, Eschman, & Zuccolotto, 2002). A standard CRT monitor was used with a refresh rate of 60 Hz. Participants responded using standard 'qwertyuiop' keyboards. Chin rests were also used to minimise head movements, these were placed at a distance of 57cm from the computer screen.

6.1.1.4. Procedure

Participants were instructed to respond to the RDK as quickly and accurately as possible by pressing the "up" key if the motion was up, and the "down" key if the motion was down (stickers were placed over the A and L keys that showed an upwards or downwards arrow). Assigned response keys were balanced across participants. Participants completed 20 practice trials of the motion direction judgement without words/sentences played over the headphones. They were told to ignore the words/sentences presented over the headphones during the experimental blocks. Presentation order for the WAV files (Up, Down, Control and Mixed) was fully randomised.

The WAV file lasted for 18 seconds (Pilot a) or 44 seconds (Pilot b). It was repeated once in each block so the total playing time was 36 seconds (Pilot a) or 88 seconds (Pilot b). The first WAV file began playing at the block onset, therefore the first item was heard 2000ms after the start of the block. The second WAV file began playing half way through the block after the 12th (Pilot a) or 26th (Pilot b) categorisation trial and after the first WAV file had finished. At the beginning of each block a blank screen was presented for 1000ms followed by a central fixation cross presented for 1000ms. During each trial, a central fixation cross appeared for a randomised delay of between 800-1000 ms (with 50ms increments in between). An RDK was then presented for 85ms at the centre of the screen. A blank screen was presented until a response, or to a time out of 3000ms. At the end of each block a grey screen was presented instructing the participant to take an optional break.

6.1.2.2. Analysis

Pilot a: Participants who scored lower than 85% accuracy for correctly categorising the RDK motion were excluded from the analysis; this removed 27 people from the analysis who were replaced with participants from the UCL subject pool. Reaction times above 50ms and below 800ms were excluded from the analysis, this removed 9.2% of the data. Errors were analysed separately.

Pilot b: Participants who scored lower than 80% accuracy for correctly categorising the RDK motion were excluded, this removed 5 participants. Reaction times above 100ms and below 800ms were excluded, this removed 9.7% of the data. Errors were analysed separately.

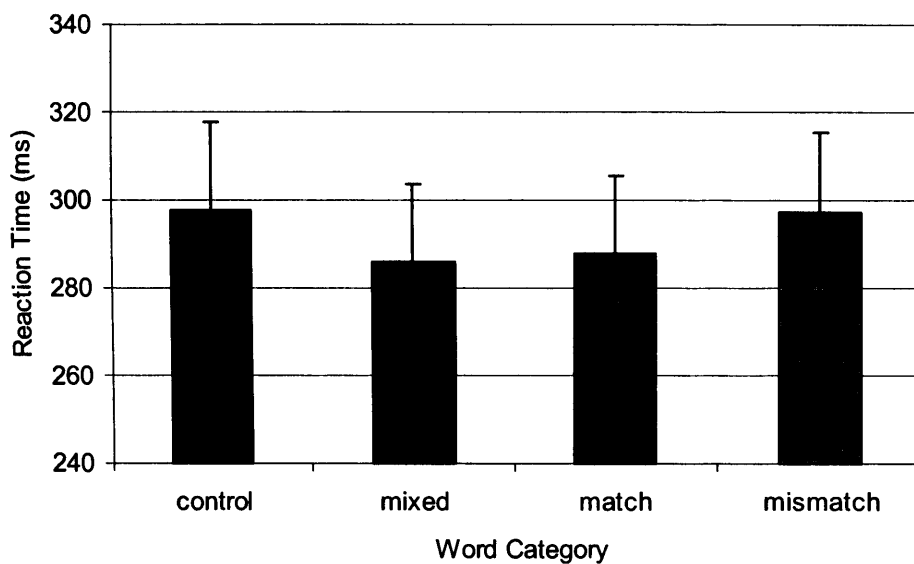
6.1.3. **Results**

6.1.3.1. Reaction Times

A 2 x 4 repeated measures ANOVA compared reaction times for the Control, Mixed, Match and Mismatch blocks across motion directions (upwards and downwards). Table 6-1 presents the reaction times for both pilot experiments.

Pilot a: There was no significant main effect of Word Category. Planned comparisons showed that the Mismatch block was marginally slower than the Mixed block ($F(1,55) = 3.560, p = 0.068, p-\eta^2 = 0.061$) and significantly slower than the Match block ($F(1,55) = 5.094, p-\eta^2 = 0.085$). As there was no significant main effect or interaction for Motion Direction, Figure 6-1 presents the reaction time data collapsed across Upwards and Downwards motion.

Figure 6-1: Mean reaction times for Pilot a for each Word Category



Pilot b: There was no significant main effect of Sentence Category or differences in the planned comparisons. There was a marginally significant main effect of Motion Direction, with Upwards motion producing longer reaction times than Downwards motion ($F(1,30) = 4.104, p = 0.052, p-\eta^2 = 0.120$).

Table 6-1: Mean reaction times for both Pilot experiments for each condition
(Standard deviation in brackets)

Experiment	Condition							
	Control		Match		Mismatch		Mixed	
	Down	Up	Down	Up	Down	Up	Down	Up
Pilot a Blocked Single Word	292.85 (133.33)	302.96 (162.29)	290.59 (123.34)	284.63 (142.54)	296.45 (140.61)	298.47 (131.71)	282.83 (132.80)	288.75 (137.56)
<i>Category Mean</i>	297.91 (147.93)		287.61 (132.72)		297.46 (135.62)*		285.79 (134.62)	
Pilot b Blocked Sentence	265.56 (75.37)	285.35 (82.14)	262.50 (74.61)	275.34 (103.98)	272.64 (98.76)	268.51 (73.85)	261.87 (75.92)	265.87 (79.92)
<i>Category Mean</i>	274.28 (79.80)		267.39 (90.60)		267.93 (86.62)		260.47 (79.20)	

*Significantly different to Match condition, $p < 0.05$

6.1.3.2. Errors

Pilot a: There were no differences in Error rates.

Pilot b: The Wilcoxon Signed Ranks Tests showed that the Control condition had marginally more errors for Downwards motion as compared to Upwards motion ($z = -1.854$, $p = 0.064$).

6.1.4. **Discussion of Pilot Results**

The pilot experiments showed that blocked single word comprehension affected the judgement of motion direction. Words congruent to the direction of motion produced shorter reaction times than words that were incongruent with the visual motion. Thus, the blocked presentation of single words referring to motion in a particular direction creates a linguistic context that is powerful enough to affect reaction times for direction judgements. In Experiment 5.1 (which looked at visuo-spatial attention), both Mismatched and Mixed blocks produced longer reaction times for shape categorisation relative to Control (but not Match) blocks. In this experiment, direction judgements in Mixed and Match blocks had faster reaction times relative to the Mismatch condition. The results are not clear cut across all conditions, in particular it is difficult to sensibly interpret what is happening in the Mixed blocks, however, the data suggest facilitation

during Matched and Mixed conditions for blocked single word presentation. Thus, when directional semantic information agrees, or partially agrees, with the location of a visuo-spatial target, participants are slightly faster to categorise the target object.

Conversely, blocked sentence presentation did not produce any consistent effect on reaction times, indicating that it failed to produce a consistent context. It is probable that each sentence, although describing motion events in the same direction, were still divergent enough that a consistent context was not created. Each sentence described idiosyncratic situations (e.g. 'The woman raised the curtain', 'The rocket flew to the moon') and the comprehension of each sentence most likely removed the effects of comprehending the previous sentence since they did not form a connected narrative. Thus the motion events described in the sentences were not continuous, being an isolated event described by each sentence. Given that presentation of the RDKs was unsynchronised to the presentation of the sentences, appearing at various points during a sentence (e.g. during the Subject, Verb or Object), it is perhaps not surprising that blocked presentation of disjointed sentences did not produce any consistent effects on direction judgements. It is possible that if RDKs were synchronised to the verb in a sentence we would have seen some congruency effect; elsewhere it has been shown that effects of motion information are localised to this region (Zwaan & Taylor, 2006).

The main methodological problem with this task is that participants made explicit judgements about the direction of motion, therefore response options were 'up' or 'down'. It is not possible to separate effects on perception from effects on response. The linguistic context may be influencing the ease with which directional motion is perceived, as predicted by embodied theories, or it may be affecting the speed with which participants can prepare and execute a response once perception is complete. This is a more trivial explanation and fits with stimulus response compatibility phenomena such as the Simon effect, in which incongruence between a stimulus location and response hand (e.g. left location and right response) produce longer reaction times than congruence (e.g. left location and left hand). Simon effects have been found for stationary targets, like RDKs, with incongruent or congruent (i.e. leftward or rightward) motion within them (Bosbach, Prinz & Kerzel, 2004). Therefore, it is entirely possible that the motion words acted as interfering stimuli during response preparation. In addition, because the single word stimuli are blocked it is possible that participants are aware of the semantic content of the words. The observed effects could

be mediated or caused by particular strategies or conscious processes, rather than an elementary effect of semantic content on visual processes. To avoid these problems, subsequent experiments turned to psychophysical methods and motion detection paradigms.

6.1.5. **Motion Detection**

Psychophysical methods used to explore motion perception are of special interest for a number of reasons. First, we can manipulate the direction of motion independently of response. Rather than making directional judgements participants judge if coherent motion is present or absent (motion detection). This reduces the likelihood that Simon-effect like influences on response preparation and execution are confounded with the dependent measures. Second, these methods have been extensively used in psychophysics to explore specific aspects of visual motion processing. Random Dot Kinematograms are known to engage a set of well characterised processes in first order motion perception. These stimuli have been shown to activate specific cortical areas, typically the MT+ complex, and in using them we focus on direction selective motion processing (e.g. Blake, Sekuler & Grossman, 2004). Third, by employing psychophysical thresholding techniques we can explore *where* effects of comprehension on motion perception are manifest. Analysis using Signal Detection Theory (SDT) allows us to separate early perceptual sensitivity (as measured by d'), decision processes (c and β) and overall response (reaction times) (see Appendix 2). In other words, by using these stimuli we can localise a *specific visual process* and interpret at what level comprehension of motion words affects the perception of visual motion. Finally, by using motion detection we avoid some of the immediate problems of strategic processing possible with blocked word presentation. If motion words affect only decision or response preparation we should not see effects on low level perception as measured by d' .

6.2. *Blocked Word Comprehension & Motion Detection*

Participants performed a coherent motion detection task with stimuli at their threshold coherence level. RDKs displayed coherent motion upwards or downwards or incoherent (random) motion. Whilst performing this task, participants heard blocks of single words over headphones. Two experiments have shown that blocked single word presentation creates an effective directional context (5.1 and 6.1). We used SDT to assess whether and at which level language comprehension affects motion perception. This analysis allowed us to separate the effects of comprehension into (i) those at the perceptual level (Sensitivity); (ii) those at the decision level (Criterion); and (iii) those at the overall response level (Reaction Time). If language comprehension only acts upon higher order processes we would predict some change in the decision variables, but not in sensitivity. If comprehension activates representations that permeate to the earliest stages of sensory processes, we would predict a change in sensitivity. In both cases we would predict reaction time changes that reflect both influences, and may also reflect conscious and strategic processing of the linguistic input. The hypothesis is non directional, although based on our previous experiments we might expect interference effects when the semantic and perceptual motion are incongruent (i.e. lower accuracy, longer response times).

6.2.1. **Participants**

20 Native English speakers (8 male) with a mean age of 25.6yrs (SD = 6yrs) were recruited from the UCL Psychology Department Subject Pool in return for monetary compensation of £6 per hour. All had normal or corrected to normal vision.

6.2.2. Method

6.2.2.1. Design

A 2x3 repeated measures design was used. The independent variables were the direction of motion of the RDK (upwards or downwards), and the Word Category (Control, Match and Mismatch). The dependent measures were the accuracy of detecting motion on each trial, and the reaction time (milliseconds) to make the judgement. Participants attended two sessions separated by 3-7 days. Within each session, direction of detected motion (upwards or downwards) was kept constant with a balanced order across participants. Participants first completed a threshold estimation routine that established the percentage coherence at which the participant detected motion with 81% accuracy (see Threshold Estimation for details).

For the Motion Detection task there were 2 runs of 16 blocks of 20 motion detection trials comprised of 10 coherent motion and 10 random noise trials presented in randomized order. Each run started with 4 practise blocks with no words followed by 12 word-blocks during which participants heard words presented over the headphones. The 12 blocks consisted of 4 different item orders for the 3 Word Categories; presentation order was pseudo-randomised using a latin squares design so that the same Category was not presented in consecutive blocks.

6.2.2.2. Materials

Item Set Two Single Words was used. There were three wordblock conditions: Up, Down, and Control. Four item orders for each group were produced and concatenated into a single WAV sound file. The sound files were 32 seconds long, a 2 second silence at the beginning followed by one word onset every second.

6.2.2.3. Graphic Display

Each RDK consisted of 1000 grey dots (width = 0.04° visual angle) against a black background covering a square area (width = 8°) centred at fixation point. During

each trial, a randomly chosen subset of the dots were vertically displaced up or down by 0.25 degree steps in 13 consecutive frames (total motion time= 150ms; speed = 21 °/s). The rest of the dots were repositioned randomly from one frame to the next. Coherently moving dots reaching either end of the display area were repositioned on the other side for the next frame. A central fixation square (width = 0.2 deg) was displayed throughout the experiment. Stimuli were generated using Cogent toolbox (www.vislab.ucl.ac.uk/Cogent/) for MATLAB (Mathworks, Inc) and presented at 60Hz using a TFT display (resolution = 800×600, 15' NEC Multiscan LCD1510+).

6.2.2.4. Apparatus

Stereo closed headphones were used to present the audio files. The experiment was conducted in a sound attenuated booth. Participants placed their chin on a rest 53cm from the computer screen.

6.2.2.5. Procedure

Threshold Estimation

Motion detection thresholds were estimated for each participant using constant stimuli and a 2-interval forced choice (2-IFC) design. In each trial, 2 RDKs - one containing a varying percentage of coherent motion and the other random noise only - were sequentially presented and the participant indicated which interval contained coherent motion. Participants pressed the left-arrow key for the first pattern, and the right arrow key for the second pattern. The coherent motion signal strength varied between 0.1-20% in 9 equal steps with each step repeated twice per block. There were 16 practise blocks, comprised of 20 trials, followed by 8 experimental blocks. In each trial, after a random delay between 500-1000ms, the first RDK was presented for 150ms, followed by a random interval between 500-1000ms, and the second RDK also for 150ms. There was no time out for response. Performance was quantified by fitting a Weibull function to the behavioural data using the `psignifit` toolbox version 2.5.6 for Matlab (<http://bootstrap-software.org/psignifit/>) which implemented the maximum-likelihood method described by Wichmann and Hill (2001). Threshold for detection was defined as the percentage of coherent motion for which the fitted psychometric function predicted

81% accuracy (50% being "chance level" performance in a 2-interval forced choice task). In each session, threshold estimation was completed at the beginning of each run. If the second threshold did not agree with the first, the lower of the two was used for the second run unless participants reported that they were experiencing significant trouble or fatigue

Motion Detection

Participants were instructed to ignore the words and concentrate on detecting the motion in the RDK patterns. WAV files began playing at the start of each experimental block and a fixation cross was presented for 2000ms (so that words were not presented before motion detection trials). In each motion detection trial, after a random delay of 500-1000ms, an RDK was displayed for 150ms followed by a time out of 1500ms for the response. Participants detected the presence of coherent motion in the RDKs by pressing the left arrow key for coherent motion and the right arrow key for random noise.

6.2.2.6. Analysis

The presentation of words and RDKs was unsynchronised and we expected semantic effects to build up over the block (e.g. Vigliocco, Vinson, Damian & Levelt, 2002), therefore the first two RDK trials from each block were removed from the analysis. Results for each condition were collapsed across runs. Reaction time data was only analysed for correctly detected coherent motion trials; values over 3 standard deviations from the mean were removed for each participant, removing 1.5% of the data. For SDT analysis, the percentage of correct detections on motion trials (hits) and incorrect detections on random motion trials (false alarms) were calculated. These values were used to calculate a measure of sensitivity (d'), and a measure of internal response criterion (c). Other SDT measures were calculated (e.g. β) but did not show significant changes, so are not reported here. Calculations were performed in Excel using equations taken from Stanislaw & Todorov (1999). 2x3 Analyses of Variance with planned tests for simple effects were then carried out on reaction times and SDT statistics.

6.2.3. Results

6.2.3.1. *d'*

Perceptual sensitivity for motion detection was significantly modulated by comprehended words. There was a significant main effect of condition ($F(2, 19) = 4.539, p-\eta^2 = 0.193$). Tests of simple effects revealed a marginal difference between Control and Mismatch conditions ($F(1, 19) = 3.761, p = .067, p-\eta^2 = 0.165$) and a significant difference between Match and Mismatch ($F(1, 19) = 7.851, p-\eta^2 = 0.292$), such that d' was lower when the motion direction and wordblock were incongruent (see Figure 6-2 & Table 6-2).

6.2.3.2. *Criterion*

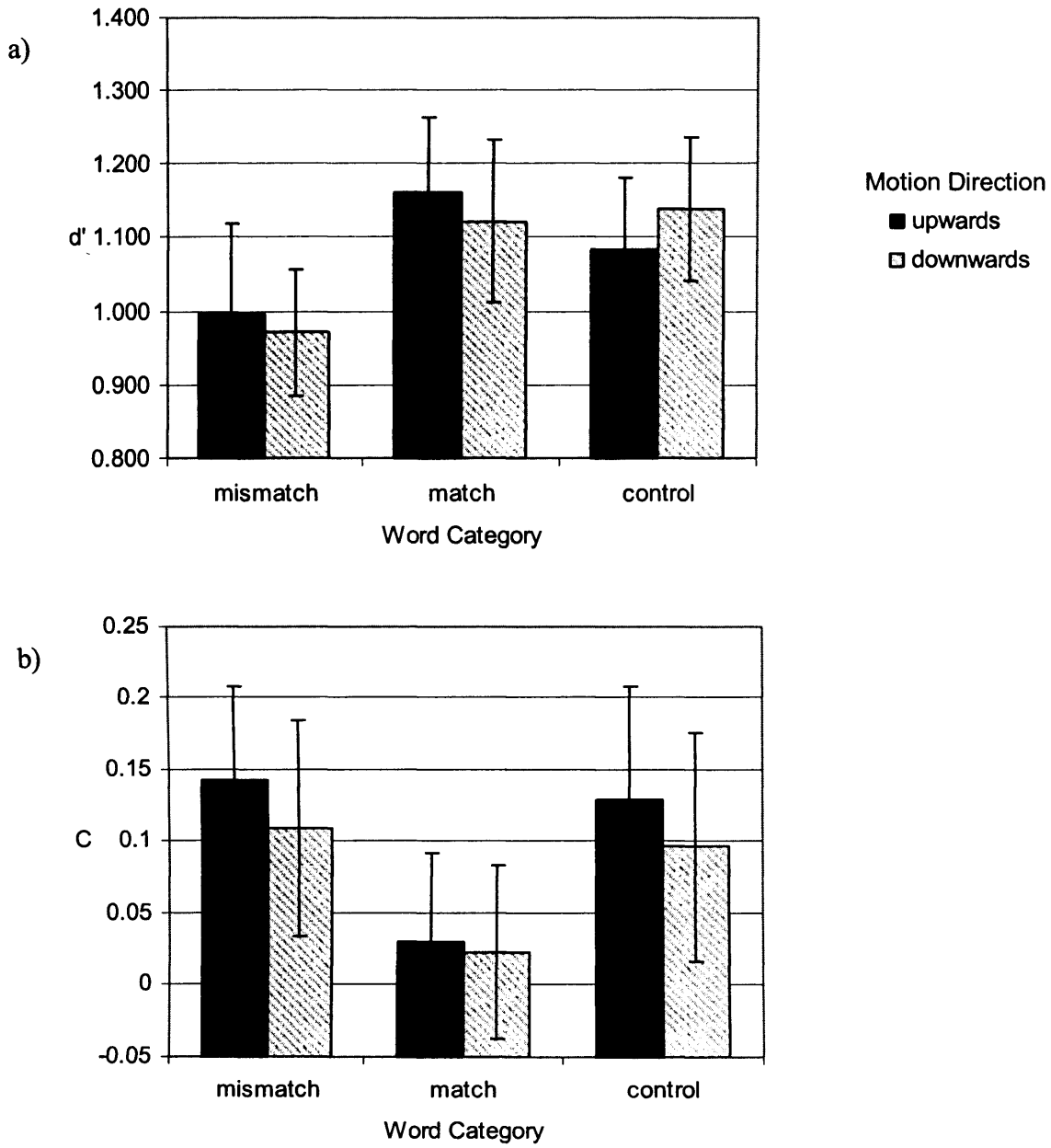
Tests of simple effects revealed a marginal difference between Control and Match ($F(1,19) = 3.250, p = .087, p-\eta^2 = 0.146$) and a significant difference between Match and Mismatch ($F(1,19) = 5.145, p-\eta^2 = 0.213$). Criterion was lower when the wordblock and motion direction were congruent (see Figure 6-3 & Table 6-3).

6.2.3.3. *Reaction Times*

There was a significant interaction between condition and RDK Motion ($F(2 19) = 4.927, p-\eta^2 = 0.206$). Tests of simple interactions showed a significant interaction between Control and Mismatch ($F(1,19) = 6.861, p-\eta^2 = .265$) and Match and Mismatch ($F(1,19) = 7.372, p-\eta^2 = .280$), in the Mismatch condition RTs were raised for upwards motion and lowered for downwards motion¹². See Table 6-5.

¹² A 2x2 ANOVA comparing Word Category (Down, Up) and Motion Direction (Upwards, Downwards) showed a significant difference between Down and Up words, such that reaction times were 13ms faster for Down words as compared to Up words ($F(1,19) = 7.372, p-\eta^2 = 0.280$).

Figure 6-2: (a) Mean d' and (b) C by Word Category and Motion Direction.



6.2.4. Discussion

The results showed that listening to motion words affected people's sensitivity to motion, their internal decision criterion and speed when performing a motion detection task. A decrease in d' indicates impaired sensory level stimulus processing resulting in a diminished separation of signal and noise distributions, hence, poorer perceptual sensitivity (Wickens, 2002). Here, we found that sensitivity was worse when the motion direction and word block were incongruent. However, when listening to non-motion words (Control) or motion words congruent with the detected motion, participants did not show any significant changes in sensitivity. Therefore, the representation that is activated upon comprehension creates specific noise and interferes with the detection of the incongruent motion signal, reducing sensitivity. When the motion signal and wordblock are congruent, no such interference occurs and the motion can be detected as well as in the control condition. This provides clear support for semantic processing penetrating to low-level sensory perception, and is also in line with semantic content that is a perceptually grounded representation of motion.

We also found that participants' Criterion (as measured by c but not Beta) was differentially modulated by word blocks such that a lower Criterion was adopted by the participants in blocks where the motion direction and word block were congruent. Criterion is a measure of the internal decision threshold set by the participant, if a stimulus is above threshold the participant will respond positively (that they perceive coherent motion), if a stimulus is below threshold the participant will respond negatively (random motion). By shifting the Criterion downwards, the participant is being less conservative. Thus, participants were more likely to decide that coherent motion was present when the detected motion and the word block were congruent. This effect looks like response priming; in which congruence between stimuli and response facilitate decision and response processes. This is in partial agreement with Pilot 6-1a, in which congruent (or partially congruent) conditions reduced reaction times. However, in this experiment facilitation did not however extend to reaction times (probably because of the response required to the target, see section 9.3), instead reaction times were raised for upwards motion and lowered for downwards motion. It is possible that downwards motion facilitated the pushing down of buttons, a weak version of the Simon effects discussed above (see section 6.1.4). Therefore, there is a separation between the effects that the words have on perceptual sensitivity (d'), in

which we see interference during incongruent conditions; decision processes (making decisions less conservative under congruent conditions) and response preparation/execution (making motor responses faster when the hand action is congruent with the direction of the words).

Table 6-2: d prime data for experiment 6.2- 6.5 by Condition (Control, Match, Mismatch) and Motion Direction (Down, Up)
(Standard deviation in brackets)

Experiment	Control		Condition		Mismatch	
	Down	Up	Down	Up	Down	Up
6.2 Blocked Word	1.14 (0.43)	1.08 (0.44)	1.12 (0.49)	1.16 (0.50)	0.97 (0.38)	1.00 (0.55)
<i>Category Mean</i>	<i>1.11 (0.43)</i>		<i>1.13 (0.47)</i>		<i>0.98 (0.47)*</i>	
6.3 Single Word	0.93 (0.36)	0.70 (0.24)	0.82 (0.27)	0.64 (0.31)	0.87 (0.36)	0.65 (0.28)
<i>Category Mean</i>	<i>0.82 (0.26) †</i>		<i>0.73 (0.23) †</i>		<i>0.76 (0.23)</i>	
6.4 Sentences	1.11 (0.55)	0.94 (0.41)	1.29 (0.54)	0.95 (0.49)	1.16 (0.59)	1.07 (0.37)
<i>Category Mean</i>	<i>1.03 (0.35) †</i>		<i>1.12 (0.39) †</i>		<i>1.11 (0.41)</i>	
6.5 Blocked Single Word	0.85 (0.48)	0.78 (0.40)	0.84 (0.40)	0.85 (0.37)	0.66 (0.44)	0.90 (0.54)
<i>Category Mean</i>	<i>0.81 (0.35)</i>		<i>0.84 (0.32)</i>		<i>0.78 (0.43)</i>	

* significant main effect driven by this condition being different to the other two

† significant difference between these two conditions in planned comparisons

This task has shown that language comprehension can influence several levels of task performance. Neither of the observed SDT patterns registered in the reaction time data, which reflect the whole of processing from perception, through decision to response execution. In contrast, SDT measures separate the levels in a decision process and provide a better foundation for discussing the origin of observed effects. It is not surprising that comprehension affected task completion in selective but varied ways. However, the finding that different levels are affected in different ways raises two important questions. First, it makes it harder to relate the findings here to previous studies that show effects in reaction times. Second, there are two possible interpretations. The different levels could be affected because language has a 'global' influence on processing that is manifest in different effects because each level of processing operates with different mechanisms. Under this description, the results are

not necessarily due to the direct engagement of sensory areas by semantic access, and the effects on d' could be due to top-down effects that ultimately filter down to the lowest levels of processing. Alternatively, language could have separable effects at different levels, with each effect somewhat independent of the others. In this case, semantic access does engage directly with low level processes (d'), decision processes (c) and response preparation/execution (reaction times).

The blocked presentation (even when passively comprehended) does not rule out strategic processes. Participants may well have been aware that each block contained 'up' or 'down' words. Therefore, having demonstrated effects with blocked single word presentation it is crucial to show similar effects following randomised trial-by-trial effects where the linguistic manipulations are hidden from the participant. The Criterion shift indicates that the comprehension of these words affected processing at a higher level than sensitivity. Therefore the first interpretation, a 'global' effect of comprehension, is not ruled out by this first experiment. Although the effects at each level were in different directions, the observed effects could be entirely contextual; the product of semantic representations modulating specific top-down processes (such as attention) which in turn affect early perceptual processes (e.g. Yantis, 2005). The changes in sensitivity could come via higher-order decision processes, rather than being an elementary consequence of semantic activation (i.e. direct engagement). It is important to note that this account does not make any claims about the semantic system; it simply offers a mechanism for the observed interaction, via task relevant expectations and draws on findings from the visual perception literature in which top-down feedback is able to alter early visual processing (e.g. Yantis, 2005). A central reason for using RDK stimuli and SDT analysis was to base the results on a sound literature in visual perception, therefore interpretation and explanation must consider what is known about the perceptual system. The other interpretation, that different levels are affected independently, is an intriguing possibility since it speaks of complex integration processes between comprehension and different levels in visual perception. However, it would require effects on perceptual sensitivity (d') without effects on decision thresholds (c and β) or similar effects on reaction times.

In the next experiments items were presented in a randomized order rather than blocked. This is essential in order to rule out consciously mediated strategic effects. When items are randomized participants should no longer be aware of the directional content of the

words. In addition, single words and sentences are used to see if a fuller event (i.e. a sentence) produces different results.

6.3. *Single Word Comprehension & Motion Detection*

Participants performed a coherent motion detection task with stimuli at their threshold coherence level. RDKs displayed coherent motion upwards or downwards or incoherent (random) motion. Each RDK was preceded by a single word presented over headphones. On filler trials, participants answered comprehension questions about the word they had just heard, making comprehension active but without the salient directional context that may have driven the effects in 6.2. SDT separated the effects of comprehension into (i) those at the perceptual level (Sensitivity); (ii) those at the decision level (Criterion); and (iii) those at the overall response level (Reaction Time). We found that single words referring to motion were not able to influence visuo-spatial attention, however, they may be sufficient to influence motion processing. The hypothesis is that if single words tap the same mechanisms as blocked presentation, we should see the same pattern of results as that found for Experiment 6.2: (i) d' will be lower when the words and the motion are incongruent (ii) Criterion (and Beta) will be reduced when the words and the motion are congruent and (iii) Reaction Times will not show any significant effects of congruency.

6.3.1. **Participants**

23 Native English speakers (6 male) with a mean age of 23.52yrs (SD = 9.9yrs) were recruited from the UCL Psychology Department Subject Pool in return for monetary compensation of £6 per hour. All had normal or corrected to normal vision.

6.3.2. **Method**

6.3.2.1. *Design*

The Design was the same as Experiment 6.2 except that after Threshold Estimation participants completed 30 practice trials of the motion detection task without hearing any words and Motion Detection was split into 3 runs of 16 blocks of 17 motion

detection trials. At the beginning of each run, participants completed 20 practise trials with words presented auditorally and 50% of these practice trials were followed by a comprehension question. Each run was split into two halves, with each item presented once in each half (8 blocks, 136 trials). Therefore, across the experiment there were a total of 816 trials, and the participant heard each word 6 times. Across each run there were an equal number of coherent and random motion trials.

6.3.2.2. Materials

Item Set Two Single Words was used and individual WAV files for each word were presented.

29 filler words were selected from the set created for that purpose, all could be followed by a comprehension question of the form “Was the word related to X?”. An additional 17 filler items were selected that were not followed by comprehension questions.

Two lists were created, one for each half of an experimental run. The lists were identical except that in List 1a each word was presented with one type of visual motion (coherent or random) and in List 1b each item was presented with the other type. The program for the experiment was designed so that it randomly selected items for presentation in each block, but presented one list in the first half-run of the experiment and the other list in the second half-run. In this way the repetition of items was kept as far apart as possible whilst maintaining randomization. For filler items, Lists 1a and 1b were duplicated (Lists 2a/2b). Items were duplicated in each list, but in List 1 half the items were followed by a comprehension question and half the items were not. In List 2, the items followed by a comprehension question were reversed. Therefore there were two possible lists for an experimental run, with different comprehension questions. The participants completed three runs of the experiment. For the first run, List 1 or 2 was presented. In the second run, the other list was presented. In the third run, the list presented on the first run was presented again. There was a minimum of 272 trials between participants receiving the same comprehension question, which was considered sufficient to maintain concentration and not produce automated responses. The presentation order of lists was balanced across participants.

6.3.2.3. Graphic Display & Apparatus

The Graphic Display and Apparatus were identical to Experiment 6.2 except that the monitor was a CRT display set to 75Hz (resolution = 800×600, 17' DELL P793). A side-light facing the wall behind the participant was used to provide lighting, this ensured that the motion stimuli were easy to discriminate but participants' eyes did not become fatigued.

6.3.2.4. Procedure

Threshold Estimation

This was identical to Experiment 6.2 except that there were 4 practice blocks and it was completed only once before the motion detection phase.

Motion Detection

This was the same as Experiment 6.2 except that motion detection trials were synchronised to single word presentation. In each trial, participants heard a word which was followed by an RDK. During experimental trials, RDKs were presented after an ISI between 27- 93ms, in 13ms steps. The delay was kept constant for each item. On filler trials, RDKs appeared at different offset delays each time the filler word was presented. Participants were instructed to listen to the words and also detect the motion in the RDK patterns. They were told that on some trials, following the RDK, they would be asked a question about the word they just heard. In each run 11% of the filler trials were followed by a comprehension question. Comprehension questions always appeared with the response options underneath, "yes" always appearing on the left and "no" on the right. Participants responded to the questions by pressing the left arrow key for 'yes' and the right arrow key for 'no'. There was a 1500ms timeout for responding to the RDKs and a 4000ms timeout for responding to the comprehension questions.

6.3.2.5. Analysis

2 subjects were removed as they did not complete the Threshold Estimation task properly, so reliable values for their threshold could not be established. 4 further

subjects were removed as they were not true Native English speakers (being trilingual and from countries where English is not the national language); this was established by more detailed questioning after the experiment. The analysis was similar to Experiment 6.2 with Word Category analysed as Control, Match and Mismatch across Motion Directions (upwards and downwards). All trials with a valid response were analysed and entered into by Subjects/by Items analyses for SDT measures. Reaction time data for Hits only was extracted and times above 2 standard deviations for each subject were excluded, this removed 5.34% of the data.

6.3.3. Results

6.3.3.1. *d'*

There was a main effect of Motion Direction, with d' for downwards motion 0.207 units greater than upwards motion ($F(1,16) = 6.828$, $p-\eta^2 = 0.299$; $F(1,87) = 23.853$, $p-\eta^2 = 0.215$). The main effect of Word Category approached significance by Subjects ($F(2,32) = 3.160$, $p = 0.056$, $\eta^2 = 0.165$; $F(2,87) = 1.459$, $p > 0.2$, $\eta^2 = 0.032$). Planned comparisons showed that the Control condition had d' values significantly higher than the Match condition ($F(1,26) = 5.449$, $\eta^2 = 0.254$; $F(1,58) = 2.6$, $p > 0.1$, $\eta^2 = 0.043$) and marginally higher than the Mismatch condition ($F(1,16) = 3.635$, $p = 0.075$, $\eta^2 = 0.185$; $F(2 < 1)$). See Table 6-2.

6.3.3.2. *Criterion*

There were no significant differences (See Table 6-3 for Word Category by Motion Direction mean values).

6.3.3.3. *Beta*

There were no significant differences (See Table 6-4 for Word Category by Motion Direction mean values).

Table 6-3: Criterion (c) data for experiments 6.2- 6.5 by Condition (Control, Match, Mismatch) and Motion Direction (Down, Up)
(Standard deviation in brackets)

Experiment	Control		Condition Match		Mismatch	
	Down	Up	Down	Up	Down	Up
6.2 Blocked Word	0.1 (0.35)	0.13 (0.097)	0.02 (0.27)	0.03 (0.28)	0.11 (0.29)	0.14 (0.34)
<i>Category Mean</i>	<i>0.11 (0.35)</i>		<i>0.027 (0.27)*</i>		<i>0.13 (0.31)</i>	
6.3 Single Word	-0.005 (0.38)	0.027 (0.38)	0.002 (0.37)	-0.04 (0.41)	0.045 (0.41)	0.022 (0.43)
<i>Category Mean</i>	<i>0.11 (0.35)</i>		<i>-0.019 (0.36)</i>		<i>0.33 (0.38)</i>	
6.4 Sentences	-0.15 (0.32)	-0.04 (0.33)	-0.14 (0.33)	-0.14 (0.28)	-0.10 (0.37)	-0.08 (0.34)
<i>Category Mean</i>	<i>-0.10 (0.27)</i>		<i>-0.14 (0.26)</i>		<i>-0.09 (0.29)</i>	
6.5 Blocked Single Word	0.02 (0.31)	-0.07 (0.31)	0.002 (0.26)	-0.013 (0.35)	0.121 (0.32)	-0.05 (0.22)
<i>Category Mean</i>	<i>-0.024 (0.28)</i>		<i>-0.006 (0.26)</i>		<i>0.036 (0.23)</i>	

*main effect driven by this condition being significantly different to both other conditions

† significant difference between these two conditions in planned comparisons

6.3.3.4. Reaction Times

There was a significant main effect of Motion Direction by Items ($F_2 < 1.3$; $F_2(1,87) = 43.857$, $p_{eta} = 0.335$) and a significant interaction between Word Category and Motion Direction by Subjects ($F_1(2,32) = 5.248$, $p_{eta} = 0.247$; $F_2 < 1.4$). Reaction times for Upwards motion were on average 23ms faster than reaction times to Downwards motion. The interaction was driven by this difference being smaller in the Mismatch condition (9ms) as compared to the Match (24ms) and Control (34ms) conditions.

6.3.4. Discussion

The results showed that perceptual sensitivity (d') was higher for downwards motion as compared to upwards motion. Reaction times were longer for downwards motion as compared to upwards motion. Therefore it looks as though upwards motion suffered from a speed-accuracy trade-off, with participants responding more quickly and making more errors as a result. Aside from the effect of direction, the main effect of Word Category was significant, with d' lower in Match and Mismatch conditions as compared to the Control. Therefore directional words of all kinds impaired perceptual sensitivity. In this experiment there was no consistent directional context and single words preceded RDK onset (although a variable ISI meant RDK onset was not entirely predictable). The immediate semantic processing of each word was more than likely over before the presentation of the motion pattern. However, comprehension was active as participants had to answer questions about the semantic content of the words in filler trials. It is possible that because the participants had to maintain the semantic content of the word during RDK presentation (in order to answer the occasional filler question), when the content was directional this interfered with motion detection. When the semantic content was non-directional, it did not affect motion detection. This suggests that there was a general demand on motion processing systems during the maintenance of semantic motion information, but the demand was not specific enough to produce congruency effects (congruent conditions produced slightly lower d' values, but this difference was not reliable). However, under this explanation we might expect the maintenance of directional semantic content to affect decision thresholds or reaction times, but it did not. These two dependent measures, which correspond to later/higher stages in decision and response processes, showed no effects of directional words in general or in terms of congruence. This suggests that the effect of directional words was via a close connection between semantic and perceptual systems, rather than a more global effect of maintenance on all stages of task execution. This demonstrates that effects on d' are independent of those at higher levels, however, given the temporal relationship between the word and the RDK, it is not safe to infer that the effects are a result of primary semantic access (i.e. the immediate semantic activation whilst the word was comprehended). These effects could be due to processes that maintain a word's meaning in memory, producing a general tax on resources in the motion processing system and reducing sensitivity. Therefore it is not clear how truthfully they reflect 'normal' semantic activity; in order to make safer interpretations about primary

semantic activity and visual processing, closer temporal contiguity between the linguistic and visual stimuli is needed.

The results from this experiment suggest that the consistent context created through blocked presentation was a major factor in producing reliable congruency effects as seen in Experiment 6.2. Perhaps this is not surprising, given that the semantic content of isolated single words may be too short-lived and the activation too weak to show similar behavioural effects – especially when the presentation of the words and the RDK were separate temporal events. In the next experiment, we looked at the effects when the same single words were presented as verbs within a sentence context. Sentences refer to complete events, therefore we might expect their more substantial semantic representations to influence motion detection. However, sentences constrain the interpretation of the verb, which may or may not influence the impact of semantic motion on perception, particularly if it shifts the focus away from the motion event (e.g. onto the object location). For example, each verb can now be indexed to an object and the representation of the motion event should be more concrete. In Chapter 5, visual stimuli were presented a sentence offset and we found effects on visuo-spatial attention. As discussed above, in order to make inferences about primary semantic access and motion perception the presentation of motion stimuli has to be contiguous with the comprehended items. Therefore, RDKs will be presented after the verb, but during the remainder of the sentence. Although the critical lexical item and motion pattern are separate temporal events, sentence comprehension – and by extension the activation of semantic representations - is still in progress.

6.4. *Sentence Comprehension & Motion Detection*

Participants performed a coherent motion detection task with stimuli at their threshold coherence level. RDKs displayed coherent motion upwards or downwards or incoherent (random) motion. Sentences were presented over headphones and in experimental trials RDKs were presented following verb offset. On filler trials, participants answered comprehension questions about the sentence they had just heard, making comprehension active. SDT separated the effects of comprehension into (i) those at the perceptual level (Sensitivity); (ii) those at the decision level (Criterion); and (iii) those at the overall response level (Reaction Time). We found that single words referring to motion did not show strong effects on motion processing; sentences may therefore provide a more substantial representation. If verbs within a sentence context tap the same mechanisms as blocked presentation, we should see the same pattern of results as that found for Experiment 6.2: (i) d' will be lower when the words and the motion are incongruent (ii) Criterion (and Beta) will be reduced when the words and the motion are congruent and (iii) Reaction Times will not show any significant effects of congruency.

6.4.1. **Participants**

22 Native English speakers (11 female) with a mean age of 25.77yrs (SD = 12.0yrs) were recruited from the UCL Psychology Department Subject Pool in return for monetary compensation of £6 per hour. All had normal or corrected to normal vision.

6.4.2. **Method**

6.4.2.1. *Design*

The Design was identical to Experiment 6.3 except that the Motion Detection was split into 2 runs of 20 blocks of 13 motion detection trials. At the beginning of each run participants completed 10 practise trials with sentences, 50% of which were followed by comprehension questions. Each run was split into two halves, with each item presented once in each half (130 trials, 40 of which were filler trials). Across the experiment there were a total of 520 trials, and the participant heard each sentence 4 times.

6.4.2.2. Materials

Item Set Two Sentences was used. 40 comprehension questions were selected from the set created for this purpose (followed by questions of the form “Did X verb Y?”, requiring a yes/no answer). Item presentation was randomised according to the list structure used in Experiment 5.4. Offset times for Experimental and Filler sentences were established for Experiment 5.4 and were re-used here.

6.4.2.3. Graphic Display & Apparatus

There were identical to Experiment 6.3.

6.4.2.4. Procedure

Threshold Estimation

This was identical to Experiment 6.3

Motion Detection

This was the same as Experiment 6.3 except that motion detection trials were synchronised to verb offset in experimental trials, and varied sentence positions in filler trials. In each trial, participants heard a sentence and judged an RDK. Participants were instructed to listen to the sentences and also detect the motion in the RDK patterns. They were told that on some trials, following the sentence and RDK, they would be asked a question about the sentence they just heard. In each run 15% of the filler trials were followed by a comprehension question. Question presentation, response keys and time-outs were the same as for Experiment 6.3

6.4.2.5. Analysis

3 participants were removed as they did not complete the Threshold Estimation correctly. Analysis was identical to Experiment 6.3. Reaction Times above 3 standard deviations for each participant were removed; this affected 1.34% of the data.

6.4.3. Results

6.4.3.1. *d'*

There was a significant main effect of Motion Direction by Items ($F_1 < 1.9$; $F_2(1,84) = 16.226$, $p-\eta^2 = 0.162$), with Upwards motion 0.173 units lower than Downwards motion. Planned comparisons showed that the Match condition had higher d' values compared to Control by Subjects ($F_1(1,18) = 6.107$, $p-\eta^2 = 0.253$; $F_2(1,56) = 2.396$, $p = 0.127$, $p-\eta^2 = 0.041$). There was also a significant interaction between Motion Direction and Word Category when the Match and Mismatch condition were compared ($F_1(1,18) = 6.261$, $p-\eta^2 = 0.258$; $F_2(1,56) = 4.424$, $p-\eta^2 = 0.073$): there was a smaller difference in d' between Upwards and Downwards motion in the Mismatch condition as compared to the Match condition (0.082 vs. 0.342) (See Table 6-2).

6.4.3.2. *Criterion*

There was a significant main effect of Word Category by Subjects that approached significance by Items ($F_1(2,36) = 3.866$, $p-\eta^2 = 0.177$; $F_2(2,84) = 2.518$, $p = 0.087$, $p-\eta^2 = 0.057$). Planned comparisons showed that the Match condition had lower Criterion values than the Control condition ($F_1(1,18) = 7.173$, $p-\eta^2 = 0.285$; $F_2(1,56) = 3.399$, $p = 0.071$, $p-\eta^2 = 0.057$) and the Mismatch condition ($F_1(1,18) = 7.842$, $p-\eta^2 = 0.303$; $F_2(1,56) = 5.042$, $p-\eta^2 = 0.083$). There was also a significant interaction between Motion Direction and Word Category when the Match condition was compared to the Control condition ($F_1(1,18) = 5.162$, $p-\eta^2 = 0.223$; $F_2(1,56) = 4.005$, $p-\eta^2 = 0.067$); this was driven by a difference in mean Criterion values for Upwards and Downwards motion in the Control condition that was not present in the Match condition (0.107 vs 0) (See Table 6-3).

Table 6-4: Beta data for experiment 6.2- 6.5 by Condition (Control, Match, Mismatch) and Motion Direction (Down, Up)

Experiment	Control		Condition Match		Mismatch	
	Down	Up	Down	Up	Down	Up
6.2 Blocked Word	1.31 (1.15)	1.25 (0.59)	1.02 (0.32)	1.16 (0.80)	1.18 (0.38)	1.27 (0.74)
<i>Category Mean</i>	<i>1.28 (0.87)</i>		<i>1.09 (0.56)</i>		<i>1.23 (0.56)</i>	
6.3 Single Word	1.07 (0.35)	1.06 (0.24)	1.06 (0.27)	1.04 (0.25)	1.06 (0.32)	1.1 (0.29)
<i>Category Mean</i>	<i>1.06 (0.28)</i>		<i>1.05 (0.24)</i>		<i>1.08 (0.21)</i>	
6.4 Sentences	0.86 (0.27)	1.11 (0.68)	0.87 (0.36)	0.98 (0.42)	1.02 (0.37)	1.00 (0.41)
<i>Category Mean</i>	<i>0.99 (0.41)</i>		<i>0.93 (0.33) †</i>		<i>1.01 (0.33)</i>	
6.5 Blocked Single Word	1.08 (0.30)	1.05 (0.26)	1.21 (0.38)	1.15 (0.59)	1.35 (0.77)	0.99 (0.18)
<i>Category Mean</i>	<i>1.06 (0.24)</i>		<i>1.14 (0.40)</i>		<i>1.17 (0.42)</i>	

† by Items difference between this and the other two conditions in planned comparisons

6.4.3.3. *Beta*

There was a significant interaction between Motion Direction and Word Category by Subjects ($F(1,2,36) = 3.656$, $p-\eta^2 = 0.169$; $F_2 < 1.6$) and a marginal main effect of Word Category by Items ($F_1 < 1.9$; $F_2(2,84) = 3.015$, $p = 0.054$, $p-\eta^2 = 0.067$). Planned comparisons showed a trend for lower Beta values in the Match condition as compared to the Control ($F_1(1,18) = 2.288$, $p = 0.148$, $p-\eta^2 = 0.113$; $F_2(1,56) = 5.444$, $p-\eta^2 = 0.089$) and Mismatch conditions ($F_1(1,18) = 6.937$, $p-\eta^2 = 0.278$; $F_2(1,56) = 2.793$, $p = 0.1$, $p-\eta^2 = 0.048$). When the Control and Mismatch conditions were compared, there was a significant interaction between Motion Direction and Word Category by Subjects ($F_1(1,18) = 4.971$, $p-\eta^2 = 0.216$; $F_2 < 1.6$); this was driven by a difference in mean Beta values for Upwards and Downwards motion in the Control condition that was not present in the Mismatch condition (0.2481 vs 0.0187). There was also a main effect of Motion Direction with higher Beta values for Upwards motion ($F_1 < 1.15$; $F_2(1,56) = 4.936$, $p-\eta^2 = 0.081$) (See Table 6-4).

6.4.3.4. Reaction Times

There was a significant main effect of Motion Direction by Items ($F_1 < 1$; $F_2(1,84) = 5.560$, $p-\eta^2 = 0.062$), with reaction times for Upwards motion 21.32ms longer than for Downwards motion. Planned comparisons showed that reaction times in the Match condition were approximately 10ms faster than the Control, this difference approached significance by Subjects ($F_1(1,18) = 4.135$, $p = 0.057$, $p-\eta^2 = 0.187$; $F_2 < 1$). See Table 6-5, which also illustrates that reaction times in this experiment were longer than those in other experiments by approximately 100ms.

Table 6-5: Reaction Time data for experiment 6.2- 6.5 by Condition (Control, Match, Mismatch) and Motion Direction (Down, Up)
Standard Deviation in Brackets

Experiment	Control		Condition Match		Mismatch	
	Down	Up	Down	Up	Down	Up
6.2 Blocked Word	670.56 (98.41)	667.74 (100.03)	669.01 (85.13)	661.74 (90.42)	682.62 (100.69)	648.61 (92.61)
<i>Category Mean</i>	<i>669.16 (94.56)</i>		<i>665.38 (87.78)</i>		<i>665.62 (96.65)</i>	
6.3 Single Word	645.50 (125.01)	622.25 (79.60)	636.50 (117.60)	612.50 (93.40)	633.63 (109.40)	624.63 (84.35)
<i>Category Mean</i>	<i>628.38 (93.70)</i>		<i>624.49 (99.01)</i>		<i>629.13 (87.99)</i>	
6.4 Sentences	744.68 (145.76)	753.08 (131.10)	734.19 (145.27)	743.78 (121.30)	744.06 (148.50)	747.80 (135.00)
<i>Category Mean</i>	<i>748.88 (130.77) *</i>		<i>738.97 (123.85) *</i>		<i>745.93 (131.37)</i>	
6.5 Blocked Single Word	579.83 (158.51)	609.57 (110.34)	637.45 (118.14)	615.31 (92.92)	639.91 (116.17)	631.40 (103.39)
<i>Category Mean</i>	<i>594.70 (107.93)</i>		<i>626.38 (95.38)</i>		<i>635.66 (98.75)</i>	

*marginal difference between these two conditions in planned comparisons.

6.4.4. **Discussion**

The d' results showed higher values for Downwards motion as compared to Upwards motion. Reaction time data also showed a difference for motion direction, with slower times for Upwards motion as compared to Downwards motion. Upwards motion was responded to more slowly and less accurately than Downwards motion. In addition, Beta showed higher values (i.e. more conservative responding) for Upwards as

compared to Downwards motion. Overall this pattern of results suggests that Upwards motion is harder to detect than downwards motion. Participants were slower, less accurate and they raised decision thresholds to avoid false alarms, indicating that upwards motion was less salient and therefore less easy to detect.

There was a significant interaction between Sentence Category and Motion Direction, when d' was compared between Match and Mismatch conditions. The Mismatch condition had a smaller difference between Downwards and Upwards motion than the Match condition. In a similar fashion, differences between criterion values for different motion directions is greater in the Control condition (when no directional words are present) than in directional conditions (Match or Mismatch). Therefore, measures of perceptual sensitivity show that incongruent words reduce the criterion disparity between different motion directions, possibly because they create more equal conditions between motion directions. Decision thresholds showed differences between motion directions in the control condition that were reduced in both directional conditions (Match and Mismatch). Therefore at the decision level directional words in general created a more equal response to different motion directions. It is not easy to interpret these changes, and it is not clear why directional words should 'equalise' response to different motion directions. However, it does provide some limited evidence for effective manipulation of congruent/directional semantics.

Most importantly, there was a trend for sensitivity to be higher in Match as compared to Control conditions, this was not due to a speed accuracy trade-off since reaction times also tended to be faster for Match as compared to Control conditions. This is in line with previous experiments which show speeded responses to visual stimuli following comprehension of congruent sentences (see Section 2.1.2.3). Therefore there was a weak but beneficial effect of congruence on perceptual sensitivity and response time. Turning to measures of decision processes, both c and β showed lower values (i.e. more liberal decision thresholds) for congruent conditions as compared to both incongruent and control conditions. This is in line with Experiment 6.2 and supports 'decision priming' under congruent conditions. Therefore, congruence improved perceptual sensitivity, lowered decision thresholds and facilitated response times.

Directional sentences created a directional context that was sufficient to affect all levels of motion detection. The RDK was perceived whilst the post-verbal parts of the sentence were being comprehended, recall that this experiment also used active

comprehension. Therefore, in Match or Mismatch conditions motion detection occurred during the construction of a semantic representation of a directional motion event. A congruent context was beneficial to all levels of task processing, supporting the sentence as a more substantial linguistic manipulation than single words. It is interesting that all effects were found in congruent conditions, this suggests that the meaning of the sentence was used to the benefit of task performance whenever possible. Sentence onset cued the appearance of the RDK at some point during that sentence, so it is possible that congruent semantic content primed motion detection – producing faster response times, lower decision thresholds and greater sensitivity. Rather than supporting a localised connection between semantic and perceptual content (as for Experiment 6.3), these results support integration of semantic and perceptual information in order to benefit task performance at every level. At this point we have three experiments, each of which show a different direction of effects on perceptual sensitivity. It is important that across all three we have seen effects on d' , since this is a measure of low level visual processes. Perceptual sensitivity can be routinely modulated by semantic content, but that modulation appears to be complex and task dependent.

So far, we have tested a strong single word context unsynchronised but overlapping with RDK presentation, a weak single word context synchronised but temporally separate to RDK presentation and a sentence context in which RDK presentation was temporally separate to the critical lexical item but during the remainder of the sentence. The results for Criterion are consistent in two experiments (6.2 and 6.4), suggesting that something like decision priming (less conservative decision thresholds) is present in congruent conditions when there is a strong enough linguistic context. This supports the criterion effects as based in strategic processing, integrating congruent information to try and benefit perception. It is not straightforward to explain the mechanisms behind criterion effects (Wickens, 2002), so we will return to these in the final chapter for further discussion.

The most interesting effects are those found for sensitivity, since it is here that strong predictions about embodiment are tested. According to the embodied direct engagement hypothesis, the semantic representation of motion words is within motion processing systems, therefore effects on sensitivity (which are thought to reflect these low level visual systems) are expected. As we have seen, blocked single words produce

reduce d' under incongruent conditions, whereas randomised single words reduce d' whenever directional words are present. It is therefore of interest whether the more selective results for experiment 6.2 were entirely due to the blocked presentation and some top-down modulation of motion processing that increased over the course of the block. This would be more in line with weak embodiment which predicts task dependent changes in the connection between semantic and perceptual information. On the other hand, the results could have been because semantic activation occurred during RDK perception; words were unsynchronised, but overlapping, with RDK presentation. This would result in a more immediate, and therefore stronger, influence of semantic access on motion processing and congruency effects have been shown in the two experiments where linguistic items did overlap with visual stimuli. This is in line with strong embodiment where interactions are always present but dependent on the temporal overlap between semantic and sensory-motor activity (e.g. Borreggine & Kaschak, 2006). To explore whether temporal contiguity or blocked context is the critical factor in producing effects on d' from single words, the final experiment presents RDKs in synchrony and contiguity with blocked single words.

6.5. *Blocked Single Word Comprehension Synchronised to Motion Detection*

Participants performed a coherent motion detection task with stimuli at their threshold coherence level. RDKs displayed coherent motion upwards or downwards or incoherent (random) motion. Single words were presented over headphones and RDKs were presented during the presentation of each word, timed to end with word offset. Words were presented in blocks (Up, Down, Control) therefore producing a salient directional context. Comprehension was passive. SDT separates the effects of comprehension into (i) those at the perceptual level (Sensitivity); (ii) those at the decision level (Criterion); and (iii) those at the overall response level (Reaction Time). Statistical analyses will be completed by dividing each block into thirds (1st, 2nd, 3rd) to look at the changes in dependent measures over the course of the block. This is to explore whether blocked presentation enhances effects that are already present for individual trials, or whether blocked presentation is the only condition under which these effects are seen. Descriptive analyses will look at the change in dependent measures over each trial in the block. The hypothesis is that if the semantic content of single words is sufficient to

influence detection, effects of congruency should be present across all trials and enhanced by blocked presentation (i.e. greater at the end of the block than at the beginning). In contrast, if blocked presentation is important in producing consistent effects there should be no stable effects of congruency but emerging differences over the course of the block. The same pattern as seen in 6.2 should be observed: (i) d' will be lower when the words and the motion are incongruent (ii) Criterion (and Beta) will be reduced when the words and the motion are congruent and (iii) Reaction Times will not show any significant effects of congruency.

6.5.1. **Participants**

26 Native English speakers (16 female) with a mean age of 25.38yrs (SD = 6.6yrs) were recruited from the UCL Psychology Department Subject Pool in return for monetary compensation of £6 per hour. All had normal or corrected to normal vision.

6.5.2. **Method**

6.5.2.1. *Design*

The Design was the same as Experiment 6.2 except that after Threshold Estimation participants completed 8 practice trials of the motion detection task at an easy level (50% coherence) and then as much practice as they wanted (in blocks of 10 trials) at their threshold level. Motion Detection was split into 3 runs of 6 blocks of 30 motion detection trials. At the beginning of each run, participants completed 16 practise trials with filler words presented binaurally. Each run was split into two halves, with each item presented once in each half (3 blocks, 90 trials). Therefore, across the experiment there were a total of 540 trials, and the participant heard each word 6 times. Across each block there were an equal number of coherent and random motion trials.

6.5.2.2. Materials

Item Set Two Single Words was used. Each half-run had 3 blocks that presented each Word Category (Up, Down and Control) in a pseudo-randomised order so that across the run the same Word Category was not presented consecutively. Two lists were created, one for each half of a run; in each list each word appeared once with coherent motion and once with random motion. The order of presentation was randomised within each Word Category.

6.5.2.3. Graphic Display & Apparatus

These were identical to Experiment 6.3.

6.5.2.4. Procedure

Threshold Estimation

This was identical to Experiment 6.3

Motion Detection

This was the similar to Experiment 6.3 except that there were no comprehension questions.

On each trial, stimuli were presented so that offset of the RDK and the WAV file coincided. By using this method, there was a different onset of the RDK from trial to trial whilst ensuring that it was never displayed outside the duration of the word. The RDK onset time was established by subtracting 150ms (the duration of the RDK) from the word's duration. This had been previously established using a custom made programme that recorded the onset and offset time for each item's WAV file; duration was then established by subtracting the onset from the offset time.

6.5.2.5. Analysis

3 subjects were removed as they did not complete the Threshold Estimation task properly, so reliable values for their threshold could not be established. One subject was removed as they did not complete the task properly (with over 70% of responses

being errors). Two subjects were removed as they were too fatigued to complete the task properly. Reaction time data above 3 standard deviations for each subject were excluded, this removed 4.77% of the data. Within each Word Category, by Subjects, data was collapsed across runs for each trial (1 to 30). This was statistically analysed by further collapsing trials into Block Sections, made up of epochs of 10 trials. Trials 1-10 made up the first Block Section, trials 11-20 made up the second Block Section and trials 21-30 made up the third Block Section. This was entered into a 3x3x2 repeated measures ANOVA comparing Word Category (Control, Match, Mismatch), Block Section (first, second, third) and Motion Direction (Upwards, Downwards). However, collapsing into Block Sections may not present a true picture of changes across individual trials (being an average), therefore to illustrate the change in dependent measures in more detail, graphs are presented which show data on a trial by trial basis for each Word Category by Subjects. A by Items analysis was not possible as each word had not appeared in each possible trial location (1-30) in a balanced fashion, being randomly determined in each block.

6.5.3. Results

6.5.3.1. d'

There was a small but significant interaction between Word Category and Block Section ($F(4,76) = 2.7$, $p-\eta^2 = 0.124$). The Control condition had a mean value of 0.850 in the first Section, 0.852 in the second Section and dropped to 0.738 in the third Section. The Mismatch condition had a mean value of 0.881 in the first Section, dropping to 0.762 in the second and then 0.704 in the third Section. The Match condition had a value of 0.777 in the first Section, rising to 0.828 in the second Section and 0.918 in the third Section (See Figure 6-3 and Table 6-2). The interaction was significant when the Match condition was compared to the Control ($F(2,38) = 3.172$, $p = 0.053$, $p-\eta^2 = 0.143$) and Mismatch conditions ($F(2,38) = 5.015$, $p-\eta^2 = 0.209$); however, when Mismatch and Control conditions were compared there was no longer a significant interaction, but a marginal main effect of Block Section ($F(2,38) = 2.991$, $p = 0.062$, $p-\eta^2 = 0.136$). Overall there was also a marginal interaction between Motion Direction and Block Section ($F(2,38) = 3.140$, $p = 0.055$, $p-\eta^2 = 0.142$). Upwards motion had slightly lower d' values than Downwards motion in the first and second Sections (a mean difference of

0.1 and 0.2 respectively), this difference was removed in the third Section (a mean difference of 0.02). (See Table 6-2 and Figure 6-3).

Table 6-6: The mean difference in Criterion and Beta for each Motion Direction by Word Category
Standard deviation in brackets.

Experiment	Condition					
	Control		Match		Mismatch	
	Down	Up	Down	Up	Down	Up
Criterion C	0.02 (0.31)	-0.07 (0.31)	0.002 (0.26)	-0.013 (0.35)	0.121 (0.32)	-0.05 (0.22)
<i>Mean difference</i>	0.09 (0.31)		0.015 (0.31)		0.126 (0.27)	
Beta	1.08 (0.30)	1.05 (0.26)	1.21 (0.38)	1.15 (0.59)	1.35 (0.77)	0.99 (0.18)
<i>Mean difference</i>	0.03 (0.28)		0.06 (0.49)		0.36 (0.48)*	

*The mean difference in the Mismatch condition produced a significant interaction between Word Category and Motion Direction.

6.5.3.2. Criterion

There were no significant differences, but the pattern was similar to that seen for Beta (See Table 6-3 and 6-6).

6.5.3.3. Beta

There was a significant interaction between Word Category and Motion Direction ($F(2,38) = 4.211$, $p = 0.0181$). The values for Upwards and Downwards motion were similar in the Control and Match conditions, but diverged in the Mismatch condition, with Upwards motion having values 0.353 units lower than Downwards motion (See Table 6-4 & 6-6 and Figure 6-4).

Figure 6-3: d prime by Block Section (a) and Trial Number (b)

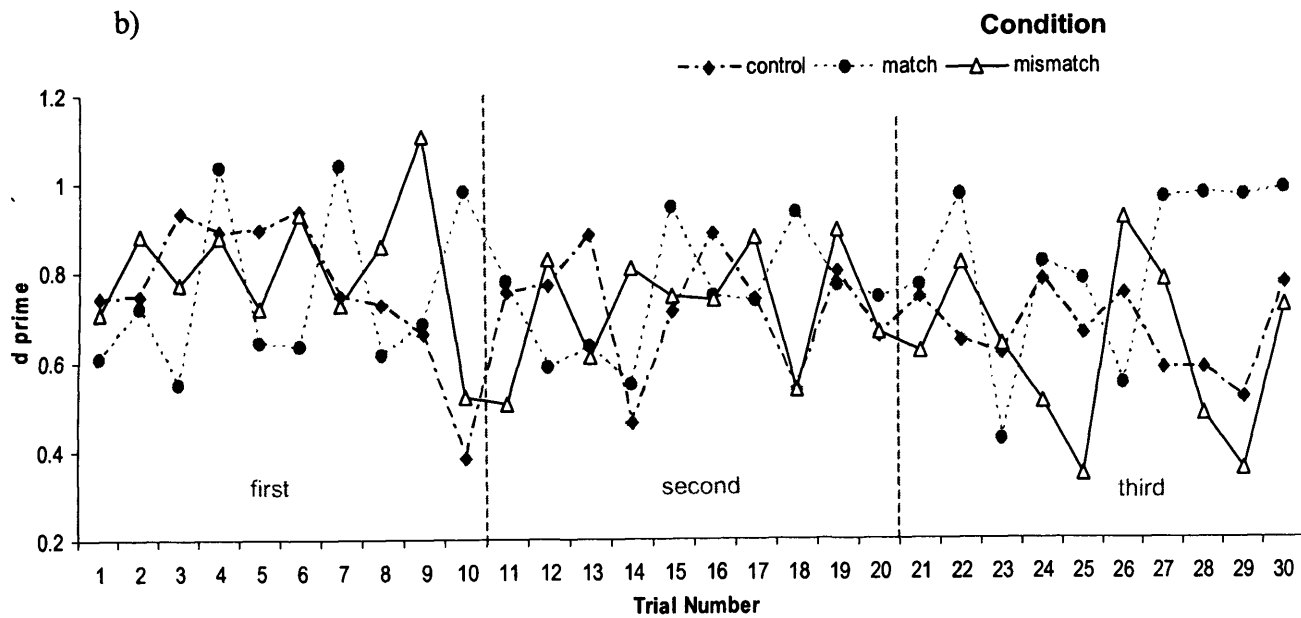
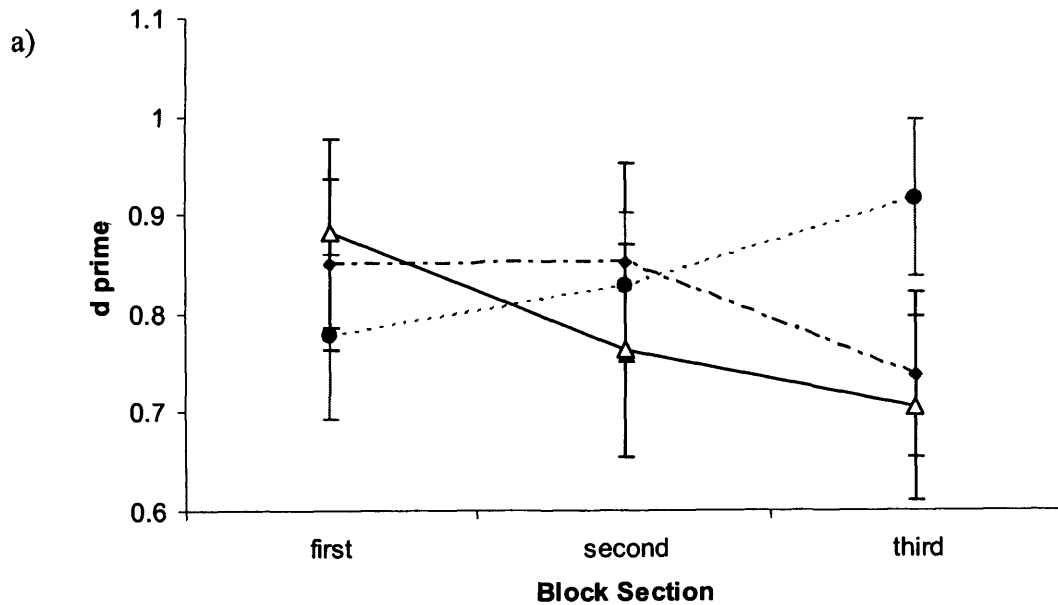


Figure 6-4: Beta by Block Section (a) and Trial Number (b)

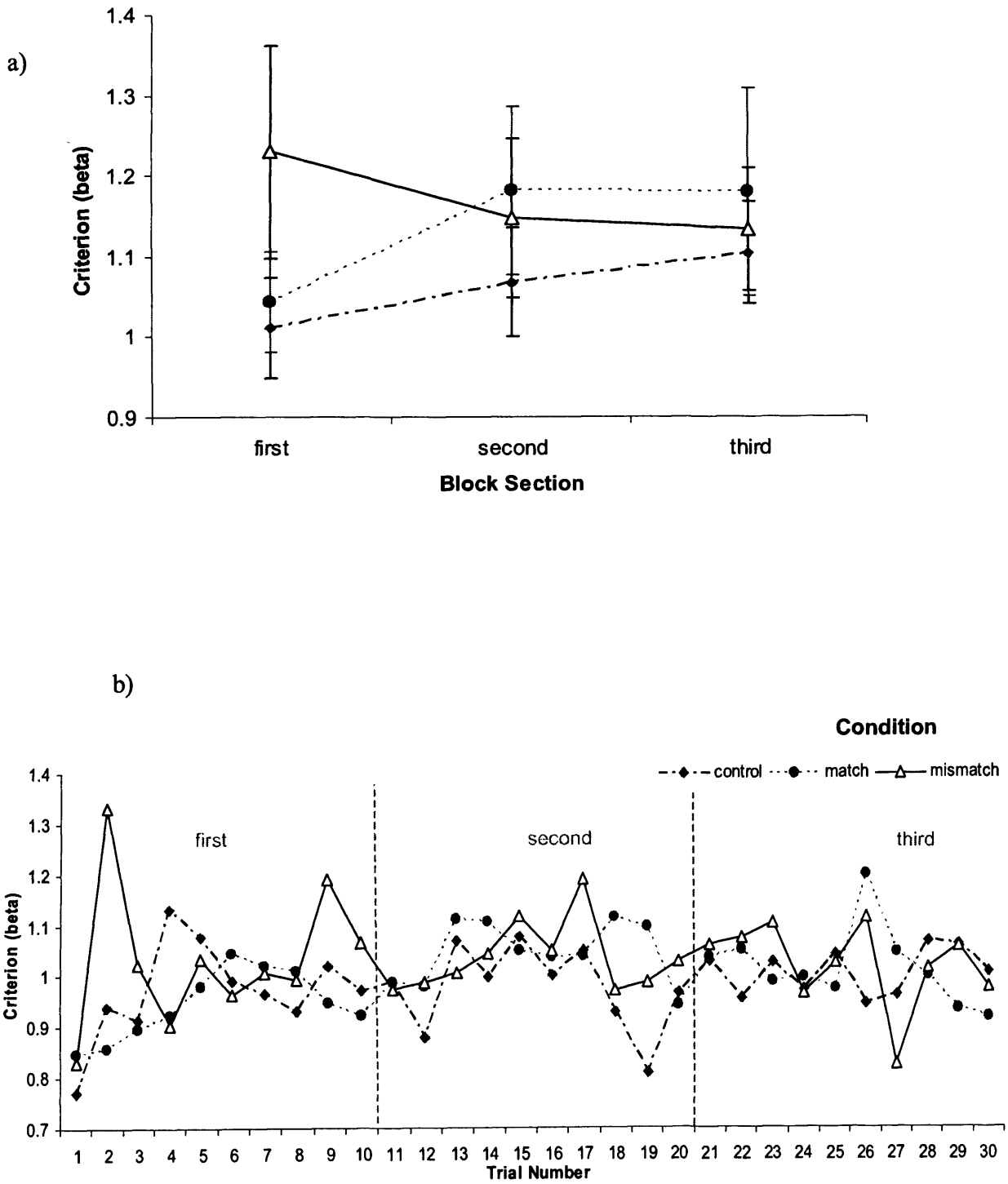
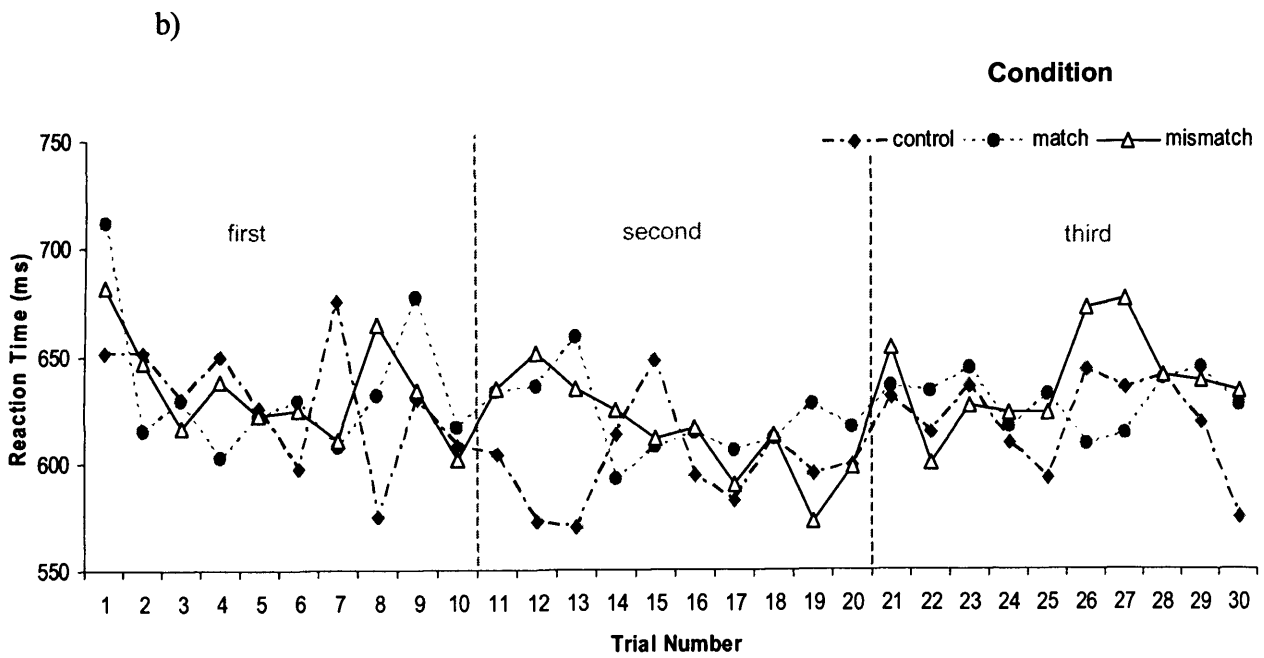
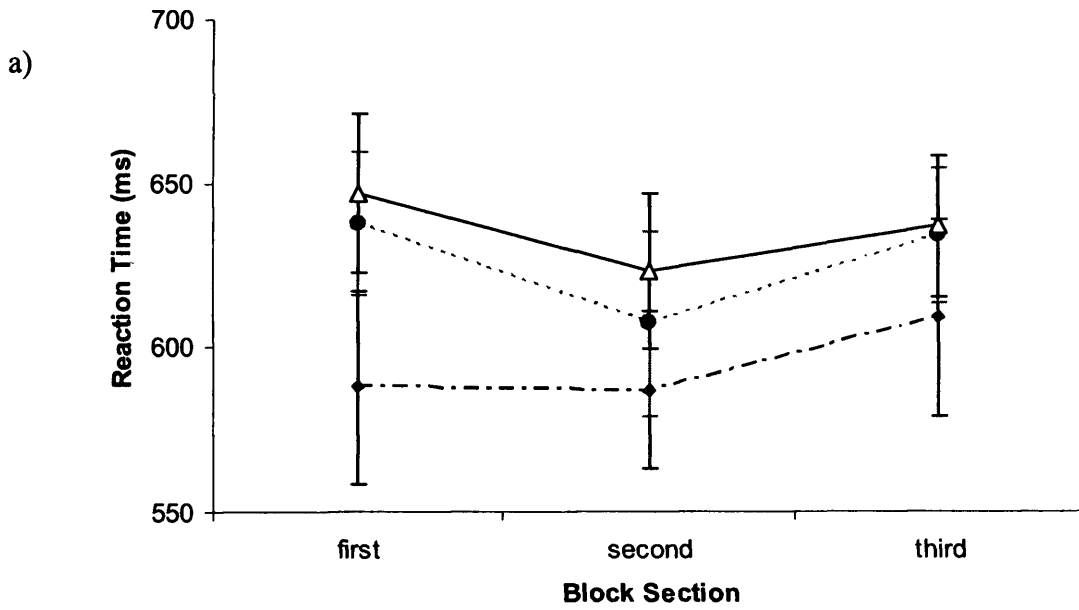


Figure 6-5: Reaction Time (ms) by Block Section (a) and Trial Number (b)



6.5.3.4. Reaction Times

There was a significant main effect of Block Section ($F(2,38) = 5.952$, $p-\eta^2 = 0.239$), with faster reaction times in the second (mean = 605.97) as compared to the first (mean = 624.11, $F(1,19) = 14.119$, $p-\eta^2 = 0.426$) and third (mean = 626.65, $F(1,19) = 8.509$, $p-\eta^2 = 0.309$) Sections. There was also a significant main effect of Word Category ($F(2,38) = 5.952$, $p-\eta^2 = 0.239$); the Control condition had the shortest reaction times (mean = 594.70), followed by times in the Match condition (mean = 626.38) and the longest reaction times were produced in the Mismatch condition (mean = 635.66). The Control condition was marginally faster than the Match condition ($F(1,19) = 3.325$, $p = 0.084$, $p-\eta^2 = 0.149$) and significantly faster than the Mismatch condition ($F(1,19) = 4.905$, $p-\eta^2 = 0.205$) (See Figure 6-5 and Table 6-5).

6.5.3.5. Descriptive Analysis

To explore the detailed time-course of changes in dependent measures across the course of the block, the data for each Trial (1-30) was collapsed across Subjects for each Word Category and illustrated in Figures 6-3b to 6-5b. Dashed vertical lines represent the epochs (first, second and third Block Section) which were used for the statistical analysis. Dependent measures that showed significant differences in the above Block Section comparison were chosen for this descriptive analysis. What is clear across all three graphs is that the mean changes seen in the statistical analysis were not a product of steady differences between Word Categories. Rather, data from trial to trial are highly variable.

Figure 6-3b illustrates the changes in d' , for which the statistical analysis showed a significant interaction between Word Category and Block Section (see Figure 6-3a). In the first Section (Trial Number 1-10), the Match condition was on average slightly lower than the Control and Mismatch conditions. The Match condition had the lowest data points on most trials, but on trials 3,7 and 10 the Match condition had mean values that were higher than both the Control and Mismatch conditions. The Control and Mismatch conditions are broadly similar, although towards the end of the first section they diverge (trials 8-10). The Mismatch condition having one very high data point in trial 9 (above 1.0) followed by a lower data point at trial 10 (below 0.6). The Control condition on the other hand decreases steadily from trials 8-10. In the second Section, all Word Categories were similar although the Mismatch condition was on average

slightly lower than the other two. In partial agreement with this, all Categories are broadly similar across Trials 11-20, although the Mismatch condition does not show consistently lower d' values, with nearly half the data points higher than Control and Match conditions (trials 12, 14, 17 and 19). Finally, in the third Section, the Match condition was higher than the Control and Mismatch condition. This pattern does not emerge until the final 4 trials (27-30), in which the Match condition is consistently higher than the other two. The Control condition shows broadly similar values across trials 21-30, whilst the Mismatch condition again shows greater variance. In the final epoch (the third Block Section) the Mismatch condition shows an almost sinusoidal pattern, with peaks (trials 22 and 26) followed by troughs (trials 25 and 29).

Although Beta did not show significant differences between different Word Categories across Block Sections, the descriptive analysis is still informative and is worth exploring the trial-to-trial changes in decision threshold. Beta has been chosen since it showed significant differences elsewhere so any other trends should also be clearer. Figure 6-4a showed the values of Beta averaged across Block Section, and Figure 6-4b shows values on a trial to trial basis. There is a trend for the Mismatch condition to have higher values than the Control condition (this can also be seen in Table 6-4). The descriptive analysis shows that this is due to several high data points across the block (e.g. at trials 2,9 and 17) for the Mismatch condition which would have raised the average value for each section. The trial-to-trial changes in the Match condition follow the averaged values (Figure 6-4b) quite faithfully; with lower values in the first 10 trials that then remain stable over the remaining 20 trials. However, one peak data point (trial 26) in the final Section pulls the average above that for the Control condition. On the whole, the Control condition had the lowest values (Figure 6-4a). This was due to low data points (trials 1, 12 and 19) that pulled averages down as the majority of trials were comparable to Match and Mismatch conditions. In sum, decision threshold was similar across all Word Categories with the occasional peak/outlier trial pulling averaged values apart.

Figure 6-5b illustrates changes in reaction times on a trial by trial basis. The averaged values (6-5a) show that the Mismatch condition has longer reaction times than the Control condition. The trial-to-trial data shows that each condition is quite variable in the first Block Section. For the Mismatch condition, no trials fall below 600ms and peak values pull the average of the Mismatch condition up (trials 1 and 8). The Match

condition shows a similar pattern with the majority of trials falling between 600-650ms and two peak trials pulling the average up (trials 1 and 9). The Control condition has low values at trials 8 & 6 which pull the average down, otherwise the majority of values are between 600-650ms. In the second Block Section the conditions diverge more clearly. The Control condition shows consistently lower values except for trials 14 & 15 which peak around 650ms. The Match condition peaks at trial 13 (650ms) and then falls to a relatively stable level across the course of the block, at approximately 600ms. The Mismatch condition starts with high values on trials 11-14 and then drops off towards the end of the block, with a trough at trial 19 (approximately 575ms). In comparison to the first block, the second Block Section therefore shows stable decreases for all conditions, as reflected in the averages. In the final section, the Control condition is on a par with the other two conditions and then drops off considerably towards the end, with trial 30 coming close to 550ms. The Match condition is quite stable, with no peaks or troughs. The Mismatch condition shows more variation, with high values at trials 21, 26 and 27. The averaged (and significant) difference between the Mismatch and Control conditions is reflected in the trial-to-trial data although there is still considerable overlap in each Block Section.

6.5.4. Discussion

There were no overall differences between conditions for perceptual sensitivity (d'), however, across the course of the block there was a significant difference in the way sensitivity changed for each condition. Under congruent conditions perceptual sensitivity increased, under incongruent conditions sensitivity decreased and under control conditions sensitivity was stable and then dropped towards the end of the block. Experiment 6.2 showed a drop in perceptual sensitivity under incongruent conditions, similar to that found here. However, whereas 6.2 showed no difference between congruent and control conditions, here the congruent condition showed an increase in perceptual sensitivity relative to the incongruent and control conditions. The incongruent condition did not show a significantly different pattern compared to the control, but it did show a sharper decrease in sensitivity (see Figure 6-3a). This was expressed in the stronger interaction between the congruent and incongruent conditions

as compared to the interaction between congruent and control conditions. A closer inspection of the changes from trial to trial show that the patterns seen in the statistical analyses were not supported by stable differences between Word Categories, e.g. the incongruent conditions being consistently lower than the congruent condition in the third block section. The descriptive analysis showed that there was considerable and regular overlap between Categories, but for the congruent condition the data did show a consistently elevated d' level towards the end of the block. Taken together these results strongly suggest that blocked presentation is essential for producing semantically modulated changes in sensitivity following single word comprehension. In other words, there has to be a consistent directional context for congruence effects on d' levels. In 6.2, when the words were unsynchronised to RDK onset, incongruent interference occurred. In this experiment when the words do predict RDK onset, congruent facilitation is seen. In the experiment where word and RDK onset were unsynchronised and RDK onset was unpredictable (6.2), semantic information cannot be used to enhance task performance, but it is able to impair performance. In contrast, when word onset predicts RDK onset, it is possible that semantic information can be used to improve perception, similar to the perceptual learning seen elsewhere in vision. When irrelevant motion patterns are presented and one particular direction is paired with target trials on a focal task, sensitivity is only increased to the direction of motion that predicts (or is associated with) target trials (Seitz & Watanabe, 2003; Watanabe, Nanez & Sasaki, 2001). These findings indicate that learning takes place for irrelevant stimuli that are positively correlated with target trials (Seitz & Watanabe, 2003). The important point to take from these studies is that the visual system selectively integrates 'irrelevant' information that is associated with target presentation. In our experiment we have presentation of irrelevant, but predictive and contextually congruent, stimuli that increase performance in the central task. In congruent blocks, words which are in the same direction as the motion being detected cue experimental trials. Over time, this improves perception of that motion direction, sensitising the visual system. Crucially, it is not simply the presentation of a word (i.e. sound) that provides an effective cue to target onset, thereby increasing sensitivity. An improvement in performance is only seen when all the words in a block refer to the same direction as the motion being detected. In contrast, when the semantic information in the words does not create a consistent context (the control condition) or creates a directional context that is incongruent with the visual motion, sensitivity drops over the course of the trial. The visual system is *always* flexibly responding to semantic information, and when that

information is congruent with the target information, perception improves. When that information is incongruent or uninformative, perception declines, as if the information was being integrated automatically but this time to the detriment of perception. As demonstrated above, perception is also impaired by incongruent information when RDK onset is unpredictable and the semantic information cannot be used for perceptual learning (Experiment 6.2). The effects on perceptual sensitivity are task dependent, with integration working to the benefit of task performance when semantic information is congruent and reliable. Note that the perceptual learning mechanisms that could produce such effects are widely thought to operate at low levels of visual processing (e.g. Seitz & Watanabe, 2003); therefore they can be combined with strong embodied explanations which predict that semantic access necessarily effects low level perception.

There were no significant differences for either measure of decision threshold. The descriptive analysis for Beta showed a trend for incongruent conditions to have higher values than control conditions, indicating more conservative decision making. However, a lack of significant effects precludes any firm conclusions. So, whilst the words have a selective effect on perceptual sensitivity, decision thresholds do not change. Therefore, the integration of semantic and perceptual information does not come via top-down mechanisms such as attention. As stated above, if effects on perceptual sensitivity come independently of similar (i.e. congruence) effects on decision thresholds or reaction times, this supports a separate mechanism. The modulation of d' appears to be a product of a more direct link between semantic and visual systems. Interestingly, although they showed the same pattern in d' (a decrease in sensitivity) the incongruent and control conditions did differ in reaction times. Participants produced the slowest reaction times under incongruent conditions as compared to the control condition. Longer reaction times support the conclusion that incongruent blocks were more disruptive than control blocks at some level, slowing the response. Experiment 6.2 did not show congruency effects on reaction times so it is interesting to find them here. In fact, this is the only experiment to show a clear congruency effect in reaction times. It is perhaps due to the tighter relationship between linguistic and motion stimuli, since the semantic processing of the word is arguably still in process during RDK presentation and therefore more able to directly interfere with the preparation and execution of a response. Reaction times also showed a main effect of block section, with participants speeding up during the middle of each block and slowing again towards the end. This is best interpreted as acceleration in response

execution with practice, followed by fatigue effects (or anticipation of the end of the block).

6.6. *General Discussion*

The pilot experiments used a motion direction judgement to explore the influence of comprehension on perception; however, this conflated the perception of directional motion with response preparation. To avoid these problems, the rest of the experiments used a motion detection task, Table 6-7 gives a descriptive summary of the results. In this task participants detected coherent motion and SDT analysis allowed us to separate level of processing. In particular, we were interested in whether comprehension would influence d' . This dependent measure represents the strength of the signal as detected by the participant, being directly proportional to the separation of distributions of signal and noise in a detector during a 2-alternative forced choice task. In this task it is therefore equated to low level visual processes that detect directional motion.

Table 6-7: Descriptive summary of results

Linguistic Stimuli	Results
Blocked single word presentation unsynchronised to target	d' : Interference from Mismatch conditions criterion : Facilitation from Match conditions RT : No congruence effects
Blocked single word presentation synchronised to target	d' : Facilitation from Match across block criterion : No significant changes RT : Interference from Mismatch conditions
Single word presentation in random order	d' : Interference from Match and Mismatch conditions criterion : No congruence effects RT : No congruence effects
Partial sentence comprehension	d' : Facilitation from Match conditions criterion : Facilitation from Match conditions RT : Trend for facilitation from Match conditions

The direct engagement hypothesis of embodied theories states that it is just such processes that should be recruited by semantic representation; those sensory areas which are used to perceive motion are recruited in the representation of semantic motion. The four experiments manipulated the strength of linguistic context and the concurrence between linguistic and motion stimuli. First, effects of congruency will be reviewed and then I will briefly discuss effects of motion direction, before concluding.

6.6.1. Congruency

In the first experiment (6.2), blocked single words were presented for passive comprehension and were unsynchronized to the presentation of the motion stimuli. The results showed that d' was reduced for incongruent conditions. In the second experiment (6.3), a randomly selected single word preceded each motion stimulus. In this case there was a trend for d' to be lower for all directional words regardless of congruence. In the third experiment (6.4), motion stimuli were presented within a sentence context at verb offset, perceived during the remainder of the sentence. Results showed a trend for higher d' in congruent as compared to control conditions. Finally, in the fourth experiment (6.5), motion stimuli were presented to coincide with word offset, and words were passively comprehended in semantically coherent blocks. Over the course of the block, d' increased in congruent conditions and decreased in incongruent and control conditions.

Therefore we have 4 experiments which show effects of motion words on d' , a low level visual process. What is intriguing is that the effects on d' often diverge from those for decision threshold and reaction times; converging in one experiment out of four (6.4). Therefore, the most important conclusion is that comprehension reliably influences low level visual processes, affecting perceptual sensitivity to motion. The relationship between semantic representation and perceptual information is automatic, close and direct, since it operates independently of other decision and response mechanisms. This supports strong embodiment which predicts that interactions between comprehension and perception occur at low levels of perception and that these interactions are always present.

However, the direction of effects changes across experiments so despite the fact that low level interactions are always present, they are more in line with a weaker version of embodiment (i.e indirect engagement that mediates activity in sensory-motor systems). There is a pattern (3 out of 4 experiments) for incongruent reduction or congruent improvement in perceptual sensitivity. In the one experiment where semantic access did not overlap with RDK presentation (6.3) there was a general interference effect from directional words, reducing d' as compared to the control condition. So, congruency effects occur when there is contiguity between primary semantic activation (i.e. immediately following comprehension) and motion stimuli. To try and better understand what this pattern means, it is worth noting what we have not found. We do

not see a situation where incongruent conditions improve perceptual sensitivity (relative to the congruent or control conditions). We also do not see a situation where semantic and perceptual information are contiguous and congruent conditions reduce perceptual sensitivity. Congruent interference has been evidenced elsewhere and explained as competition for the same resources (Kaschak et al, 2005; 2006; Bergen et al, in press). As in the experiments that looked at visuo-spatial attention, our results do not support this explanation from strong embodiment. However, congruent facilitation and incongruent interference can be explained by a common mechanism. In the motion detection tasks we use, the RDK stimuli present ambiguous evidence for coherent motion to the visual system. A decision has to be made about whether the dots move in a given direction or not, i.e. whether a particular motion event has occurred. The experimental lexical items access or activate semantic information which in essence refers to a motion event. When the referenced motion event is the same as the motion event to be detected, the visual system improves. The semantic information primes the visual system and therefore facilitates processing of the motion signal. When the referenced motion event conflicts with the target event, the semantic information may prime the visual system in the 'wrong' direction and therefore impede processing of the motion signal. This happens when the semantic information is currently active as a consequence of immediate comprehension. This process results in incongruent interference when there is no predictive relationship between the stimuli, so it interferes with processing. However, perceptual learning may take place when there is a predictive relationship (experiment 6.5), with semantic content priming perception. It's possible that the effects from semantics are short lived and not comparable to re-activation of that same semantic content if a word meaning has to be maintained or indexed to answer comprehension questions (as is the case in 6.3).

The crucial question is whether the priming of the visual system occurs *without a motion detection task*. This would support direct engagement as it implies that the motion system is necessarily affected following comprehension. Conversely, if effects are not seen this implies that the visual system has to be engaged by a motion task in order to be affected by semantic motion. It would be possible to test this by using a contrast judgment with moving gratings, here the task is to judge which of two patterns are brighter and so it does not involve attention to motion. If participants are exposed to one direction of motion, they consequently show a higher contrast threshold for that direction, since directionally selective neurons have been fatigued (Levinson & Sekuler,

1975; 1976; 1980). An interesting experiment would be to assess whether exposure to motion words affects contrast sensitivity to congruent or incongruent motion signals. If motion words could influence the perceptual sensitivity to the contrast of moving gratings, this would support a direct effect of motion words on motion processing areas. Crucially, it would remove the possibility that attention to motion (or whatever target perceptual event) is necessary for an interaction between semantic and perceptual motion to occur¹³.

In two experiments there is congruent facilitation of d' , and in one experiment incongruent interference. The changes in d' are not consistent across tasks (i.e. always in the same conditions) so it is interesting to look at how the visual system integrates perceptual information during tasks. Learning mechanisms within the visual system may be playing an essential role; at best selectively integrating any information that can benefit task performance and at worst unable to counteract information that conflicts with perception¹⁴. We found that when words were unsynchronised to RDK onset, incongruent interference occurred (6.2). In contrast, when the linguistic stimuli do predict RDK onset, congruent facilitation is found (6.4 and 6.5). When word onset predicts RDK onset, semantic information can be used for perceptual learning. The visual system selectively integrates 'irrelevant' information that is associated with target presentation (Seitz & Watanabe, 2003; Watanabe, Nanez & Sasaki, 2001) and a predictive relationship (contiguity) allows congruent information to be exploited for the benefit of perception. A non-predictive relationship means that congruent information cannot be exploited, but that incongruent information interferes. In the attention literature, target detection (d') improves when the location of the target is cued and declines when the target is preceded by an invalid cue (Hawkins et al, 1990). The results for criterion are less consistent, but it has been suggested that attentional effects such as these increase the salience of information at attended locations, resulting in more accurate processing and a lower criterion (Hawkins, et al, 1990; Blaser, Sperling & Lu, 1999). Thus, external cues that are relevant for target detection are able to influence low level visual processes so it is possible that congruent and incongruent semantic information acts in a similar way, modulating detection in a top-down manner by increasing the salience of a particular motion direction. This is not to say that the effects of comprehension are isomorphic with attention (i.e. they affect attention which

¹³ My thanks to Professor Alan Johnston for suggesting this task.

¹⁴ As noted above, a key question is whether these learning mechanisms are causing the influence of the

then affects perception), but they may operate in a similar manner that is not a product of direct engagement. However, perceptual learning mechanisms that operate on low level motion processes could combine with the embodied simulation of motion to produce the effects we have seen. Therefore, it is even more critical to see if the influences on motion are tied to motion detection; if they are not, this is very strong evidence for strong embodiment.

Moving on to measures of decision threshold (c and β), in two experiments (6.3 and 6.5) there were no effects on criterion. It is not clear what conditions produce no stable effects on criterion and it is possible that trends in these experiments were not reliable due to a lack of power. For example, effects on criterion may only be seen when the linguistic manipulation is strong enough; they are found for blocked single word presentation when RDK presentation is unpredictable and partial sentence comprehension. It is not surprising that randomly ordered single words that are temporally separate to RDKs (6.3) do not affect criterion and it may be that the blocked single words synchronized to RDK presentation (6.5) is also not a strong enough context. Note that the d' effects we found in 6.5 did not show main effects between conditions, but only changes over the course of a block; suggesting that the linguistic context was weaker than in 6.2. In two experiments, effects for criterion were consistent (6.2 and 6.4), with congruent conditions producing reduced criterion values. In 6.4 (sentences) the β results corroborated the patterns for c with reduced values for congruent conditions. This again appears similar to what has been found in the attention literature. There is some evidence that when a location is cued, criterion is lowered for that location relative to a neutral or invalid cue (e.g. Hawkins et al, 1990). If we apply this logic to our experiment, the congruent context provides a similar benefit for motion detection, with a criterion shift that reflects less effortful processing. This suggests that congruent conditions produce something like decision priming. A lowered decision threshold means that target events have to breach a lower value to be detected, so decisions are biased towards a positive response. It may be that under congruent conditions less sensory evidence is needed to make a decision so criterion values are lowered (e.g. Sperling, 1984). This is supported in experiment 6.4 since there was also a tendency for reaction times to be faster under congruent conditions, indicating that less time was needed to decide that a target appeared (i.e. less evidence needed to be collected). It is interesting to note that these kinds of decision biases may

produce the facilitation that is seen in reaction times studies elsewhere (see Section 2.1.2.3). This point will be discussed in more detail in the final chapter.

Finally, we turn to reaction time. This measure reflects how long it takes for all processes from perception to response. It could therefore reflect incremental increases/decreases across all decision levels or substantial increases/decreases in one condition, or a combination of these. Thus, reaction times measure overall task difficulty. Changes in response preparation/execution may affect reaction times more strongly than changes in the ease of perceptual/decision processes but the reaction times do not consistently mirror the congruency effects in d' or criterion (doing so in only one experiment, 6.4). However, reaction times do show effects of motion direction (6.3, 6.4) which were also present in other measures; therefore we can say that reaction times do reflect strong effects at perception/decision stages but are more noisy when it comes to the weaker congruency effects. This is probably due to the response preparation/execution stage that is influenced by other factors. For example, in 6.2 reaction times were faster to Down words, this has been tentatively interpreted as Down words priming the downward pressing of buttons. The alternative explanation is that Down words were less interfering than Up words, although it is not clear why this should be the case. In experiment 6.4 there was a tendency for congruent conditions to be faster than control conditions and in 6.5 the control condition was faster than incongruent conditions. Similar to the pattern in d' , reaction times show interference from conflicting information and facilitation from concordant information. In 6.4 congruence effects were present at all stages (perceptual sensitivity, criterion and reaction times). In 6.5 reaction times showed a divergence where d' (and criterion measures) did not. Participants were faster during the control condition than in incongruent conditions, however there were no differences in d' and a weak trend in criterion. Responses may have been slowed down to avoid making mistakes, i.e. to counteract the incongruent words. In contrast, no such compensation is needed in the control condition so response times were faster. This would be a strategic manipulation of response times for the benefit of task performance, and one that did not result in better accuracy. As incongruent or neutral information could not be used to the benefit of perception this did not result in improved accuracy for incongruent conditions (i.e. similar to the speed-accuracy trade-off seen in 6.3, see below). Alternatively, it may have taken longer to prepare and execute a response in incongruent conditions because the participant was responding that one direction of motion was present whilst the

words referred to the opposite direction. This would be a similar result to the pilot experiments where incongruent conditions extended response times, i.e. the Simon effect. The most important point is that reaction times do not consistently show effects that are similar to those found for d' or criterion. Where reaction times do show similar effects it is not safe to interpret them as direct reflections of lower processing levels; especially since other explanations, that take into account the response stage, do just as well.

In conclusion, the results can only be explained by some form of embodiment, interactions between comprehension and low level motion processing are consistent across all experiments. This supports strong embodiment and suggests that the connection between semantic and perceptual content is based at this low level. Effects on criterion and reaction time were more variable and dependent on stronger linguistic contexts. However, some aspects of the results are best interpreted by task based processes such as perceptual learning; this is in line with both weak and strong versions of embodiment as perceptual learning operates at low levels of visual processing. It is easier to explain the results with an automatic and low level connection between semantics and perception than to invoke top-down factors which should have also affected criterion and reaction time measures more consistently.

6.6.2. **Motion Direction**

In experiments 6.3, 6.4 and 6.5 (see Figure 6-3) there were effects motion direction on d' ; in all cases the pattern was for downwards motion to have higher d' values than upwards motion. It appears that at least in one case this (6.3) this may have been due to a speed accuracy trade-off, participants were faster to respond to Upwards motion and less accurate as a result. However, in two other experiments the tendency for Upwards motion to have lower d' cannot be explained by a strategic trade-off with response speed. In one experiment (6.4) upwards motion also had higher β values in two out of three conditions (with this difference attenuated in the Control condition), suggesting that it was less salient. Criterion was raised as more information was needed to establish that upwards motion was present. The picture that is emerging shows that Upwards motion is harder to perceive and there is limited evidence that participants are more conservative when they respond to it. Elsewhere in the visual perception literature there is no supporting evidence for a difference between Upwards and Downwards

motion, so the effects may be a product of the dual-task methods we use. Dual-task methods necessarily split attention and processing resources across different tasks; when fewer resources are available, latent differences could turn into processing advantages. Downwards motion is more ecologically frequent than upwards motion, since gravity causes things to fall. This increased exposure to downwards motion may result in a greater sensitivity to it, in line with perceptual learning following practice (Gold, Sekuler & Bennett, 2004). At low coherence, upwards motion may appear more 'random' than downwards motion and participants would therefore be less sensitive and more conservative during detection (See Chapter 7 for additional evidence for this argument).

7. Transcranial Magnetic Stimulation

The RDK stimuli that we use engage a set of well characterised processes in first order visual motion processing and are also known to activate particular cortical areas (Blake, Sekuler & Grossman; 2004). The area V5/MT+ is the principle cortical area implicated in motion processing (e.g. Zeki et al, 1991). For example, the activity of V5/MT+ increases as the level of coherence of an RDK increases (Rees, Friston & Koch, 2000); this supports its role as the primary area responsible for detecting coherent motion signals. As stated already, the direct engagement hypothesis of strong embodied theories proposes that sensory and motor areas are recruited for semantic representation. As V5/MT+ is the cortical area widely held to be responsible for early motion processing, the prediction from strong embodiment is that the semantic representation of motion words is based in V5/MT+ (e.g. see similar arguments about the involvement of the primary motor strip in words and sentences that refer to motor actions; Pulvermüller, 2001). As d' measures perceptual sensitivity, it implicates early motion processes like those at V5/MT+. For example, double pulse TMS at V5/MT+ has been shown to reduce d' for motion detection with RDK stimuli (Silvanto, Lavie & Walsh, 2005). Therefore, effects of comprehension on d' are effects of semantic content on early motion processes, possibly including those at V5/MT+. In the previous chapter, experiment 6.2 demonstrated that the passive comprehension of blocked single words produced a decrease in d' when the direction of the words was incongruent to the direction of visual motion. The experiment reported here sought to replicate that experiment with the addition of TMS at V5/MT+. In this way, the direct involvement of the cortical region involved in motion processing could be assessed. The general hypothesis was that if the semantic content of motion words has some direct effect on motion processing areas, the interference in motion processing (reduced d') caused by TMS should interact with the interference during incongruent blocks that was observed in 6.2. The specific hypotheses are spelt out in more detail below. A few changes were made to the methodology so it was more amenable to TMS, in particular a shorter RDK duration (e.g. Silvanto, Lavie & Walsh, 2005; Beckers & Zeki, 1995) and presentation of the motion stimuli only in the right hemi-field with TMS stimulation only to the left hemisphere V5/MT+ (e.g. Beckers & Zeki, 1995). Right hemi-field presentation is consistent with the lateralisation of language in the left hemisphere (for most right handed people), therefore effects of language on perception are more likely to be observed for stimuli in the right hemi-field (Gilbert, Regier, Kay, & Ivry, 2006; 2007).

7.1. *TMS during Motion Detection and Passive Comprehension of Blocked Single Words*

As demonstrated in 6.2, passive comprehension of motion words impaired perceptual sensitivity during incongruent blocks (Meteyard et al, in press). This experiment used the same methodology as 6.2 with the addition of double pulse TMS time-locked to RDK presentation, on alternating blocks. NoTMS blocks (in which sham stimulation was applied) were used as a control. In line with previous findings we expect TMS to reduce d' for motion detection (Silvanto, Lavie & Walsh, 2005) and there are three possible ways that this direct manipulation of activity at V5/MT+ could interact with the incongruent interference (reduced d') produced by comprehension: (i) The interfering effect of incongruent motion words is increased relative to the Control and congruent conditions. This would suggest that the motion words primary effect is outside V5/MT+ as the influence of the words is greater when V5/MT+ is disrupted. (ii) The interference of incongruent words is decreased. This would support an effect of comprehension that does involve V5/MT+, as this effect is attenuated when activity at V5/MT+ is disrupted. (iii) There is no change and the interference effect is of the same magnitude with or without TMS. This would also support an effect of comprehension that is outside V5/MT+, since direct changes in activity at this location do not change the effect of comprehension. It is not clear what effect TMS will have on decision processes (as measured by c and β). Experiment 6.2 showed that C was lower for congruent conditions, we expect to replicate this finding in the NoTMS blocks. In the TMS blocks, changes in criterion measures will shed light on how changes at low levels of perception affect decision processes.

7.1.1. **Participants**

14 Native English speakers (5 female) with a mean age of 28.8yrs (SD = 8.6yrs) were recruited from the UCL Psychology Department Subject Pool in return for monetary compensation of £7.50 per hour. All were right handed and had normal or corrected to normal vision. 12 participants attended both experimental sessions and 2 participants attended the first session only.

7.1.2. **Method**

7.1.2.1. *Design*

A 2x3x2 repeated measures design was adapted from Experiment 6.2. The independent variables were the direction of motion of the RDK (upwards or downwards), the Word Category (Control, Match and Mismatch) and the application of TMS (TMS, no TMS). The dependent measures were the accuracy of detecting motion on each trial, and the reaction time (milliseconds) to make the judgement. Participants attended two sessions separated by 3-7 days. Within each session, direction of detected motion (upwards or downwards) was kept constant with a balanced order across participants.

Participants completed the Phosphene Threshold Estimation, Motion Threshold Estimation and Motion Detection tasks in that order. During Motion Detection, participants completed 2 blocks of 20 RDKs (10 coherent and 10 random motion) set at an easy high coherence of 50-70%. They were then given as many practice trials as they needed (in blocks of 6) with the RDK at their threshold value. For the Motion Detection task there were 2 runs of 12 blocks of 20 motion detection trials comprised of 10 coherent motion and 10 random noise trials presented in randomized order. The 12 blocks consisted of 4 different item orders for the 3 Word Categories. Presentation order was pseudo-randomised using a Latin squares design so that the same Category was not presented in consecutive blocks and TMS was applied to each word Category in a balanced manner.

7.1.2.2. *Materials*

These were identical to Experiment 6.2

7.1.2.3. *Graphic Display*

Stimuli were identical to Experiment 6.2 except that the duration was 5 consecutive frames (total motion time= 60ms; speed = 21 °/s) and they were centred 5° to the right of the central fixation point. Viewing distance was 50cm. Stimuli were

generated using Cogent toolbox (www.vislab.ucl.ac.uk/Cogent/) for MATLAB (Mathworks, Inc) and presented at 85Hz using a CRT display (resolution = 800×600, 14" Sony Multiscan 110ES).

7.1.2.4. Procedure & Apparatus

TMS stimulation and location parameters

The stimulator was a Medtronic MagPro Model X100 (Medtronic A/S, Tonsbakken, Denmark) machine delivering current to a MC-B70 figure-of-eight coil (diameter = 70mm). In TMS blocks, the coil was fixed tangential to the skull with the handle pointing anteriorly at ~90 degrees to the axis of the neck. In the sham blocks, the TMS coil was rotated 90 degrees and its far edge was left touching the MT spot in order to mimic the auditory and somatic sensation of TMS without delivering any magnetic stimulus to the brain

We localized MT using a functional method (e.g. Silvanto et al. 2006). A 3×3 grid of potential target points (centre-to-centre distance = 1.0cm) was centered at 3cm dorsal and 5cm left lateral from theinion. TMS was delivered to the left hemisphere only. Double pulse TMS (intensity = 70% machine output) was delivered over each of the nine points sequentially to determine the site most reliably giving rise to large visual phosphenes in the contra-lateral hemifield, which we then chose, as the stimulation site for the experiment. Functional localization is valid as it takes into account the variation in extent and location of V5/MT+ across human participants (Malikovic et al., 2007) and has been supported by studies that show comparable TMS disruption of motion perception following fMRI motion localizers (Sack et al., 2006) or functional localization (Silvanto et al. 2005) of V5/MT+. The use of double pulse TMS allowed us to increase the chances of disruptive TMS effects (compared to single pulse) while still maintaining good temporal resolution (O'Shea et al, 2004). In each Experimental trial, double pulse TMS with an inter pulse interval 50ms was applied 100ms from motion onset (i.e. one pulse at 100ms and one pulse at 150ms following stimulus onset).

Phosphene Threshold Estimation

Once V5/MT+ was functionally localized, participants then went through a staircase procedure to determine their phosphene threshold to be used as the baseline for TMS pulse intensity in Experimental trials. Double pulse TMS was applied and participants

reported the presence or absence of a phosphene. TMS intensity was raised following a negative response and lowered following a positive response. The threshold was taken as the TMS intensity that induced V5/MT+ phosphenes in the absence of concurrent visual stimuli on 50% of stimulation trials. On experimental trials, double pulse TMS was delivered at 110% of participant-specific phosphene threshold. The average Phosphene Threshold was 39.6% (SD = 8.9) of the maximum output intensity. To ensure that stimulation was effective, intensity was set to 35% of maximum output if a participant's Phosphene Threshold resulted in a 110% value that was less than 35% (this was the case for 2 participants in both sessions, and 3 participants in one session only).

Eye movement control

Horizontal eye movements were recorded using infrared light transducers in the Skalar IRIS 6500 system attached to the head rest (IRIS; Skalar Medical, Delft, the Netherlands; sampling rate: 1000 Hz - Analog-to-digital converter card Type PCM-DAS 16d/12, Computerboards, Pittsburgh, PA) and analyzed offline with DASyLab 5.0 (DATALOG, Mönchengladbach, Germany). Eye traces were recorded for -200 to 200ms peri-stimulus time on every trial and the equipment was recalibrated between blocks. Online monitoring during the experiment ensured that in > 90% trials participants maintained fixation. A break of fixation was defined as > 2 degrees deviation from fixation point or a blink artifact in the critical (+/-200 ms) period.

Motion Threshold Estimation

This was identical to Experiment 6.2 except that the RDKs were 60ms duration and centred 5 ° to the right of fixation (see Graphic Display).

Motion Detection

During practice trials a tone sounded if they judged the RDK coherence wrongly. Experimental blocks were identical to Experiment 6.2 except that on alternating blocks double pulse TMS or Sham TMS was applied in synchrony with the presentation of the RDKs (see TMS stimulation and location parameters). The alternation of TMS and No TMS blocks was randomly selected for each run. To minimise fatigue, participants had

short breaks at the end of each block and refreshments during the gap between the 2 runs of 12 blocks.

7.1.2.5. Analysis

2 subjects did not complete the second session so they were not included in the overall analysis. However, they were included in the post-hoc analysis of Session 1 effects only (see Results below). Analysis was identical to Experiment 6.2 except that values were entered into a 2 (TMS or No TMS) x 2 (upwards or downwards motion direction) x 3 (control, match or mismatch condition) repeated measures ANOVA. Reaction Times above 2 standard deviations for each participant were removed; this affected 2.36% of the data. There were no significant differences in criterion (as measured by c or β) and no significant differences in reaction times so the values are not reported.

7.1.3. **Results**

7.1.3.1. d'

There was a significant main effect of Motion Direction ($F(1,11) = 8.327$, $p-\eta^2 = 0.431$), with Downwards motion having mean d' values 0.313 units higher than Upwards motion.

There was also a significant main effect of Word Category ($F(2,22) = 4.723$, $p-\eta^2 = 0.300$), planned comparisons showed that this was driven by the Control condition having d' values 0.171 units higher than the Match condition ($F(1,11) = 11.636$, $p-\eta^2 = 0.514$). Figure 7-1 illustrates this main effect across TMS Application and testing Session.

There was no main effect of TMS. To find out whether TMS was ineffective or variable across testing sessions, we compared the d' values for TMS and No TMS blocks across the First and Second testing session. This showed that there was a significant interaction between TMS and Session ($F(1,11) = 14.280$, $p-\eta^2 = 0.565$). As compared to the No TMS condition, d' values were lowered by the application of TMS in the first session, but raised in the second session. The interaction is principally

driven by the change in d' for No TMS, rather than TMS, blocks. In the first session, participants mean d' value in No TMS conditions is 1.304, in the second session it falls to 0.976 ($t(11) = 2.098, p=0.058$). In contrast, mean d' in TMS conditions is 1.129 in session 1 and 1.205 in session 2, see Table 7-1.

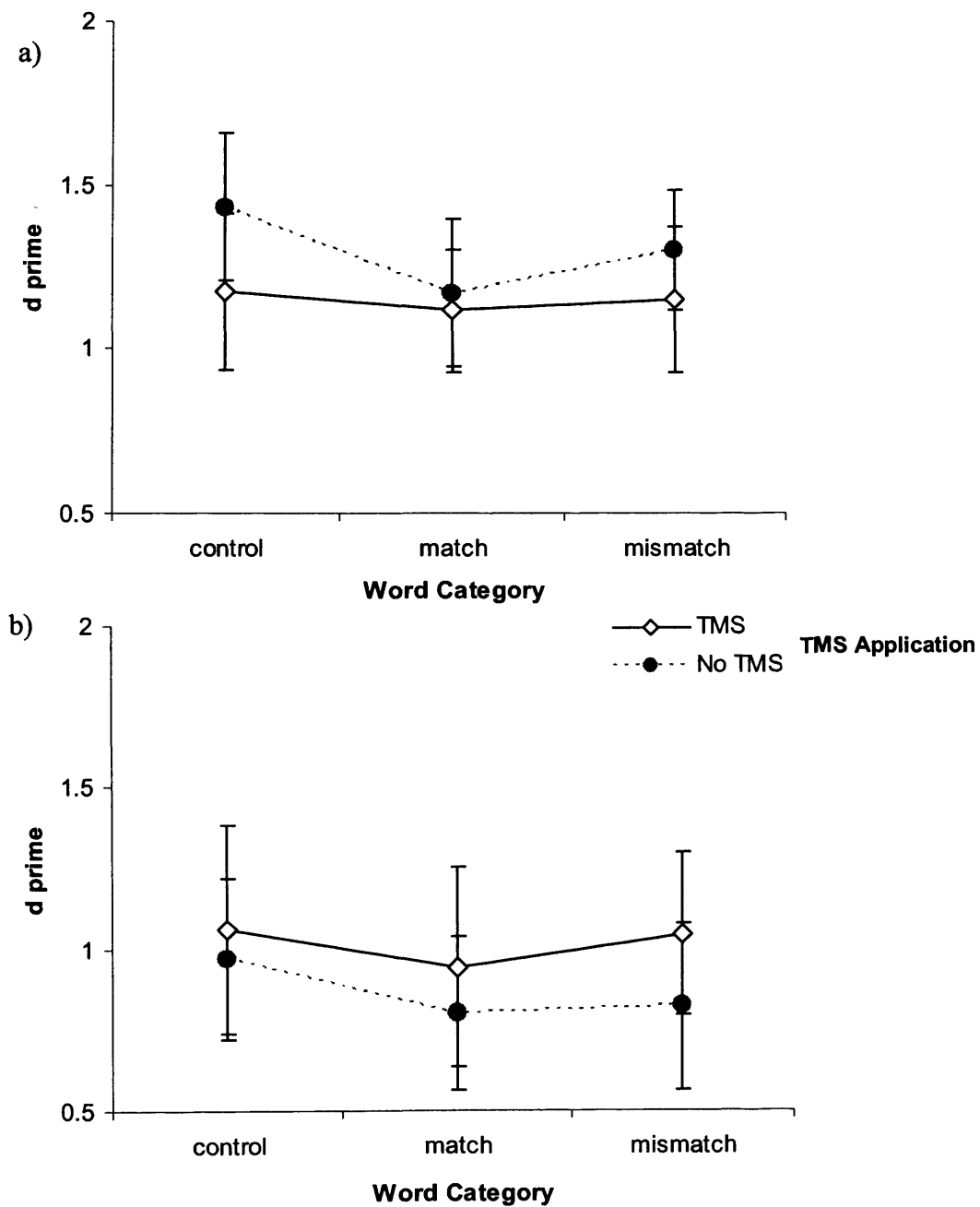
Table 7-1: Mean d' values by TMS Application and Session*

Session	TMS Application		<i>Difference</i>
	TMS	No TMS	
First / 1	1.129 (0.68)	1.304 (0.78)	0.175 (0.73)
Second / 2	1.20 (0.60)	0.976 (0.49)	-0.224 (0.55)
<i>Difference</i>	0.071 (0.64)	0.331 (0.64)†	

*Standard deviation in brackets

† this difference drove the interaction between TMS and Session

Figure 7-1: d' for Session 1 (a) and Session 2 (b) by TMS/No TMS and Control, Match, Mismatch conditions



We then performed follow up comparisons comparing TMS and NoTMS blocks within session. In Session 1, the main effect of TMS was not significant ($F(1,11) = 3.316$, $p = 0.096$, $p-\eta^2 = 0.232$) although there was a trend for TMS blocks to have mean d' values 0.175 units lower than No TMS blocks. In Session 2, the main effect of TMS was significant ($F(1,11) = 13.821$, $p < 0.005$, $p-\eta^2 = 0.557$), with TMS blocks on average 0.229 units higher than No TMS blocks. Figure 7-1 shows TMS and NoTMS blocks across Control, Match and Mismatch conditions, for Session 1 (a) and 2 (b). A post-hoc analysis on d' values for Session 1 only (including the 2 extra participants who had only completed one session) showed trends, but no significant differences. The main effect of TMS approached significance for an uncorrected p value of 0.05 ($F(1,13) = 3.511$, $p = 0.084$, $p-\eta^2 = 0.213$) as did the main effect of Word Category ($F(2,26) = 2.596$, $p = 0.094$, $p-\eta^2 = 0.175$). These results are likely produced by a lack of power since half the data set was excluded for most participants.

7.1.3.2. Criterion, Beta and Reaction Times

There were no significant differences.

7.1.3.3. Threshold Values

To explore why participants baseline d' values varied between Session 1 and Session 2, the individual threshold values (expressed as the percentage of dots moving in one direction) were compared using a paired samples t -test. There was a significant difference between mean coherence thresholds for Session 1 at 18.5% and Session 2 at 14.4% ($t(11) = 4.041$, $p < 0.005$); participants coherence thresholds dropped by 4.1% between the first and second testing sessions. There was no difference in thresholds between upwards and downwards motion ($t < 1$).

Table 7-2: Mean coherence thresholds (%) by Session and Motion Direction for Experiment 6.2 - 7.2

Experiment	Condition			
	Session		Direction	
	1	2	Up	Down
6.2 Blocked Word	6.1 (1.9)	7.5 (3.0)	7.6 (3.1)	6.0 (1.6)
<i>Mean Difference</i>	1.47 (3.0)*		1.51 (3.0)*	
6.3 Single Word	7.6 (2.4)	7.0 (2.2)	7.3 (2.3)	7.3 (2.2)
<i>Mean Difference</i>	0.12 (3.5)		0.84 (3.4)	
6.4 Sentences	7.5 (2.2)	7.4 (3.7)	7.9 (3.7)	7.0 (2.1)
<i>Mean Difference</i>	0.12 (3.5)		0.84 (3.4)	
6.5 Blocked Single Word	6.9 (2.5)	6.3 (1.4)	7.0 (2.4)	6.2 (1.5)
<i>Mean Difference</i>	0.5 (2.5)		0.8 (2.4)	
7.1 TMS Blocked Word	18.5 (7.1)	14.4 (4.9)	16.0 (7.4)	17.0 (5.3)
<i>Mean Difference</i>	4.1 (3.5)*		1.0 (5.4)	

To check whether this pattern held for other Experiments, the thresholds for Session 1 were compared to session 2, and the thresholds for Upwards motion were compared to Downwards motion, using paired sample t-tests. See Table 7-2 for the results. These showed that changes in threshold levels only occurred for Experiment 6.2 (i.e. the experiment we sought to replicate here). In that experiment participants' thresholds were on average higher in session 2 than session 1 (by approximately 1.5%), and thresholds for upwards motion were on average higher than thresholds for downwards motion (again by approximately 1.5%). Therefore there is no tendency across experiments for participant's thresholds to drop between session 1 and 2.

7.1.4. Discussion

There was a significant main effect of Motion Direction, with Downwards motion having higher d' values than Upwards motion. This is a pattern evidenced in experiment 6.3 and 6.4, and supports the increased salience of downwards motion in the motion detection task (see Section 6.6 for a tentative explanation of this finding). There was also a significant main effect of Word Category, with congruent conditions producing decreased sensitivity (d') compared to Control conditions.

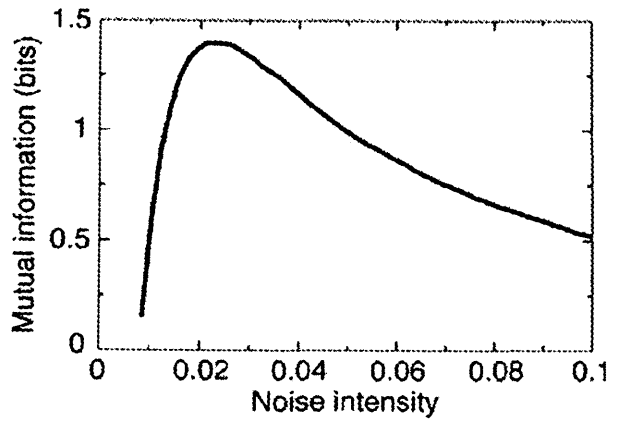
Both of the main effects were independent of TMS application. Based on the previous literature, we expected the application of TMS to reduce perceptual sensitivity to motion (as evidenced by lower d' values) (Blake, Sekuler & Grossman, 2004). However, we found that TMS marginally reduced sensitivity in the first testing session, but increased sensitivity in the second testing session. This produced a significant interaction between TMS Application and Session. Curiously, the interaction was driven by lower d' values in the baseline No TMS condition when Session 1 was compared to Session 2 (mean difference = 0.331). In support of this difference in baseline performance across Sessions, participants coherence thresholds were significantly lower in Session 2, falling by 4.1%. It is worth pointing out that because RDK stimuli were presented peripherally, both the motion thresholding and detection tasks were harder. It is difficult to maintain attention at a location other than fixation and some participants did report that it was hard to fixate on the centre whilst attending to the periphery. The reduction in threshold values is probably due to a practice effect across the two sessions, such that participants improved at attending to the periphery and had lower coherence thresholds in Session 2. The comparative nature of the motion thresholding task is different to the singular judgement required during motion detection; therefore better performance (lower coherence thresholds) in motion thresholding may make motion detection harder. A weak coherent signal that can be discriminated in comparison to a noise pattern may not be as easily detectable when it has to be extracted without comparison. In motion detection, the participant has to infer what a random pattern looks like and judge the perceived stimulus against an inferred, rather than a visible, alternative. It is possible that the lower threshold values in Session 2 produced worse performance in the baseline detection task, and in line with this the data shows a drop in baseline (NoTMS) performance. In contrast, values in the TMS condition were raised by a mean of 0.071 between Session 1 and 2. Thus, the

application of TMS resulted in broadly similar levels of sensitivity across both Sessions. Given that TMS introduces noise into cortical processing, the interaction between TMS and Session can be explained by Stochastic Resonance (Moss et al, 2004).

Stochastic resonance refers to the sometimes beneficial effect of noise during the processing of signal information, therefore being directly applicable to the perception of sensory signals (Moss et al, 2004). The focused magnetic field of TMS creates non-specific electrical activity in the cortical surface underneath it, essentially introducing noise to the activity of a particular cortical area (Blake, Sekuler & Grossman, 2004). In both Sessions, introduced noise (TMS) produced similar levels of sensitivity (d') during motion perception. However, because of the changes in the No TMS baseline conditions, this was evidenced as a decrease (Session 1) or increase (Session 2) in sensitivity. In both Session 1 and Session 2, participant thresholds corresponded to 81% accuracy on the thresholding task. However, in Session 2 participants were better at discriminating signal from noise during thresholding, leading to reduced coherence thresholds. This was likely due to practice effects with peripheral presentation; due to the stochastic nature of sensory systems, thresholds are not absolute or fixed (Moss et al, 2004).

As summarised above, a reduced signal during detection of a single event is more detrimental since there is no direct comparison. Therefore, in Session 1, participants' performance detecting signals around their threshold was reduced by the introduction of noise from the TMS pulse. In this case, the signal was at threshold so the TMS noise reduced sensitivity. Conversely, in Session 2, the signal was sub-threshold so the introduction of noise from the TMS pulse improved performance. According to Stochastic Resonance Theory a moderate amount of noise acts to amplify the signal as well as the noise, pushing it above the perceptual threshold. Conversely, smaller amounts of noise do not effectively amplify the signal and greater amounts of noise instead disrupt signal processing (as it swamps the signal). This creates an inverted 'U' function where, within a particular range, additional noise is actually beneficial to signal processing (see Figure 7-2).

Figure 7-2: Inverted U noise detection graph produced from a model of stochastic resonance. Moderate amounts of summated noise maximise sensitivity (from Moss et al, 2004)



For example, Kitajo et al (2003) found that participants' accuracy in monitoring the changing grey level of a visual stimulus in one hemi-field was improved by the addition of Gaussian noise in visual stimuli presented to the contra or ipsi-lateral hemi-field. The grey level stimulus was sub-threshold, therefore supporting beneficial role of stochastic resonance in detecting sub-threshold stimuli. Stochastic resonance occurs when there is a threshold, a sub-threshold stimulus and noise. In dynamic systems (like the natural world) the sub-threshold stimulus does not have a fixed value, varying over time in how close it comes to the threshold value, but never breaching it. The crucial finding is that the addition of moderate noise allows the signal to breach the threshold, for example at those time points when the distance between the sub-threshold signal and threshold are reduced. With moderate noise, threshold breaches are not random but constrained by the signal; therefore, signal information is preserved and improved detection of the sub-threshold stimulus is possible. In contrast, small amounts of noise do not increase power sufficiently so that the signal breaches the threshold level and large amounts of noise swamp the signal so that threshold breaches are essentially random (see Figure 7-3).

The use of two sessions (one for each motion direction) created unforeseen problems for the application of Transcranial Magnetic Stimulation. The tentative explanation is that practice effects which improved performance in the Thresholding task resulted in reduced performance during the Detection task. The moderate noise created by TMS then interacted with the sub-threshold signal in Session 2. The original hypothesis was that the application of TMS to the motion processing area V5/MT+ would provide some

insight into the interactions between the comprehension of motion words and the perception of visual motion.

Figure 7-3: An illustration of low (a), moderate (b) and high (c) noise added to a sub-threshold stimulus

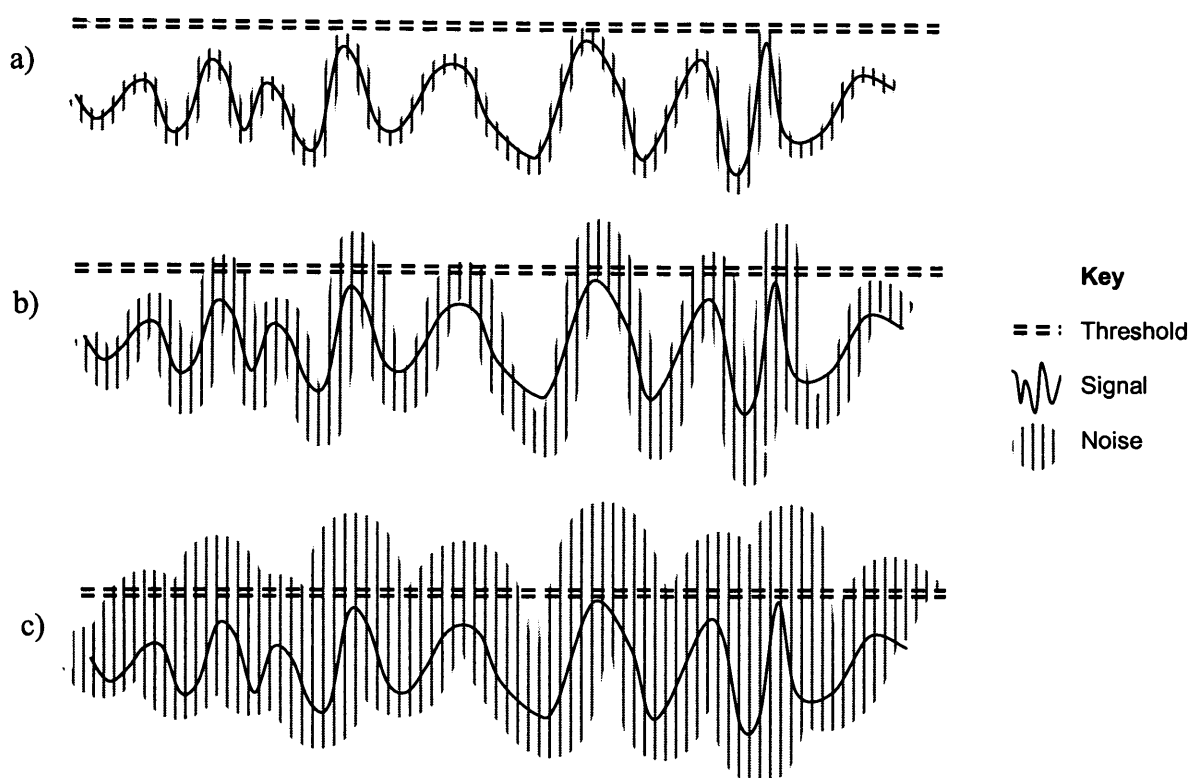


Figure 7-3 (adapted from Moss et al, 2004): The dashed line represents the threshold, the curved black line the sub-threshold stimulus and the vertical grey lines the noise (a) When small amounts of noise are added the sub-threshold stimulus does not breach the threshold (b) When moderate noise is added, the sub-threshold stimulus is able to breach the threshold whilst preserving characteristics (e.g. the phase) of the original signal, allowing better detection (c) When large amounts of noise are added the threshold is breached, but now little signal information is preserved and threshold breaches are essentially random and uninformative.

As the critical manipulation (TMS versus NoTMS) was ineffective we are unable to draw any firm conclusions about how TMS affected the interactions. However, regardless of TMS application and whether it reduced or increased d' , there was a significant main effect of Word Category; congruent conditions produced lower perceptual sensitivity. The effect of words would appear to be independent of direct stimulation on V5/MT+, and therefore independent of different levels of cortical activity at the area responsible for processing the coherent motion signals present in the

RDKs (Blake, Sekuler & Grossman, 2004). The semantic modulation of sensitivity could be outside the low level motion processing areas, such as V5/MT+, but pervasive enough to affect d' (rather than obligatorily affecting both sensitivity and decision processes). The interaction between Word Category and TMS never approached significance and Figure 7-1 illustrates that the main effect of Word Category was fairly consistent. Whilst the pattern suggests that the influence of directional words is independent of TMS, it is possible that the effects are additive *at V5/MT+*, such that motion words do directly engage V5/MT+ and the noise from TMS is summated on to the semantic effect. With different TMS parameters, perhaps causing more severe disruption of V5/MT+ processes, we may have seen an interaction between the two. The explanation of stochastic interference supports the fact that the TMS noise was moderate. Theoretically, if V5/MT+ processing is disrupted through catastrophic levels of noise, rather than impaired through moderate noise, the motion words should not be able to modulate V5/MT+ activity at all. If congruency effects are found on sensitivity when more disruptive TMS parameters are used, this would present quite strong evidence for their effects originating outside V5/MT+.

The main effect of Word Category contrasted with previous results: blocked word presentation produced effects on d' , however, the effect is in a different condition to that found in Experiment 6.2. There, the Mismatch condition had lower d' values compared to the Match and Control conditions. In this experiment, the Mismatch condition was not significantly different from the Match or Control conditions, falling in between the two, but the Match condition had reduced sensitivity relative to the Control. There were two methodological differences between the original method in 6.2 and its adaptation to TMS. First, the motion patterns were 60ms long (rather than 150ms) and they were presented in the right visual hemi-field (rather than centrally). Both changes increased the difficulty of the task, as demonstrated by the increased thresholds for this experiment, as compared to those for Experiments 6.2-6.5 (see Table 7-2). Either of these two changes in method could have produced the different congruency effects.

It is unlikely that shortening the motion pattern would produce this difference in conditions. The unsynchronised presentation of words means that their influence across the block is continuous and therefore it should affect perception of RDKs regardless of their length. However, since presentation was unsynchronised the very short duration means that RDKs may have been less likely to overlap with the presented words;

making the paradigm more like consecutive presentation than concurrent presentation. In this case, we might expect congruency effects to appear in different conditions. Elsewhere it has been proposed that congruent visual motion interferes with comprehension when the two are presented simultaneously, but facilitates comprehension when the two are presented consecutively (Kaschak et al, 2006). During simultaneous presentation congruent semantic and perceptual information compete for the same resources (i.e. direction selective neurons in motion processing areas), whereas consecutive presentation produces priming since semantic and perceptual information are temporally separate and access the same resource in turn. This explanation should also hold for comprehension interfering with motion, since the resources are the same. Here we find incongruent interference in one experiment (6.2) and congruent interference in another (the present experiment). Therefore the pattern is always interference, but the condition changes. It is not clear why this should be case and it is not straightforward to apply the same logic as that used in Kaschak et al (2006). If the shorter RDK patterns had produced something similar to consecutive presentation we should see some facilitation (congruent or incongruent) if the priming account is to be adopted. In sum, the shorter RDK presentation does not provide a straightforward account of the different congruency effects.

The major change in method, and the one that most likely led to the TMS x Session interaction, was peripheral presentation. It is thought that the neural representation of peripheral motion is different to central motion. In the monkey MT (the analogue to the human motion processing area V5/MT+), larger receptive field sizes are present with increased eccentricity (Blake, Sekuler & Grossman, 2004). It is assumed that the same is true for the human V5/MT+ (it has been found for human V1; Kastner et al, 2001). Neurones respond to motion across larger areas of peripheral vision than they do for central vision, therefore central motion is at a higher resolution relative to peripheral motion. Given the reduced visual acuity in the periphery, it is sensible that motion signals are collated across larger visual regions: when information is not perceived in as much detail it is pooled across larger areas. Therefore, it is possible that the influence of motion words on peripheral visual motion is different to the influence of motion words on central visual motion, due to differences in the neural representation of visual motion by eccentricity. Central motion is at a greater resolution so may be disrupted by incongruent semantic information (but not congruent information). When the visual signal is more detailed (central presentation) it may be less open to disruption from

semantic/contextual information so we only see a reduction when semantic information is opposed to the visual direction. In contrast, peripheral motion is at a lower resolution so may be disrupted by both congruent and incongruent semantic information; the signal has less detail so any extraneous direction information may affect processing. Note that although the incongruent condition was not significantly different to congruent or control conditions, as shown in Figure 7-1 there was a trend for the Mismatch condition to pattern with the Match condition (as in 6.3).

Thus, these results cannot be used to inform the original hypotheses. However, an important result was that there were no significant effects on measures of decision threshold or reaction times. Therefore, the selective influence of semantic content on perception was localised to low level visual perception. This supports the effect of motion words on low level visual processes independently of higher order decision or response processes. This finding is in line with other experiments that have shown effects on d' without effects on decision processes (6.5) or without effects on either decision processes or reaction times (6.3). As discussed previously, Signal Detection Theory treats measures of criterion (c and β) as independent of perceptual sensitivity (d'); however, it is possible that the influence of motion words operates pervasively on attention, decision threshold, response speed and perceptual sensitivity. By finding that d' can be influenced without any other changes in dependent measures, the effect of motion words on perception is decoupled from effects on higher order processes. This provides further evidence for a direct and automatic connection between semantic motion and the perception of motion.

8. Perception of visual motion on comprehension

As we have seen, the comprehension of motion words can influence low levels of visual perception. However, those influences are variable across small changes in task parameters, in particular, the strength of the linguistic context and the temporal relationship between lexical items and visual motion. The previous experiments have also shown that language affects different levels of task performance in different ways. This highlights the fact that whenever influences of comprehension upon perception are found, it is difficult to exclude other top-down mechanisms on perception that have themselves been affected by comprehension. The influences of comprehension on perception may be mediated, coming from changes in attention or working memory, rather than from the direct engagement of perceptual systems during semantic access. Language is instrumental in modulating attention and the content of working memory; these are two of the most ordinary consequences of comprehension. Stronger evidence for embodiment comes from influences of perception on comprehension. In this case, irrelevant perceptual stimuli influence linguistic tasks (such as lexical decision or sentence judgements) in a semantically modulated manner. Here, it is less likely that uncontrolled top-down mechanisms mediate the interaction between perception and comprehension. If motion perception, which activates particular perceptual processes, interacts with comprehension this is more likely to be because those perceptual processes have directly influenced semantic access. This is because higher-order cognitive processes such as attention and working memory are canonically held to modulate perceptual processes (e.g. Yantis, 2005) rather than be modulated by them. Therefore, in this chapter a number of experiments are presented which explore the influence of motion perception on the comprehension of words and sentences that refer to motion.

The first two experiments (8.1) were designed as an extension of Kaschak et al (2005), in which visual motion was perceived in blocks whilst sentences were judged for sensibility and grammaticality. To avoid problems that might occur when visual motion was blocked (in particular, the motion after effect) we presented perceptual and lexical stimuli on a trial by trial basis. In addition, we adapted the paradigm for use with both single words (lexical decision) and sentences (sensibility judgements). The final two experiments presented sentences (8.2) and words (8.3) visually rather than aurally. This was to counter uncontrolled linguistic variables that may have been covering effects for

experiments 8.1; this method of presentation also ensured participants perceived the visual motion. If single words and sentences referring to motion automatically engage the perceptual systems involved in motion processing, the reaction times and error rates for judging stimuli should differ (1) in comparison to the control condition in which linguistic stimuli do not refer to motion and (2) depending on whether the visual and semantic motion are congruent or incongruent.

8.1. *Trial by Trial perception of motion on single word and sentence comprehension*

Participants were presented with single words for lexical decision (8.1a) or sentences for sensibility judgements (8.1b). Visual motion began before item onset and ended shortly after offset, therefore covering the duration of the item without needing blocked presentation. The motion animations were RDKs, rather than black and white lines, with 60% coherent motion. These stimuli were used for two reasons, first, to further reduce the chance that the MAE would be produced by presenting salient but not overwhelming visual motion and second, to make the stimuli comparable to the motion used in experiments covered in Chapters 6 and 7. In line with the finding from Kaschak et al (2005) the hypotheses are as follows: when the semantic and visual motion are congruent, participants will be slower to judge the words or sentences (as compared to congruent and control conditions).

8.1.1. Participants

8.1a) 33 native English speakers took part in return for payment. Mean age was 25.7 years (standard deviation = 10.6 years), there were 19 females.

8.1b) 33 native English speakers took part in return for payment. Mean age was 24.6 years (standard deviation = 8.1 years), there were 22 females.

8.1.2. **Method**

8.1.2.1. *Design*

A 3 x 2 repeated measures design was used. The independent variables were Word/Sentence Category (Up, Down and Control) and Dot Motion (Upwards, Downwards). The dependent variables were the reaction time (ms) and error rates for lexical decision/sentence judgement. The experiment was divided into 2 parts, the first part comprised the first presentation of each item, and the second part comprised the second presentation of each item. Each item was presented with both upwards and downwards motion. Presentation of parts, blocks and trials was fully randomised.

8.1a) Each half contained 10 blocks, each block contained 21 lexical decision trials, making a total of 420 trials.

8.1b) Each half contained 10 blocks, each block contained 15 sentence judgement trials, making a total of 300 trials.

8.1.2.2. *Materials*

8.1a) Item Set Two Single Words was used. 103 nonsense words were selected from the set created for this purpose.

8.1b) Item Set Two Sentences was used. 60 nonsense sentences were selected from the set created for this purposes.

Two lists were created, one which paired each item with one direction of motion and one which paired each item with the other direction of motion. One list was presented in the first half of each experiments and the other list was presented in the second half.

8.1.2.3. *Graphic Display*

The motion display was created by displaying a sequence of 70 bitmap files as frames, which were displayed for 16 ms each. The sequence was created so that it could repeat

smoothly and give the impression of continuous motion; one complete cycle of the 70 frames took 1.12seconds. Moving dots which left the frame area along the y-axis were replaced. Each frame subtended $9.9 \times 9.9^\circ$ of visual angle, and contained 225 white dots on a black background. Each dot subtended 0.14° , all dots covered less than 0.05% of the total area. 135 of the dots (60%) were displaced in one direction by $9^\circ/s$, the remaining dots were displaced to random locations between each frame. A fixation cross presented centrally (0.42°) was presented throughout.

8.1.2.4. Apparatus

Experiments were presented using E-Prime 1.1 software (Schneider, Eschman & Zuccolotto; 2002). The experiment was run on PC computers (Intel Pentium II), using 15" flat screen LCD monitors. Chin rests were placed at a distance of 57cm from the computer screen. Stereo closed headphones were used to present the audio files binaurally.

8.1.2.5. Procedure

Participants were seated in a sound attenuated booth, and instructed to press the 'M' key if the word/sentence they heard was a real meaningful word/sensible sentence, and the 'N' key if the it was a non-word/nonsense sentence. Participants were seated in front of the computer screen and rested their chins on a chin rest which minimized head movements and placed their eyes level with the centre of the computer screen.

On each trial a blue fixation cross presented for 1000 ms to signal the beginning of each trial, the motion animation was displayed for 1100 ms before the word (8.2a)/sentence (8.2b) was presented and continued playing for 1100/2200ms during the presentation of the word/sentence. A red fixation cross was then presented for 1000ms. There was a 3000/5000ms time out following the offset of the word/sentence.

8.1.2.6. Analysis

To reduce the influence of sentence/word length on response times, the length of the word or sentence was subtracted from the reaction time measured from sentence onset; this produced reaction times that were relative to sentence offset.

8.1a) Two subjects were removed from the analysis as their accuracy rates were below 60%, all other subjects had accuracy rates above 90%. Reaction times below 0 ms (before word offset), above 2 standard deviations per subject or above 1500ms were excluded. This removed 5.9% of the data. Errors were analysed separately.

8.1b) Four subjects were removed from the analysis as their accuracy rates were below 70%, all other subjects had accuracy rates above this cut-off. Reaction times below -200ms (200ms before sentence offset), above 2 standard deviations per subject or above 1000ms were removed. This removed 3.34% of the data. Errors were analysed separately.

8.1.3. Results

8.1.3.1. Reaction Times

8.1a) There was a significant main effect of Word Category by Subjects ($F(2,60) = 8.877, p-\eta^2 = 0.228; F_2 < 1.9$) and the main effect of Motion Direction also approached significance by Subjects ($F(1,30) = 3.536, p = 0.07, p-\eta^2 = 0.105; F_2 < 1$). The mean reaction time for Control words was 409ms, for the Mismatch condition 391ms and for the Match condition 374ms. Planned comparisons showed that the Control condition had significantly longer reaction times than the Mismatch ($F(1,30) = 6.866, p-\eta^2 = 0.186; F_2 < 1$) and Match conditions ($F(1,30) = 18.779, p-\eta^2 = 0.385; F_2(1,58) = 3.362, p = 0.072, p-\eta^2 = 0.055$). There was also a significant interaction between Word Category and Motion Direction ($F(2,60) = 11.374, p-\eta^2 = 0.275; F_2(2,87) = 2.315, p = 0.105, p-\eta^2 = 0.051$). Upwards motion produced reaction times that were around 11ms faster than Downwards motion in the Control and Match conditions; however, in the Mismatch condition Downwards motion was 27ms faster than Upwards motion. The interaction between Word Category and Motion Direction was significant when the Mismatch condition was compared to the Control ($F(1,30) = 9.247, p-\eta^2 = 0.236; F_2(1,58) = 2.580, p = 0.114, p-\eta^2 = 0.043$) and Match conditions ($F(1,30) = 19.646, p-\eta^2 = 0.396; F_2(1,58) = 3.144, p = 0.081, p-\eta^2 = 0.051$)¹⁵. See Table 8-1 and Figure 8-1a.

¹⁵ When a 2x2 repeated measures ANOVA was used to compare Word Category (Down and Up) by

8.1b) There was a significant main effect of Sentence Category ($F(2,56) = 39.647$, $p-\eta^2 = 0.596$; $F(2,84) = 7.345$, $p-\eta^2 = 0.149$). Planned comparisons showed that the Control condition had reaction times 66ms faster than the Match condition ($F(1,28) = 54.520$, $p-\eta^2 = 0.661$; $F(1,56) = 9.987$, $p-\eta^2 = 0.151$) and 61ms faster than the Mismatch condition ($F(1,28) = 59.584$, $p-\eta^2 = 0.680$; $F(1,56) = 7.240$, $p-\eta^2 = 0.114$). There was also a significant interaction between Sentence Category and Motion Direction ($F(2,56) = 4.2$, $p-\eta^2 = 0.130$; $F(2,84) < 1$). In the Control condition, Downwards motion was 22ms slower than Upwards motion; this in contrast to the Match ($F(1,28) = 5.192$, $p-\eta^2 = 0.156$; $F(1,56) < 1$) and Mismatch condition ($F(1,28) = 6.118$, $p-\eta^2 = 0.179$; $F(1,56) < 1.7$) for which Downwards motion was 15ms faster than Upwards motion. See Table 8-1 and Figure 8-1b.

8.1.3.2. Errors

8.1a) A Wilcoxon Signed Ranks test showed there were more errors for the Mismatch condition when words were paired with Upwards motion ($z = -2.109$, $p < 0.05$); this difference was reliable by Subjects only.

8.1b) There was a significant difference in error rates between the Control, Match and Mismatch conditions (Friedman test by Subjects $X^2 = 14.150$, $df = 2$; Kruskal Wallis by Items $X^2 = 5.917$, $df = 2$, $p = 0.052$). There were significantly more errors in the Match (mean = 8.7; $z_1 = -2.778$, $z_2 = -5.484$) and Mismatch conditions (mean = 8.0; $z_1 = -2.431$, $z_2 = -5.030$) in comparison to the Control condition (mean = 6.4).

8.1.4. Discussion

Control words produced significantly longer reaction times as compared to directional words. In addition, Down words were found to have longer reaction times than Up words. This set of items was matched for length in phonemes and phonological neighbourhood size, in an attempt to reduce the variability in lexical decision times. However, there are still differences between Word Categories that produce different reaction times regardless of visual motion. The fact that most comparisons were

Motion Direction (Downwards and Upwards), a significant main effect of Word Category ($F(1,30) = 19.646$, $p < 0.0001$, $p-\eta^2 = 0.396$) was found, with Up words faster than Down words by a mean of

significant by Subjects but not by Items supports the inference that there were uncontrolled factors on an item-by-item basis that increased the overall variance. Differences between Word Categories (Control, Down and Up) could be due to uncontrolled variables that affect spoken words recognition.

Table 8-1: Mean Response Time for Sentence Judgement in Experiment 8.1b and 8.2b
(Standard deviation in brackets)

Experiment	Condition					
	Control		Match		Mismatch	
	Downwards	Upwards	Downwards	Upwards	Downwards	Upwards
Exp 8.1a						
Single Words	417.80	399.29	394.59	353.54	376.91	404.08
Trial dot motion	(94.09)	(113.41)	(86.81)	(103.67)	(119.34)	(103.29)
Mean	408.54 (103.76)*		374.07 (97.06)		390.50 (111.53)	
Exp 8.1b						
Sentences	252.76	230.96	303.78	312.48	293.18	313.26
Trial dot motion	(74.45)	(60.89)	(81.64)	(82.16)	(78.20)	(77.00)
Mean	241.86 (68.30)*		308.13 (81.30)		303.22 (77.59)	

* Significantly different to the other two conditions, $p < 0.05$

For example, auditory lexical decision is sensitive to phonotactic probability (Vitevich et al, 1999), with higher probabilities (i.e. more frequent phonemes and sequences of phonemes) producing faster reaction times. Another factor that is known to influence auditory lexical decision times is the unique recognition point of each word. RTs were calculated relative to word offset, however, there was likely considerable variability in the recognition points for each word that would increase variability across items (for a review see Goldinger, 1996). This increases the noise in the data and may mask any congruency effects.

As the central task was to judge words presented aurally, it is possible that the visual motion was not consistently viewed (at least for the lexical decision experiment, where motion patterns were of a shorter duration than for sentence judgements, see below). When single words were presented, there were some effects of motion direction that interacted with congruency, however these were driven by differences between Down and Up categories rather than by strong effects of visual motion. There are a number of ways in which this problem could be counteracted. First, eye-movements could be monitored to ensure fixation. Second, a dual task method could be employed so that

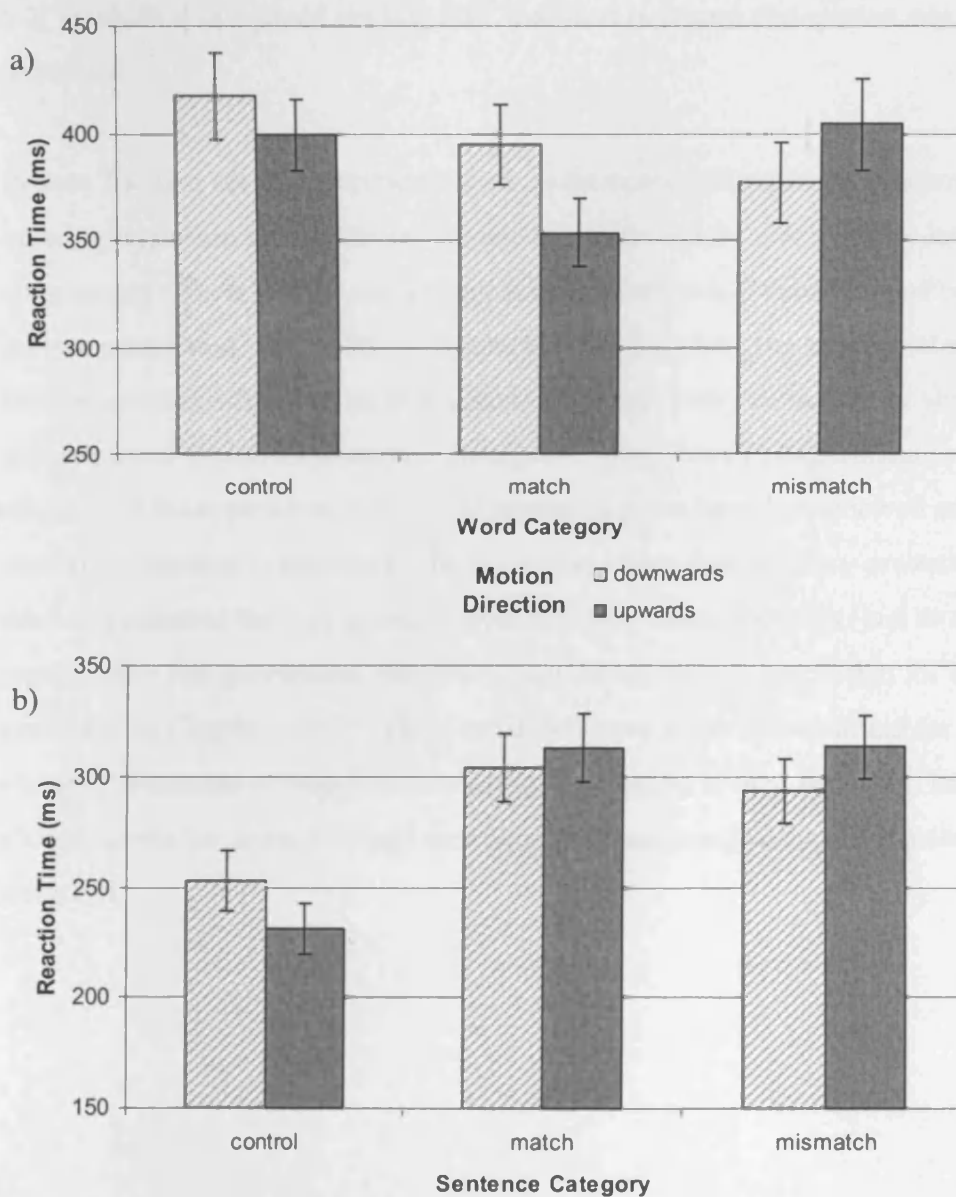
34ms. Therefore when collapsed into Match versus Mismatch conditions an interaction emerges.

attendance to the visual display was necessary (e.g. Zwaan & Taylor, 2006, Experiment 3). Finally, visual presentation of the linguistic stimuli could be used so that they are overlaid on RDK motion. In this case, perception of the motion is passive but guaranteed during the lexical decision task; this was the approach taken in Experiment 8.3.

Reaction times for sentence judgements also showed differences between the Control and directional items, such that Control sentences were judged more quickly than directional sentences. In addition, error rates were higher for directional sentences, so a speed-accuracy trade-off was not responsible for the reaction time differences (which would have produced more errors for the Control condition as compared to directional sentences). Therefore directional sentences were responded to more slowly and with more errors, regardless of congruence. In addition, the mean response times for each direction of motion differed between Control and directional sentences. In the Control condition, upwards motion produced faster responses than downwards motion; the opposite was true for directional sentences with faster responses when downwards motion was perceived. This indicates that the motion patterns were being perceived and it reinforces the conclusion that directional sentences behaved differently to the control sentences. However, it is not possible to interpret the interaction itself in a sensible or motivated manner, therefore it will not be discussed any further. Although there was no effect of congruence, it is possible that this result is based on the influence of visual motion on the comprehension of motion sentences. We could make a tentative inference of general interference caused by the perception of visual motion on the comprehension of directional sentences. This is against the original findings of Kaschak et al (2005) as the interference is not modulated by congruency. Recall that in that experiment, congruence led to longer reaction times. Our results instead support more difficult comprehension of sentences referring to vertical motion whenever the visual system perceives vertical motion. Kaschak et al (2005) presented a 30 second motion animation whilst a number of aurally presented sentences were judged; therefore the motion patterns lasted much longer. There are two possibilities: First, the longer motion duration in Kaschak et al (ibid) produced qualitatively different motion processing than the shorter animations we presented. Exposure to strong motion for durations of around 30 seconds produces the motion after effect, demonstrating that directionally selective neurones have become fatigued (Levinson & Sekuler, 1976). Thus, congruency effects may only appear when motion processing is heavily taxed.

Directionally selective neurones may be able to process visual and semantic motion reasonably well (producing general interference for any semantic motion, congruent or incongruent) unless one task demands substantially more resources for directionally selective processes (when congruence effects emerge). This would also be in line with the results from our motion detection experiments, in which directionally selective processes are heavily taxed through the presentation of ambiguous (i.e. difficult) motion stimuli.

Figure 8-1: Reaction Times to judge stimuli for 8.1(a) Lexical Decision and 8.1(b) Sentence sensibility



Under this explanation, we should find congruency effects with longer durations of motion (similar to Kaschak et al, *ibid*) but not for shorter durations (in which we should replicate our current results)¹⁶. Second, a more trivial explanation is that in Kaschak et al (*ibid*) the longer motion duration could mean that participants were more likely to consistently perceive motion, producing congruency effects. In our experiment, the shorter motion duration (operating on a trial by trial basis) may have led to less consistent viewing of the motion patterns and a general interference from motion (since the visual motion signal was weaker due to inattention). This explanation means that if we monitored eye-movements or ensured participants were viewing the patterns, we would also find congruency effects. This second explanation was tested in Experiment 8.2, in which a self-paced reading task was used to ensure that motion was consistently perceived.

In sum, the first set of experiments show some consistent effects of motion perception on comprehension for sentences, but not for single words. For auditory lexical decision, congruency effects may be masked by uncontrolled lexical variables and/or inconsistent motion perception. For auditory sentence comprehension, the general interference for motion sentences (in contrast to Kaschak et al, 2005) may be due to the shorter motion animations or inconsistent motion perception. The next two experiments sought to counter the main problem, that visual motion may not have been viewed consistently enough to produce congruency effects. Visual rather than auditory presentation was used to guarantee the perception of motion during comprehension (just as auditory presentation had guaranteed comprehension during motion perception for experiments presented in Chapter 6 & 7). There are differences in the effects found for auditory and visual presentation of linguistic items (e.g. Goldinger, 1996). However, both forms should access the same semantic representation, and congruency effects should be observed.

¹⁶ See Appendix 3 for an attempted replication and extension of Kaschak et al (2005)

8.2. *Self-paced reading with motion perception*

Participants were presented with sentences (phrase by phrase) in a self-paced reading task. Behind the box in which phrases were displayed, upwards or downwards motion was presented in the form of black and white lines. Self-paced reading was used for two reasons. First, it is hoped that visual presentation will remove some of the variation in response times present with auditory comprehension. Second, as the participants have to read the sentences they also have to perceive the visual motion present in the periphery. It has been previously demonstrated that illusory rotational motion influences the reading times for sentences referring to rotation (Zwaan & Taylor, 2006). In line with their findings, one hypothesis is that when the visual and semantic motion are congruent, reading times will be faster in comparison to incongruent visual motion and the control condition (in which the sentences do not refer to a consistent direction of motion). They found that only reading times for the verb (which implied rotation) were facilitated, therefore the addition to the main hypothesis is that faster reading times will only be present for the verb, rather than the subject or object phrase. However, since Zwaan & Taylor (2006) used sentences that described manual rotation (so the visual motion was implied rather than directly described), results may be different for sentences that explicitly describe visual motion. Thus, a second hypothesis is that we may replicate in reading times what Kaschak et al (2005) found for auditory sentence judgement: congruent conditions increase reaction times as the semantic and visual motion compete for the same resources. Finally, an alternative to these congruency effects is the hypothesis that we will replicate the general interference from motion found for experiment 8.1b; this would support the inference that trial-by-trial motion perception has a different effect on comprehension as compared to blocked motion perception (see discussion of experiment 8.1b above).

8.2.1. **Participants**

38 native English speakers (20 female) took part in return for payment. Mean age was 22.45 years (standard deviation = 7.95 years).

8.2.2. Method

8.2.2.1. Design

A 3 x 3 x 2 repeated measures design was used. The independent variables were Sentence Category (Up, Down and Control), Sentence Phrase (First, Second or Third) and Motion Direction (Upwards, Downwards). The dependent variable was the reaction time to read each Phrase (milliseconds). The experiment was divided into 2 parts, each containing 9 blocks of 15 trials, making a total of 270 sentence judgement trials. The first part of the experiment comprised the first presentation of each item; the second part comprised the second presentation of each item. Each sentence was presented whilst the participant viewed upwards or downwards motion. Presentation of parts, blocks and trials was fully randomised.

8.2.2.2. Materials

Item Set Two Sentences was used. 45 Filler items (followed by yes/no comprehension questions of the form “Did the X verb the Y?”) were selected from the set created for that purpose.

For the purposes of presentation, all sentences were broken down into three phrases: 1. subject phrase, 2. verb phrase, and 3. prepositional/ object phrase (e.g. 1. the man 2. walked 3. to the river). Two lists were created for each half of the experiment, pairing each sentence with each direction of motion. Each filler item was followed by a comprehension question in one list only, resulting in comprehension questions on 17% of trials.

8.2.2.3. Graphic Display

Motion stimuli were created by displaying a sequence of 8 bitmap files as frames, each displayed for 100ms. The sequence was created so that it could repeat smoothly and give the impression of continuous motion. Each file filled the screen; presenting 24 thin black lines (approx 0.115°) separated by thicker white lines (approx 1°). The lines

moved at around 1.152 °/s. The sentences were presented on a white frame overlaying the lines, centred at fixation (approximately 18 x 14° visual angle). The first and second phrases were usually located on the same line directly above the third phrase. Thus, behind the central white frame and filling the rest of the screen were horizontal lines that moved either upward or downwards

8.2.2.4. Apparatus

The experiment was carried out using E-Prime 1.0 software on Intel Pentium III computers. The display was presented on a 15" LCD monitor with a refresh rate of 75Hz. Participants responded using a standard 'qwerty' keyboard and sat approximately 50cm from the screen during presentations

8.2.2.5. Procedure

Participants were instructed to press the spacebar to make each part of the sentence appear (first, second and third phrase) and to read them at their own pace. On some trials, they would be required to answer questions about the sentence they had just read, pressing Y for 'yes' and N for 'no'. During each trial, a fixation cross was presented for 500ms. A sentence was then presented with all letters and punctuation (e.g. apostrophes) replaced by the capital letter X. On the first press the first phrase appeared whilst all other letters/phrases were still replaced by Xs. Upon each press the previous phrase was again replaced with Xs and the next phrase appeared. When the spacebar was pressed after presentation of the third (and final) phrase, the screen was replaced by a fixation cross for 500ms. On filler trials a comprehension question then appeared, with 6000ms time out for responses. After each block participants were given an optional break.

8.2.2.6. Analysis

Three participants were eliminated from the study for having accuracy scores lower than 85%. The reaction times were measured from the time taken to press the space bar after the appearance of each phrase. Invalid trials were removed (when a double press had occurred and reaction times were therefore recorded as negative). Reaction times below

100ms and above 1000ms were removed as well as times over 2 standard deviations from each participant's mean response time. This removed 5.39% of the data.

Remaining response times were analysed using a 3 (phrase) x 3 (sentence category) x 2 (motion direction) ANOVA by-subjects and by-items. Planned simple comparisons were then performed for Sentence Category x Motion Direction within each Phrase.

8.2.3. Results

There was a significant main effect of Phrase ($F(1,68) = 24.782$, $p-\eta^2 = 0.422$; $F(2,168) = 641.890$, $p-\eta^2 = 0.884$). Reading times for Phrase 1 (mean = 409ms) and Phrase 2 (414ms) were significantly faster than Phrase 3 (525ms). There was also a significant interaction between Phrase and Motion Direction ($F(2,68) = 3.507$, $p-\eta^2 = 0.093$; $F_2 < 1.3$). For Phrase 1 and 2, upwards motion produced faster reading times than downwards motion; this pattern was reversed for Phrase 3. There was also a significant 3-way interaction between Phrase, Sentence Category and Motion ($F(4,136) = 7.645$, $p-\eta^2 = 0.184$; $F(4,168) = 3.531$, $p-\eta^2 = 0.078$). Phrase 1 and 2 behaved similarly, with Control and Match conditions having faster times for Upwards motion, this was reversed in the Mismatch condition with reading times faster for Downwards motion. These differences produced a significant interaction between Sentence Category and Motion Direction for Phrase 2 ($F(1,68) = 3.830$, $p-\eta^2 = 0.101$; $F_2 < 1$) but not Phrase 1 (all $F_s < 1$). Phrase 3 had faster reading times in the Control and Match conditions for Downwards motion. In contrast, the Mismatch condition had faster reading times for Upwards motion. These differences produced a significant interaction between Sentence Category and Motion Direction for Phrase 3 ($F(1,68) = 4.558$, $p-\eta^2 = 0.118$; $F_2 < 1$).

Table 8-2: Mean Reading Time (ms) for each Phrase by Condition
(Standard deviation in brackets)

Phrase	Condition					
	Control		Match		Mismatch	
	Downwards	Upwards	Downwards	Upwards	Downwards	Upwards
Phrase 1	410.04 (122.31)	399.41 (118.65)	416.44 (131.06)	411.09 (134.01)	407.35 (120.91)	411.05 (127.11)
<i>Mean</i>	404.73 (119.74) †		413.77 (131.61)		409.20 (123.16)	
Phrase 2	413.09 (136.48)	402.22 (126.71)	426.40 (153.08)	411.04 (138.09)	415.51 (133.98)	419.11 (134.77)
<i>Mean</i>	407.65 (130.85) *		418.72 (144.92)		417.31 (133.41)	
Phrase 3	509.48 (244.48)	523.51 (245.23)	521.70 (253.24)	542.24 (270.40)	535.10 (254.61)	518.36 (236.72)
<i>Mean</i>	516.50 (243.18)		531.97 (260.16)		526.73 (244.18)	

†Significantly different to the Match condition.

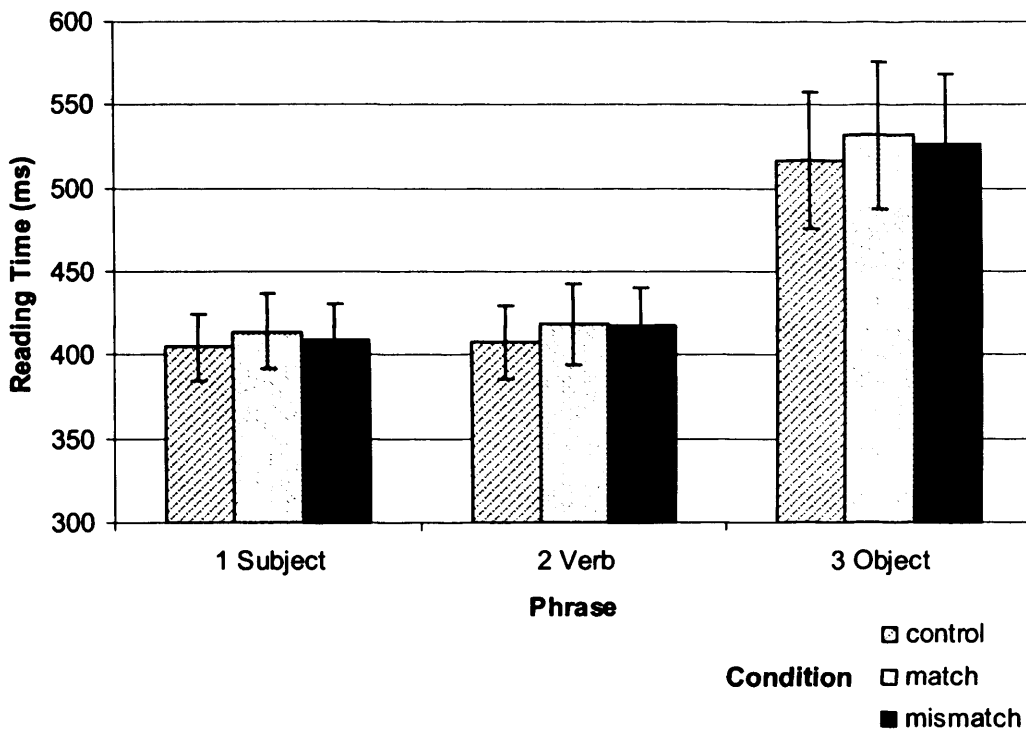
*Significantly different to both other conditions.

The main effect of Sentence Category was also significant ($F(2,68) = 4.065$, $p-\eta^2 = 0.107$; $F_2 < 1$). Mean reading times were 443ms for the Control condition, 451ms for the Mismatch conditions and 455ms for the Match condition. Phrase 1 showed a marginal main effect of Sentence Category ($F(2,68) = 2.875$, $p = 0.063$, $p-\eta^2 = 0.078$; $F_2 < 1$), driven by the difference between Control (mean = 405ms) and Match conditions (414ms). Phrase 2 showed a significant main effect of Sentence Category ($F(2,68) = 5.224$, $p-\eta^2 = 0.133$; $F_2 < 1$), with the Control condition (mean = 408ms) significantly faster than the Match (419ms) and Mismatch condition (417ms). Phrase 3 did not show a significant main effect of Sentence Category ($F_1 < 1.8$; $F_2 < 1$)¹⁷. See Table 8-3 for reading times in all conditions and Figure 8-3 for reading times for each Phrase and Sentence Category, collapsed across Motion Direction.

¹⁷ When Control, Down and Up conditions were compared by Subjects, a main effect of Sentence Category ($F(2,68) = 4.514$, $p-\eta^2 = 0.117$) was driven by the Down and Up conditions producing longer reading times as compared to the Control condition ($F(1,34) = 10.380$, $p-\eta^2 = 0.234$; $F(1,34) = 5.229$, $p-\eta^2 = 0.133$ respectively). They were not significantly different to each other ($F < 1$).

Figure 8-2: Mean reading times by Phrase and Condition*

*Error bars are one standard error of the mean



8.2.4. Discussion

There was a significant main effect of Sentence Category, with Control sentences producing faster reaction times as compared to the Match condition in Phrase 1, and the Match and Mismatch conditions for Phrase 2. It is possible that the effect at Phrase 2, the verb, is due to the directional content of the words, however it is more likely that the effect present in Phrase 1 carried over and was enhanced by Phrase 2. In other words, general differences between the Control and directional sentences appeared early (following the subject phrase) and were still present when the verb was comprehended. The changes in reaction times indicate that the Control sentences were easier to comprehend than the directional sentences. This result fits with the data from auditory sentence comprehension (8.1b) that showed faster response times to Control sentences as compared to directional sentences. Therefore, at least some of the difference in reaction times between control and directional sentences can be attributed to general differences in sentence composition rather than directional content. No differences were present between Sentence Categories at Phrase 3, which also showed reading

times substantially longer than the other two phrases. This result is probably due to the active comprehension task employed to ensure participants read the sentences properly. On filler trials, participants had to respond 'yes' or 'no' to a simple question about the content of the sentence. For example, after reading the sentence "The man drove to work" they might be asked the question "Did the man walk to work?" or "Did the man drive home?". The longer reading times at Phrase 3 suggest that participants read the sentence quickly but slowed down at Phrase 3 to either recall the sentence or maintain it in working memory should they be asked a question. One possibility that might rescue the interpretation that motion (rather than general comprehension difficulty) affected directional sentences in experiment 8.1b, is that visual presentation for reading forced the participants to attend to the different elements of the sentence, decreasing the focus on the motion event (the verb). This returns us to the point about single words presenting type information (more schematic, but less constrained) in comparison to sentences which present token information (more concrete, but more constrained). When presented as phrases the sentences will be comprehended in chunks that have to be integrated and participants would have become used to the separation of each sentence into 3 phrases. This may remove the salience of the motion event and increase the salience of the subject and object. Note that if the sentences had been presented word-by-word (as in Zwaan & Taylor, 2006) this might have reduced this problem as all words would have to be given the same attention.

The longer reading times at Phrase 3 also explain why no differences in Sentence Categories were found here, as this strategic slowing down was present for all sentences and would swamp any small differences that might have been present on the initial reading of the final phrase. The whole sentence was initially presented with all the letters replaced by Xs; therefore participants were obviously aware of where the sentence ended. It is perhaps not surprising that this influenced reading times in combination with comprehension questions. This interpretation also implies that the comprehension questions were not an effective manipulation, resulting in a shallow reading of the sentences. The surface form of the target sentence was only needed to correctly respond to the comprehension questions as they typically differed from the target in only one word. Three changes to this method may improve results. First, sentences could be lengthened and presented one word at a time, rather than in phrases. This would increase the sensitivity of the paradigm and produce a longer exposure to visual motion. For example, each sentence could be preceded by a context sentence,

e.g. “Once he had finished his breakfast, the man drove to work” (e.g. Zwaan & Taylor, 2006). It would also make it less clear where the sentence ended. Second, a grammaticality or sensibility judgement could be used instead of a comprehension question. With a careful selection of filler items, this would make the task harder and require a deeper reading of each sentence. Third, motion direction could be blocked, with each block starting with one or two filler sentences. In the present design, sentence presentation and motion direction were randomised; therefore the length of exposure to visual motion was limited. This is especially true here as programming constraints required each frame of the motion animation to be presented for 100ms, making the speed of vertical motion slower than in all other experiments. This has the unfortunate result that directional motion may not have been salient enough (given that it was also presented in the periphery) to influence reading times.

There were also interactions between Sentence Category and Motion Direction, with Phrase 3 and the Mismatch condition behaving in opposition to the Control and Match conditions for Phrase 1 and 2. The behaviour of the Mismatch condition cannot be attributed to general differences between the Down and Up sentences that produced interactions when analysed as Match vs. Mismatch (see Footnote 3), so it is unclear how to interpret them. The differences for Phrase 3 may well be due to the longer reading times evidenced here distorting effects seen in Phrase 1 and 2. However, what the interactions do demonstrate is that the direction of motion did influence reading times. For Phrase 1 and 2, upwards motion generally resulted in faster reading times as compared to downwards motion. For Phrase 3 this was reversed. These results do not agree with previous data that show upwards motion as harder to perceive/more interfering than downwards motion. However, they do reinforce the conclusion that visual motion was perhaps not salient enough to produce congruency effects in reading times. Based on a presentation speed of one frame every 100ms, mean reading times for Phrase 1 and 2 would allow 4 frames to pass, whilst in Phrase 3 this is increased to 5 or 6 frames. Only in Phrase 3, when reading times were substantially longer, did upwards motion show the longer response times that we have seen more reliably elsewhere.

In sum, the self-paced reading task was beset by methodological problems. Therefore no conclusions can be drawn about the influence of motion perception on reading times for directional sentences.

8.3. *Visual Lexical Decision with Salient and Ambiguous motion*

Participants are presented with single words in a visual lexical decision task. RDK motion patterns are presented behind the word, with motion that is either supra-threshold and salient or at threshold and therefore ambiguous. The manipulation of motion coherence is motivated by a recent finding in visual neuroscience. Tsushima et al (2006) gave participants a visual target identification task (reporting two digits in a rapid serial visual stream of letters) whilst presenting irrelevant motion in the periphery. They found that when the coherence of the irrelevant motion was close to threshold, performance on the target identification task dropped. A follow up fMRI experiment showed that when the irrelevant motion was salient, MT+ showed reduced activity and the Dorso Lateral Prefrontal Cortex (DLPFC) showed increased activity. In contrast, when the irrelevant motion was ambiguous, MT+ showed increased activity and the DLPFC showed reduced activity. The authors concluded that when the motion is salient, the DLPFC suppresses activity at MT+, preventing the salient motion signal from interfering with task performance (by taking resources). In contrast, an ambiguous signal does not activate suppression and is processed; thus it takes resources and reduces performance in a visual target identification task. By using this manipulation (ambiguous vs. salient motion) we can explore how a bottom-up manipulation of the visual system influences semantic processing in a language task. If the manipulation is effective, we have strong evidence of integration between semantic and visual motion.

We hypothesised that the same pattern should be true for a visual lexical decision task, with ambiguous motion interfering with task execution whilst salient motion does not. However, if semantic motion information is integrated with the visual motion signal, the interference from the ambiguous motion should be semantically modulated. We should see a difference between control words and motion words and we should see some effect of congruency. The ambiguous motion should interfere with lexical decision (producing longer reaction times) but it is not clear how congruency will interact with this interference effect. We predict one of two patterns: the interference will only be evident in the incongruent condition (as the ambiguous motion conflicts with semantic motion, making lexical decision harder); alternatively, the interference will be present for control and incongruent conditions, but reduced in the congruent condition (as the ambiguous motion benefits semantic processing).

8.3.1. **Participants**

8.3a) 28 native English speakers, 16 female, were recruited from the subject pool at University College London, they were paid £6 for their time. The average age was 23.8 years (standard deviation 5.9), 4 were left handed.

8.3b) 28 native English speakers, 11 female, were recruited from the subject pool at University College London, they were paid £3 for their time. The average age was 25.64 years (standard deviation 6.31), 4 were left handed.

8.3.2. **Method**

8.3.2.1. *Design*

A 3x2 repeated measures design was used. The independent variables were Word Category (Control, Up, Down) and Motion Direction (Upwards, Downwards). The dependent variables were the reaction time (ms) and accuracy in the lexical decision task. Participants completed 12 blocks of 34 trials. There were 6 blocks in each half, with each word presented once in each half. Within each half and block, items presentation was randomised. For 8.3a the experimental sessions were made up of two parts: the thresholding procedure and the lexical decision task. For 8.3b participants only completed the lexical decision task. Participants were tested individually in a small testing room. A side-light facing the wall behind the participant was used to provide lighting, this ensured that the motion stimuli were easy to discriminate but participants' eyes did not become fatigued.

8.3.2.2. *Materials*

Item Set Two Single Words was used. 96 non-words were selected from the set created for that purpose. For the experimental items two lists were created. The lists were identical except that in List1 the first presentation of each word was with one motion direction, whereas in List 2 the first presentation was with the other direction of motion.

The use of these lists was balanced across participants. Each list had two halves (A & B), with each item presented once in each half. The two halves were identical except that in A each item was paired with one motion direction and in B with the other direction; therefore over the course of the experiment each item was presented twice, once with each motion direction. The program for the experiment was designed so that it randomly selected items for presentation in each block, but presented A in the one half-run of the experiment and B in the other half-run (balanced across participants). In this way the repetition of items was kept as far apart as possible whilst maintaining randomization.

8.3.2.3. Graphic Display

Stimuli were identical to Experiment 6.2 except that they were 200ms in duration (15 consecutive frames; speed = 21 °/s).

8.3.2.4. Apparatus

The apparatus was identical to Experiment 6.3.

8.3.2.5. Procedure

For experiment 8.3a participants underwent a thresholding procedure for each direction of motion (upwards and downwards), the structure of the thresholding task was identical to Experiment 6.5. The order of thresholding for each motion direction was balanced across participants. Individual threshold values for each motion direction (upwards and downwards) were used as the coherent motion signal for the RDKs. For 8.3b all RDKs were set at 30% coherence (5 times higher than average threshold from Experiment 1, see Section 8.3.3.1). For the lexical decision task, participants completed 12 practice trials (half words and half nonwords), auditory feedback was given on mistakes in both the practise and experimental sessions to encourage concentration. Participants then completed the experimental blocks. On each trial, a central fixation square (width = 0.2°) was presented for 500ms. This was replaced by the item for lexical decision (Arial, white, 14pt) and an RDK behind the item for 200msec (SOA of 0ms). The word then remained on screen until a response or time out to 1500ms. Participants were

instructed to ignore the motion patterns and complete the lexical decision task; throughout all tasks they placed their chin on a rest 57cm from the screen.

8.3.2.6. *Analysis*

Subjects whose total errors were 2 standard deviations above the group mean were excluded; this removed 3 participants from experiment 8.3a and 2 participants from 8.3b. For the reaction time analysis only correct responses were included, errors on experimental trials were analysed separately. To remove outliers from the reaction time data, times under 100ms, over 1000ms or 2 standard deviations above the mean for each subject were excluded; this removed 7.7% of the data for Experiment 8.3a and 7% of the data for 8.3b. In a follow up analysis to compare ambiguous (8.3a) and salient (8.3b) motion, reaction times were normalised as z-scores relative to the control condition in each experiment. The mean and standard deviations were taken from the control condition for a given motion direction (upwards and downwards) and applied to the mean reaction times (by Subject or by Items) for Control, Up and Down words for the same motion direction. This was to remove any influence of a particular motion direction on overall reaction times.

8.3.3. Results

8.3.3.1. Threshold Values from Experiment 8.3a

The mean threshold value was 6.9% (SD = 2.5%) for upwards motion and 6.3% (SD = 2.7%) for downwards motion, expressed as the percentage of coherently moving dots in the RDK. The thresholds for upward and downward motion were similar ($t(27) = 1.238, p > 0.2$ two-tailed).

Table 8-3: Mean Reaction Times by Condition and Experiment for 8.3a and 8.3b
(Standard Deviation in brackets)

Experiment	Condition					
	Control		Match		Mismatch	
	Down	Up	Down	Up	Down	Up
Experiment 8.3a: Threshold	555.23 (53.32)	554.26 (57.94)	562.80 (65.22)	549.34 (66.72)	572.83 (61.82)	574.43 (68.64)
<i>Mean</i>	554.75 (55.80)		556.07 (66.22)		573.63 (64.36)*	
Experiment 8.3b: Supra-threshold	576.92 (56.89)	566.63 (56.76)	564.86 (51.65)	582.56† (56.64)	581.73† (57.06)	572.27 (61.93)
<i>Mean</i>	571.77 (56.50)		573.71 (54.41)		577.00 (59.15)	

*Significantly different to other conditions, $p < 0.05$

†Up words significantly longer than down words, $p < 0.05$

8.3.3.2. Reaction Times

8.3a) There was a significant main effect of Word Category ($F(1,24) = 3.815, p - \eta^2 = 0.137$; $F(2,87) = 3.161, p - \eta^2 = 0.068$). Planned comparisons showed that the Mismatch condition had significantly longer reaction times as compared to the Control ($F(1,24) = 6.989, p - \eta^2 = 0.226$; $F(2,87) = 7.929, p - \eta^2 = 0.120$) and Match ($F(1,24) = 4.964, p - \eta^2 = 0.171$; $F(2,87) = 3.589, p = 0.063, p - \eta^2 = 0.058$) conditions. The mean reaction time for Mismatch conditions was 574ms, compared to 555ms for Control and 556ms for Match conditions.

8.3b) There was a significant interaction between Word Category and Motion Direction ($F(1,25) = 3.398, p - \eta^2 = 0.120$; $F(2,87) = 2.631, p = 0.078, p - \eta^2 = 0.057$). Planned comparisons showed that the interaction was significant for the Match condition as compared to the Control ($F(1,25) = 6.015, p - \eta^2 = 0.194$; $F(2,87) = 5.169, p - \eta^2 =$

0.082) and Mismatch conditions ($F(1,25) = 4.271$, $p-\eta^2 = 0.146$; $F(1,58) = 3.103$, $p = 0.083$, $p-\eta^2 = 0.051$). In the Match condition Upwards motion produced reaction times 18ms slower than downwards motion, this was reversed for the Control and Mismatch conditions, in which downwards motion was on average 10ms slower than Upwards motion.¹⁸

Table 8-3 presents the reaction times for each experiment in each condition.

8.3.3.3. Normalised Reaction Times

The interaction between Word Category and Experiment was marginal by Subjects ($F(2,98) = 3.046$, $p = 0.052$, $p-\eta^2 = 0.059$; $F_2 < 1$). Mismatch RTs were significantly longer in Experiment 8.3a but not Experiment 8.3b (see Figure 8-4). Planned comparisons showed a significant interaction between Word Category and Experiment for the Mismatch as compared to Match ($F(1,49) = 5.082$, $p-\eta^2 = 0.094$; $F_2 < 1$) and Control ($F(1,49) = 7.247$, $p-\eta^2 = 0.130$; $F_2 < 1$) conditions, with substantially longer reaction times for the Mismatch condition in Experiment 8.3a only. These comparisons also showed a significant main effect of Word Category, with longer RTs in the Mismatch as compared to Match ($F(1,49) = 5.112$, $p-\eta^2 = 0.094$; $F_2 < 1$) and Control conditions ($F(1,49) = 7.447$, $p-\eta^2 = 0.132$; $F_2 < 1$). Reaction times in the Mismatch condition were on average 0.179 SDs above those for the Match (mean = 0.067) and Control (0.0) conditions.

8.3.3.4. Errors

Separate Friedman tests compared the number of errors in the Control, Match and Mismatch conditions for each Experiment. There was no overall difference in errors for Experiment 1 ($X^2 < 2$, $df = 2$, $p > 0.4$) but the comparison approached significance for Experiment 2 ($X^2 = 5.247$, $df = 2$, $p = 0.073$). Paired sample comparisons using the Wilcoxon Signed Ranks test showed that in Experiment 1 there were significantly more errors in the Match than the Mismatch condition ($z = -2.245$, $p < 0.05$). In Experiment 2 there were significantly more errors in both the Match ($z = -2.030$, $p < 0.05$) and

¹⁸ A simple 2x2 comparison between Word Category (Down, Up) and Motion Direction (Downwards, Upwards) showed a significant main effect of Word Category ($F(1,25) = 4.271$, $p-\eta^2 = 0.146$; $F(1,58) = 2.250$, $p = 0.139$, $p-\eta^2 = 0.037$) such that Down words were on average 14ms faster than Up words. This produced the observed interaction when collapsed into Match and Mismatch

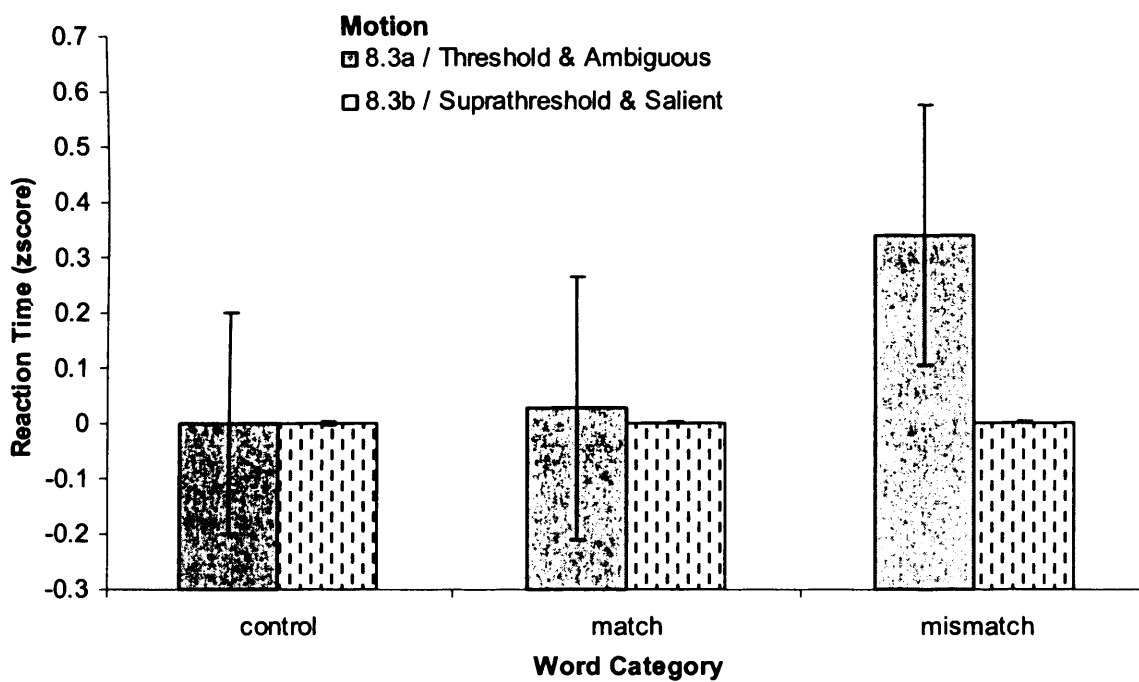
Mismatch conditions ($z = 2.158, p < 0.05$) as compared to Control. See Table 8-4 for the raw error scores (the same pattern was found when error rates were compared as percentages).

Table 8-4: Errors for each Experiment by Condition*

Experiment	Condition		
	Control	Match	Mismatch
Experiment 8.3a: Threshold	5.61 (4.02)	6.71 (4.71)	5.39 (3.77)
Experiment 8.3b: Supra-threshold	3.77 (3.93)	4.85 (2.82)	4.81 (3.54)

*Standard deviation in brackets

8-3: Normalised reaction times for Experiment 8.3a and 8.3b by Condition*



*Error bars are one standard error of the mean

8.3.3.5. *Further Analysis*¹⁹

The error rates suggest that the reaction time changes for experiment 8.3a may have been partially due to a speed-accuracy trade off in the Match condition, which has more errors (and faster reaction times) than the Mismatch condition. Therefore response times in the Match condition may actually be more similar to those found for the Mismatch condition, if accuracy rates between the two conditions were equal. To remove the influence of a speed accuracy trade-off, Down and Up items which showed large differences in error rates across the two presentations (with Upwards or Downwards motion) were removed to create a set of items whose mean difference in error rates across the two conditions was close to 0 ($t < 1$). This reduced the number of items in the Match and Mismatch conditions from 30 to 23²⁰. Following the removal of these items, errors showed a main effect of Experiment, with 1% more errors in Experiment 8.3a (threshold) as compared to 8.3b (suprathreshold) ($F(1,49) = 4.331$, $p-\eta^2 = 0.081$). When error rates within each experiment were compared, there were no significant differences; therefore, any changes in reaction times can now be attributed to the selective effect of visual motion.

Table 8-5 Mean Reaction Times (after Item removal) by Condition and Experiment for 8.3a and 8.3b

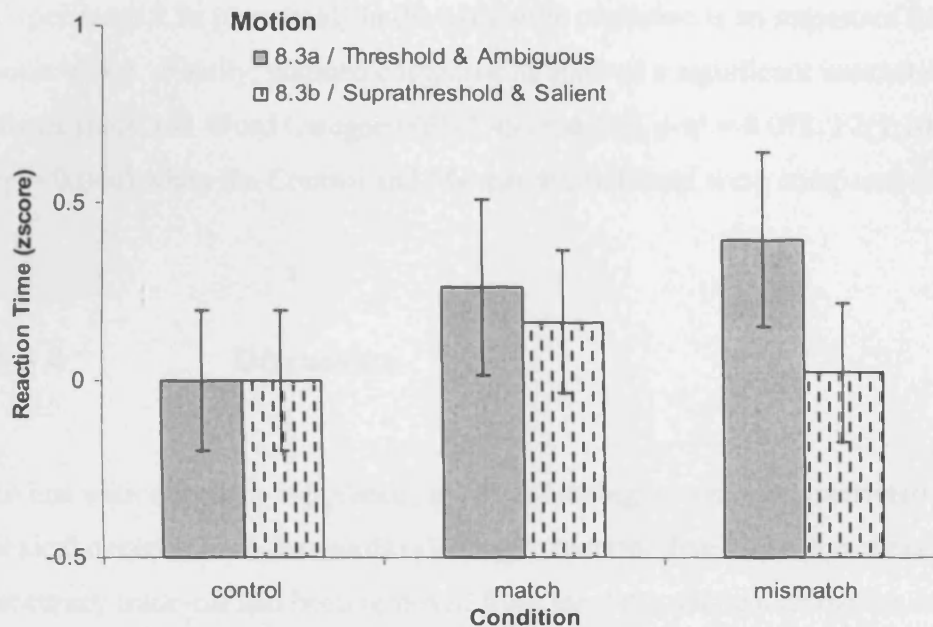
Experiment	Condition					
	Control		Match		Mismatch	
	Down	Up	Down	Up	Down	Up
Experiment 8.3a: Threshold	555.23 (53.32)	554.26 (57.94)	576.27 (65.05)	561.93 (72.88)	573.82 (70.12)	580.21 (66.77)
<i>Mean</i>	554.75 (55.11)		569.11 (68.74)		577.02 (67.84)*	
Experiment 8.3b: Supra-threshold	573.54 (53.28)	570.00 (60.56)	575.91 (64.65)	587.25 (50.14)	580.44 (57.29)	564.86 (56.20)
<i>Mean</i>	571.77 (56.50)		581.58 (57.57)		572.65 (56.73)	

*Significantly different to control, $p < 0.05$

¹⁹ Following completion of the thesis work, this experiment was submitted for publication and following reviewers comments extra participants were run. The pattern indicating a speed accuracy trade-off did not hold.

²⁰ 7 items were removed from the Down category and 9 from the Up category. When re-paired as Match and Mismatch conditions this produced 21 pairs (i.e. data points for Downwards and Upwards motion) with 2 Down items left unpaired. For the items analysis the missing values were replaced with the group mean.

Figure 8-4 Normalised reaction times (after Item removal) for Experiment 8.3a and 8.3b by Condition*



*Error bars are one standard error of the mean

Reaction times were then re-analysed for both experiments. For experiment 8.3a (threshold) there was a significant main effect of Word Category ($F(1,24) = 3.815$, $p-\eta^2 = 0.137$; $F(2,73) = 3.161$, $p-\eta^2 = 0.068$). Reaction times in the Mismatch condition were 23ms longer than those in the Control ($F(1,24) = 6.989$, $p-\eta^2 = 0.226$; $F(2,58) = 7.929$, $p-\eta^2 = 0.120$) and 8ms longer than those in the Match condition. For experiment 8.3b (suprathreshold), there were no significant differences (see Table 8-5).

The reaction times were then normalised and re-analysed across both experiments. This showed a main effect of Word Category ($F(1,98) = 3.553$, $p-\eta^2 = 0.068$; $F(2,146) = 3.824$, $p-\eta^2 = 0.05$). Planned comparisons showed that reaction times in the Match condition were 0.176 SDs longer than those in the Control ($F(1,49) = 5.850$, $p-\eta^2 = 0.107$; $F(2,102) = 5.551$, $p-\eta^2 = 0.052$) and that reaction times in the Mismatch condition were 0.169 SDs longer than those in the Control condition ($F(1,49) = 5.271$, $p-\eta^2 = 0.097$; $F(2,102) = 5.734$, $p-\eta^2 = 0.053$). There was also a main effect of Experiment by Items ($F(1,146) < 1$; $F(2,146) = 5.701$, $p-\eta^2 = 0.033$) driven by the longer reaction times in the Match and Mismatch condition for Experiment 8.3a (threshold) when compared to Experiment 8.3b (suprathreshold). This main effect was significant

in planned comparisons between each Experiment for Mismatch as compared to Match ($F_1 < 1$; $F_2(1,88) = 6.301$, $p-\eta^2 = 0.067$) and Control ($F_1 < 1$; $F_2(1,102) = 4.950$, $p-\eta^2 = 0.046$) conditions. However, it was not significant for comparisons between the Control and Match condition ($F_1 < 0.5$; $F_2 < 1$), signalling that the long reaction times for Experiment 8.3a (threshold) in the Mismatch condition is an important factor in this main effect. Finally, planned comparisons showed a significant interaction between Experiment and Word Category ($F_1(1,49) = 4.152$, $p-\eta^2 = 0.078$; $F_2(1,102) = 4.950$, $p-\eta^2 = 0.046$) when the Control and Mismatch conditions were compared (see Figure 8-4).

8.3.4. Discussion

In line with our main hypothesis, irrelevant ambiguous motion interfered with the lexical decision time for words referring to motion. Even after the effects of a speed-accuracy trade-off had been removed from the data, ambiguous motion interfered more than salient motion in incongruent conditions when compared to the Control condition. However, due to a loss of power (as we had to remove roughly a third of the critical experimental items) this interaction did not survive in the between subjects' analysis across Experiments. The increased reaction times for ambiguous motion in the Mismatch condition produced a main effect of Experiment by Items, lending some extra support to the changes in this condition as compared to Control and Match conditions. Therefore, interference from ambiguous motion is semantically modulated, increasing when the visual motion direction was incongruent with the semantic motion direction. This result provides clear support for a close relationship between semantic representations of motion and the perception of visual motion. The visual motion was entirely irrelevant to the lexical decision task and it is highly unlikely that participants consistently noticed the direction, as it was at the threshold of conscious perception. Finally, the linguistic stimuli were single words, not sentences, and therefore provide a stronger test of embodied semantics. Sentence comprehension requires the integration of semantic content across individual words; therefore it is possible that modality specific information is only recruited during the formation of these more substantial semantic representations. However, we find that incongruent perceptual information does influence the semantic processing of single words. Therefore, the interference comes from the automatic integration of a single word's semantic content with

perceptual information. This is difficult to explain without an embodied account where semantic representations of motion are either based upon, or intimately associated with, the perceptual processing of motion. We know that the ambiguous motion is being processed, and we know that the semantic content of motion words is being accessed. The motion pattern was 200ms long, a length chosen so that it was likely to coincide with semantic access (e.g. Pulvermüller, 2000, 2001). It appears as though the lexical decision cannot be made until the obfuscating influence of the visual motion has passed, so it takes longer to ascertain that the lexical item is genuine. This effect is most pronounced when semantic and visual motion are incongruent: relative to the Control condition, ambiguous motion increases reaction times whereas salient motion does not. When the semantic and visual motion are congruent, we see a slight increase in reaction times for both salient and ambiguous motion. For the Match condition both Salient and Ambiguous motion increased reaction times (as shown in a main effect across both experiments), however, this increase was only significant in the overall comparison and was not significant in either experiment individually.

The fact that ambiguous motion was able to interfere with an irrelevant task supports the original finding of Tsushima et al (2006). The original interpretation, supported by fMRI evidence, was that salient motion signals activate inhibition mechanisms that reduce the influence of irrelevant stimuli, in contrast, ambiguous motion does not provide a strong enough signal to initiate suppression. According to a strong embodied account, the semantic representation of motion words is produced by activity at MT+; therefore suppression of this area during the perception of irrelevant salient motion should cause problems for the comprehension of motion words. If suppression of MT+ does cause problems for motion words, this may be reflected in increased activity in the IFG (indicating harder lexical retrieval) (Thompson-Schill, Bedney & Goldbery, 2005).

One way to expand these results would be to use MEG or fMRI techniques. For MEG, the time course of activation between incongruent and congruent conditions may shed some light on why one condition produces reaction time changes and the other does not. More generally, we know from Tsushima et al (2006) to expect differences in activation for the DLPFC and MT+ when salient or ambiguous motion is present. In addition, we might expect greater activity in the Inferior Frontal Areas (BA44/45) when lexical retrieval is more difficult (Thompson-Schill, Bedney & Goldbery, 2005) during Mismatch/ambiguous motion conditions. We would also expect activity in the temporal

lobes during semantic access and it may be similar for all conditions (Halgren, et al 2002). Of particular interest would be the relative activity of IFG, temporal and MT+ regions during the Match condition as compared to the Control condition. In one case we know that the visual motion should not impact upon semantic access (Control) and in the other we know that there may be some contact between the two (Match). Detailed time-course information or the relative level of activation for different cortical areas may show differences where reaction times do not.

8.4. *General Discussion*

Table 8-6 summarises the results from the experiments presented in this Chapter. Two experiments looking at single word recognition used auditory (8.1a) and then visual (8.3) presentation, with the passive viewing of visual motion. Reaction times for auditory lexical decision were longer for Control words as compared to directional words, regardless of congruence. It is not clear whether this difference was due to a general facilitation from viewing salient visual motion for motion words, or due to uncontrolled lexical variables. It is also possible that congruence effects were present, but they were either too weak to be reliable (e.g. if participants had not consistently viewed the visual motion) or they were masked by noise (e.g. because reaction times were not measured from each word's uniqueness point). When the perception of motion was better controlled, in the study with visual lexical decision, congruence effects were found. It is also difficult to interpret why auditory presentation would produce general facilitation for motion words and visual presentation would produce incongruent interference (for ambiguous motion). Explanations touch on how the brain integrates information from different modalities (e.g. phonological word forms and visual motion versus orthographic word forms and visual motion); however, the semantic representation activated by both phonological and orthographic word forms should be similar (if not the same) and *this* is where congruency effects should be located.

Table 8-6: Descriptive Summary of results

Linguistic Stimuli	Results
Auditory Lexical Decision	Motion words responded to more quickly than control words for both Match and Mismatch conditions.
Visual Lexical Decision	Mismatch interference with ambiguous motion
Auditory sentence comprehension	Match & Mismatch sentences responded to more slowly and less accurately than control sentences.
Visual sentence comprehension (reading)	No effects of congruence

The difference in congruency effects could be genuine and due to the different time courses of semantic access in auditory versus visual word recognition. During auditory lexical decision, participants viewed motion before, during and after the presentation of the item. Visual motion engages motion processing systems, pre-activating them. When semantic access also taps motion processes, this could have produced non-specific priming for all motion words. During visual presentation motion was presented at word onset for a brief time. No reliable effects for salient visual motion were found, but ambiguous motion produced incongruent interference. The shorter duration of motion may have meant that motion processing and semantic access were concurrent, creating incongruent interference as both tap the same directionally selective processes. This explanation relies on some shaky assumptions about the time course of semantic access, the time course of motion processing and the interaction between the two. For example, the general facilitation found for auditory comprehension of motion words requires an account where directionally selective motion processing is able to prime motion processing generally, this is distinctly not the case in the visual perception literature (e.g. Seitz & Watanabe, 2003). The auditory lexical decision experiment had some methodological problems; therefore it is much more sensible to avoid integrating the results and concentrate on conclusions that can be drawn from the visual lexical decision experiment. Here, in line with Tsushima et al (2006), we found that ambiguous motion interfered with a target task whereas salient motion did not. Crucially, the interference for ambiguous motion was semantically modulated, being greatest in incongruent conditions. When the visual system was unable to suppress the motion signal, it interfered with the semantic access for incongruent semantic motion. This result is particularly striking since manipulation of the visual system (i.e. whether or not the motion signal could be suppressed) affected the integration between semantic and perceptual motion. If the result had shown congruent interference, we could have concluded that directionally selective neurons were busy processing the ambiguous

motion signal and were unavailable to represent semantic motion in the same direction (in an extension of the explanation by Kaschak et al, 2005). However, it is when visual motion is incongruent that directionally selective motion processing slows the verification of semantic content: it takes longer to decide that the word is genuine because the ambiguous motion cannot be suppressed and it conflicts with the semantic representation of the item. It looks as though for a brief time, the semantic system cannot distinguish perceptual from semantic motion information. This suggests two things: first, a very close relationship between the perception of motion and the semantic representation of motion; second, a semantic system that can recruit sensory areas in a flexible manner. Suppression of the salient motion signal does not appear to affect semantic access, even though suppression of motion processing areas might be expected to disrupt any semantic representation that is also occurring at that site. For example, direct stimulation of the motor strip influences lexical decision times for motor words (Pulvermüller, Hauk, Nikulin, & Ilmoniemi, 2005). The fact that suppression does not disrupt semantic processing as reflected by behavioural measures (speed and accuracy) suggests that the visual and semantic motion are being integrated *outside* the visual system. When the visual system suppresses motion processing, no integration takes place. However, when motion processing is obligatory (because suppression cannot be initiated) the visual signal is automatically integrated with relevant semantic content (that refers to a direction of motion). This integration is unproblematic when the two are congruent, but disruptive when they conflict.

For auditory sentence comprehension, directional sentences had longer reaction times than control sentences; more errors were also made to directional sentences as compared to control sentences. This result supports a general interference from viewing visual motion. A follow up using self-paced reading did not show any effects of congruence, most likely because of methodological problems. The general interference effect found for auditory comprehension does not support previous results which showed longer reaction times to motion sentences when congruent visual motion was perceived in comparison to incongruent motion (Kaschak et al, 2005). It is possible that congruency effects were weak or not present due to a lack of power. The participant numbers are not hugely different; the original experiment had 37 (Experiment 1) and 48 (Experiment 2) participants, in comparison to the 33 participants who completed 8.1b. Although we had fewer participants, there was roughly twice the amount of directional sentences in comparison to the original experiment (60 versus 32). It is also possible that the

blocked presentation of motion in Kaschak et al (2005) was important in producing congruency effects. Long exposure to one direction of visual motion fatigues directionally selective neurons, therefore it may be that congruency effects only arise when motion processing systems are heavily taxed. When salient motion is perceived during sentence comprehension but is otherwise transient (presented on a trial-by-trial basis), it may be that it can be effectively suppressed and so it does not produce congruency effects. However, the suppression of motion processing could then cause problems for the comprehension of directional sentences, producing longer reaction times and more errors (as seen in 8.1b). This pattern of general interference may be seen for sentences (8.1b) but not single words (8.3b) since sentences produce a more substantial semantic representation with more specific information (a token); thus the demand on motion processing systems may be greater and less transient than the demand created by single word comprehension. Thus, when the motion processing system is suppressed, it shows up as a behavioural impairment for sentences rather than single words (although it is present for both). This explanation is speculative, and would require a follow up experiment that manipulated visual motion in three ways. First, blocked salient visual motion and sentence comprehension should replicate Kaschak et al (2005), showing congruent interference. Second, randomised salient visual motion should produce similar effects to those seen here, showing general interference for motion sentences. Third, ambiguous motion should produce congruency effects since it cannot be suppressed and does consume motion processing resources.

The experiments in this chapter do not allow any strong conclusions about sentence and single word comprehension, at least in comparison to each other. However, the key finding is that low-level visual processes are essential to whether congruency effects are observed. By presenting salient and ambiguous motion, bottom-up modulation of visual processing influences the integration between semantic and perceptual information. When the visual information cannot be suppressed it is automatically integrated with semantic information; causing problems for comprehension when the two conflict. It would be particularly interesting to see if this holds for auditory comprehension, when the central task does not involve visual processing²¹. A more sophisticated version of

²¹ In Tsushima et al (2006) the implication is that visual motion is suppressed during the performance of a concurrent visual task, to liberate visual processing resources. It is an open question whether the same suppression would occur when the central task is not in the visual modality.

the self-paced reading task, employing salient and ambiguous motion presentation, would also allow this effect to be explored for a sentence context which describes a complete motion event (e.g. Zwaan & Taylor, 2006). The results from the lexical decision task cannot be reduced to strategic effects; they therefore provide a clear demonstration of a close and automatic connection between semantic and perceptual motion.

9. Discussion and Conclusion

The four questions posed at the beginning of this investigation will be presented and discussed with reference to the results presented in chapters 5-8. Following the discussion of these questions, two further questions will be posed. The first of these will ask what can be concluded once the results are integrated with the available literature. The second, and final, question will ask what the main conclusions of the investigation are and what these conclusions mean for theories of semantic representation. Further experiments and extensions of the current work will be integrated into the discussion, rather than given in a separate section.

At the beginning of this investigation we defined the sensory domain of vision as being a good candidate to explore effects of embodied semantics. The motor domain has consistently shown effects of embodied semantic representation, both in measures of behaviour (e.g. Glenberg & Kaschak, 2003) and neural activity (e.g. Pulvermüller et al, 2001; 2005). Results for the visual domain are more varied and it was unclear whether low-level visual processes were influenced by semantic access (as predicted by strong embodied theories). Recall that it is possible that the motor (efferent) and perceptual (afferent) domains may respond differently to top-down influences such as language, especially since perception is more directly linked to bottom-up influences from perceptual stimuli. Thus, a systematic investigation of the visual system was necessary. It was also noted that the visual system may respond more variably to semantic information since it routinely integrates information in a task-dependent manner (e.g. Watanabe et al, 2001, Seitz & Watanabe, 2003). In order to explore the variability of effects we looked at different perceptual processes; crucially, there was behavioural work which could motivate our experiments in visuo-spatial attention (Richardson et al, 2003; Bergen et al, in press) and motion perception (Kaschak et al, 2005).

9.1. *Question 1: Are embodied effects present for single words and sentences?*

According to embodied theories, the semantic representation for both sentences and single words is based in simulation. The empirical work was divided into three broad areas: comprehension on visuo-spatial attention, comprehension on motion perception and motion perception on comprehension. Results will be compared within each of these areas in order to assess whether sentences and single words produced similar effects. The results will also be discussed with reference to the general type representations produced by single words as compared to the more specific token representations produced by sentences. The type representations activated by single words may not be substantial enough to interact with perception, so sentences that specify particular events may be more likely to produce consistent results.

Alternatively, the specific information in sentences may constrain the interpretation to such a degree that broader influences are lost: a sentence may shift the focus onto a particular dimension and away from the perceptual event of interest. In this case, single words that present schematic information will produce more consistent results because their salient interpretation is focused on the event of interest.

9.1.1. **Visuo-spatial attention**

In these experiments, the task was to categorise a shape that was presented in the upper or lower visual field. Single word comprehension produced incongruent interference (longer reaction times) when presented in semantically coherent blocks, but not when randomly ordered. The blocked presentation means that it is possible effects were mediated by/a product of, strategic processes: participants may have been aware of the directional content of the words. However, in this case one could also expect congruent facilitation, since strategies should primarily work to the benefit of task performance and capitalise on any relevant information. Average reaction times were approximately 100-200ms faster than categorisation in experiments where the target appeared after active comprehension and 500ms faster than the final experiment where the target appeared during active comprehension (see Table 5-2). It may not have been possible for further improvement in categorisation speed, so any facilitation was hidden as performance was close to ceiling. However, this does not explain away the interference

effect that we found and congruent conditions actually showed a slight increase in reaction times, about half the size of the interference found for incongruent and mixed conditions. Congruent conditions were marginally faster than incongruent conditions, but not faster than mixed blocks. Therefore, congruent conditions also appear to produce some weak interference, rather than any trend for facilitation. The results show that categorisation processes could not exclude the confusing influence of directional semantic content. When the directional information was inconsistent or opposed to the required allocation of visuo-spatial attention, categorisation was slowed down. Congruent conditions, rather than benefiting the allocation of attention, interfered a little bit less. Interestingly, when categorisation followed the active comprehension of a sentence, both congruent facilitation and incongruent interference were found, relative to the control condition. Sentences were randomly ordered and it was unlikely that participants noticed their directional content, since fillers and control sentences were present. Therefore, whether or not participants were aware of the directional content, language which referred to a direction of motion opposed to the required direction of attention slowed categorisation. In both cases the directional context was established, either because comprehension had finished or there had been a sufficiently long exposure to directional words. In contrast, when the directional context was not established, there was no interference: when the target appeared after verb offset within a sentence, directional sentences facilitated reaction times relative to control sentences and when the target appeared after single word offset without semantically coherent blocks, there were no effects of comprehension. Therefore, both single words and sentences interfered with categorisation when they produced a substantive directional context. The general facilitation (when categorisation was during the sentence) may be explained by implicit expectations so is not a strong result, however, congruent facilitation was seen for a complete sentential context (when categorisation was at sentence offset). The results suggest that the token information provided by sentences, in particular the concrete referents that move and therefore have a location, is necessary to facilitate the allocation of attention (as well as interfere with it). In contrast, single words can bias attention over time when a strong directional context is created through blocked presentation and in this case we only see interference. Here, the type information (a direction of motion) still influences attention but in a more general and disruptive manner; this may be because it does not provide any concrete information that can be used immediately by visuo-spatial attention.

The fact that single words and sentences can both produce interference sheds light on the mechanisms at work. Semantic representations of direction and space automatically hinder the effective allocation of visuo-spatial attention to places that conflict with that semantic representation. This suggests that either that semantic representation acts as a cue to a spatial location, or represents the spatial location itself. This second possibility is the embodied explanation. Previous research has found congruent interference; this fits with an embodied explanation that referred to results from mental imagery, which show similar effects (Richardson et al, 2003; Bergen et al, in press). The semantic representation acts like a mental image or a visual representation of a location which interfere with detection at that location (Perky, 1910).

However, our results do not show this pattern, so the explanation from mental imagery is less germane. Similar results are seen in the attention literature, where congruent distracters facilitate target identification and incongruent distracters impair target detection, relative to a neutral condition (Eriksen & Eriksen, 1974). Similarly, congruent cues to a target location facilitate identification and incongruent cues impair identification (Posner, 1980). Therefore, whether a visual distracter or cue is presented, the attention literature demonstrates a beneficial effect of congruence, rather than its interference. This pattern is in line with our results and suggests that semantic content is influencing attention in a similar manner, perhaps like a cue (since it preceded target onset). Single words which refer to direction but *not* a spatial location do not affect attention unless they can form a substantial context. This context biases attention over time, rather than producing an immediate effect as the sentences do. Only sentences show facilitation and only sentences provide concrete information about moving objects. We find both incongruent interference and congruent facilitation when the location is fully specified by a sentence and there is a general benefit to attention when information is incomplete but directional. In this case the semantic content could act like a non-specific cue to target appearance along a particular axis. We know from the motion detection experiments that motion representations are active at this point; but there is no complete event and therefore no end point for the motion, i.e. a concrete destination. As concluded at the end of Chapter 5, the fact that some concrete spatial information is needed before language can affect visuo-spatial attention is in line with previous findings (Bergen et al, in press). This interpretation is also in line with eye-movement studies which show that linguistic referents are indexed quickly to relevant locations in the environment (Richardson & Spivey, 2002) and that the eyes follow the

direction of described events (Spivey, Richardson, Tyler & Young, 2000). Therefore, it is possible that eye-movements to the target location were facilitated by congruent sentences. In line with embodied explanations, it has been proposed that attention is facilitated when semantic and visual objects are similar in appearance: visual simulations of the referents are produced during semantic access and these facilitate the visual identification of those same referents (Estes et al, submitted). In our task that would mean that the language described the location of circles and squares, which it did not. The sentences referred to all kinds of objects and we still found facilitation. This suggests a more abstract facilitation, one that is dependent on valid information about a specific location. Language is routinely used to veridically direct attention to locations in space and quickly respond to objects in the environment, so it may also be routine to integrate semantic information into the allocation of visuo-spatial attention. If this is the case and the allocation of visual attention is open to information from semantics, these results suggest that semantic information is always recruited for this purpose, even to the detriment of attention when the semantic and visual information conflict.

9.1.2. Comprehension on Motion Perception

The motion detection experiments used a task that was more sensitive to low-level visual processing. Participants were presented with patterns of moving dots and had to judge whether the dots moved coherently (in one direction) or randomly. By presenting motion that was at the threshold of visual perception, we were able to look at very subtle changes caused by comprehension. I will first discuss the effects on perceptual sensitivity (d'), and then the effects on measures of criterion (c and β). Reaction time results were, for the most part, uninformative; therefore they will not be discussed at length, but brought in to the overall discussion when relevant.

When blocked single words were unsynchronised to the presentation of motion stimuli, perceptual sensitivity was reduced for incongruent conditions. When synchronised to motion stimuli, blocked congruent words increased perceptual sensitivity. When motion stimuli were presented at verb offset during a sentence, congruent conditions increased perceptual sensitivity. Therefore, both single words and sentences are able to facilitate motion detection²². As for visuo-spatial attention, there are particular

²² Randomly ordered single words presented just prior to motion stimuli reduced perceptual sensitivity in

conditions in which single word and sentence presentation influenced motion detection in a similar way. First, the onset of linguistic stimuli predicts RDK onset (to a greater or lesser degree); second, primary semantic access coincides with RDK presentation and third, there is a substantive motion representation (either from blocked presentation or a sentence). As discussed at the end of Chapter 6, this suggests that perceptual learning may be playing some role in the integration between semantic and visual information (see section 9.2 below for a fuller discussion). When the context is strong (i.e. blocked presentation) but is not predictive of RDK onset, congruent information cannot be used to benefit perception, but incongruent information does interfere. This supports the conclusion that semantic and perceptual information are always integrated, and therefore likely to share a common representation (at least at some point in visual processing). The finding of incongruent interference or congruent facilitation suggests that a common mechanism is at work; benefiting perception when reliable and interfering otherwise.

In order to reliably conclude that single words and sentences do operate in the same way, further experiments are necessary. First, randomly ordered single words with RDKs timed to their offset (i.e. coinciding with primary semantic access) could establish whether a single word's semantic content is enough to affect motion detection. The results found when RDKs were presented after word offset are likely due to the maintenance of semantic content in working memory. Second, to fully explore the importance of a sentence context, a variety of manipulations could be used. First, sentence comprehension with RDKs presented after sentence offset. We might not expect RDKs following sentence offset to show any effects, indicating that motion representations are dynamic and time-locked to the lexical items that specify them (Zwaan & Taylor, 2006). It would also be interesting to look at sentences that are 'empty' in comparison to the ones we have used here, e.g. "It climbed" or "It fell". In this case, results would highlight how necessary concrete referents are for motion representations²³; these sentences provide a syntactic frame without constraining the verb with a particular object. We have only observed incongruent interference for blocked, unsynchronised single words; producing the hypothesis that perceptual

congruent and incongruent conditions. In this case, comprehension of single words preceded motion detection so the general interference from directional words was probably due to the maintenance of semantic information in working memory. This differs from the other experiments where semantic access was contiguous with motion perception, so it will not be discussed.

²³ This manipulation would also shed light on how necessary concrete referents are for visuo-spatial attention.

learning could not operate under these conditions. In that case, it is also necessary to try and produce similar conditions for sentences and see if the same pattern emerges (e.g. passive comprehension and fillers reducing the predictive relationship between sentences and RDKs).

Turning to measures of criterion (c and β), two experiments showed effects of comprehension and two did not. The decision criterion was lowered during congruent conditions when RDK onset was unsynchronised to blocked single words or was after verb-offset within a sentence. The similarity between these two experiments is the relative predictability of RDK onset. When words were unsynchronised to RDK onset, there was no predictive relationship between the two stimuli. During sentence comprehension, fillers were added to increase the variability of RDK onset so that it was not completely predictable; occurring at some point during the sentence. Congruent information can benefit the accuracy of detection (d') when it has a predictive relationship to the target stimulus, when it does not it can still be used to enhance decision processes. It is sensible to think that a congruent context would facilitate decision making, without necessarily improving accuracy. If semantic information agrees with the (unpredictable) target stimulus, the participant is more likely to respond that the target was present. One interpretation is that lowered decision thresholds mean less sensory evidence is required to detect the target (i.e. to make a decision) (e.g. Hawkins et al, 1990). Here, the target information appears to be more salient so the decision is actually easier (rather than being a product of contextual priming between semantic and visual information). During sentence comprehension, reaction times were slightly faster during congruent conditions, lending further support to the fact that decisions were easier to make. We will return to this when discussing the levels of visual processing that are influenced by comprehension (Question 2). For this question, the important point is that both single words and sentences affected criterion setting in a similar way, when the target event was relatively unpredictable.

For motion detection, both blocked single words and sentences that were contiguous with RDK presentation influenced motion detection. This supports a common motion representation for both single words and sentences. It is unclear whether randomly ordered single words with contiguous RDKs could produce similar effects; thus, it remains to be seen how effective the type motion from single words is in comparison to the token motion from sentences. Note that the influence of sentences was weak but

beneficial. This may be because the token information made the representation more specific, shifting the focus away from the motion event per se and onto the objects that were moving (e.g. their path and location, as found for visuo-spatial attention).

9.1.3. **Motion perception on Comprehension**

In the final set of experiments, the influence of passively perceiving motion on comprehension was explored. Methodological problems meant that no conclusions could be drawn about how motion perception influenced the comprehension of motion sentences. In the auditory sentence judgement task the results suggested that there was general interference from motion perception on all directional sentences (producing slower reaction times and higher error rates), however this result conflicts with the existing literature (possibly because of methodological differences) and it was not backed up in other experiments (see experiment 8.2 and Appendix 3). In the final experiment, a visual lexical decision task showed that ambiguous motion slowed reaction times when incongruent with the target word. It is an open question whether similar effects would be found for words within a sentence context, and further experiments that extend this ambiguous/salient motion paradigm to sentence comprehension are needed. The existing literature shows that visual motion perception interferes with sentence comprehension during congruent conditions (Kaschak et al, 2005); therefore it is possible that single words and sentences are influenced in different ways. No firm conclusions can be drawn since Kaschak et al used salient motion with auditory presentation and it is not clear how important auditory versus visual presentation is for the different effects. It would be interesting to explore why salient motion did have an effect in Kaschak et al (2005) but not in the lexical decision experiment; this could be due to the different demands on comprehension for single words and sentences. During sentence comprehension the semantic content of individual items has to be integrated, therefore perceptual information may also be more likely to be integrated. In contrast, visual lexical decision is a recognition process so salient visual motion that interferes with orthographic identification is suppressed. Experiments that use visual presentation (e.g. a more sophisticated self-paced reading task and/or visual lexical decision) whilst manipulating the salience of visual motion are required to pull apart the different possibilities.

9.1.4. Summary and Conclusions

The experiments that explored the influence of comprehension on perception showed that, under certain conditions, single words and sentences influenced visual processing in similar ways. When single words are presented in a way that allows them to build up a directional context over time, the results are broadly similar to when a sentence is presented. It could be that very small effects are present with randomly ordered single words, but are not detected in these experiments due to a lack of power. Experiments that present randomly ordered single words that are contiguous with target presentation would help to address this point; time-locking semantic access to target presentation seems particularly important for motion perception, so it may be more likely to tap effects in both visual domains. However, the present results show that the important thing is not just a semantic representation; that representation has to be reliable and robust. When single words achieve that through blocked presentation, we see effects on visual perception that are similar to a sentence context. When robust semantic information is reliable (i.e. predictive of target onset), the visual system shows facilitation. When it is unreliable, the visual system shows interference. The results from measures of criterion also show that a robust context primes decision making, further supporting the integration of an established semantic context with visual perception. However, the importance of a robust semantic context is a hypothesis rather than a conclusion. Further experiments that manipulate the contiguity between linguistic and motion stimuli, and the content of sentences (e.g. using 'empty' sentences), are needed before it can be concluded that a robust semantic representation is the crucial factor.

The results from the final lexical decision experiment, which demonstrated that the unconscious perception of incongruent motion interfered with the recognition of motion words, suggest that a robust directional context is not needed in order for perception to influence comprehension. The results from the comprehension of randomly ordered single words on motion detection also showed *some* influence. Thus suggests that, *for motion*, contiguity between semantic and visual information may be the critical factor, rather than a strong directional context. When the two occur simultaneously, we see effects between visual and semantic processes; this supports the dynamic nature of semantic representations and the importance of tapping primary semantic access. For

visuo-spatial attention, it may well be that a more robust semantic context is needed since representations of space and location are being tapped (see Question 3 below).

The results support analogous foundations for the semantic representations of single words and sentences but the type / token distinction that is present between single words and sentences also makes a difference. For visuo-spatial attention, the concrete referents provided by sentences are crucial in producing immediate, beneficial effects. This is in line with the fact that single words referring to *objects* do influence visuo-spatial attention (Estes et al, submitted) and supports the view that language is routinely used to guide attention to objects and locations in the environment. The results from motion detection can be taken to reinforce the results from visuo-spatial attention. In this case, the effect of sentence stimuli on motion may have been attenuated precisely because the sentences provided more information (e.g. about objects and locations) and shifted the focus away from the motion event itself. In contrast, blocked single words produced a strong motion representation that was not constrained by any particular object or event. Thus, the difference between the two (type vs. token) still has consequences for the way in which perception is affected. This demonstrates a remarkable subtlety in the interactions between comprehension and perception.

The results support a common representation between single words and sentences and embodied effects are found for both. The results vary depending on the specific information (type vs. token) contained within the semantic representation, this emphasises the importance of detailed semantic content for interactions between comprehension and perception. The comparison between single words and sentences supports strong embodied predictions about the content of semantic representations (i.e. lexico-semantic and sentential representations show similar effects). Weaker versions of embodiment can also explain these results, as long as they allow substantial detail in the modal content that is engaged during semantic representation. Given the complexity of modality specific information that is likely to be present at the conceptual and/or perceptual level, this is not necessarily a problem.

9.2. *Question 2: Are low-level visual processes implicated in interactions between comprehension and perception?*

Previous studies which used reaction time as the dependent measure implicate some level of visual processing but it is not clear whether the effects are due to low-level visual processes (e.g. visual signal processing) or effects at later stages (e.g. decision and response processing). One of the main aims of this investigation was to isolate specific visual processes and separate the levels of processing that are being influenced. Therefore the key results are from the experiments which explored the effect of comprehension on motion detection and used Signal Detection Theory to separate levels of task processing. For these experiments we compared the effects on d' (perceptual sensitivity), c and β (where the decision threshold is set) and reaction time (the entire time to complete the task). In addition, Experiment 8.3 manipulated low-level visual processes (presenting ambiguous or salient motion) and looked at effects on lexical decision; therefore results from this study shows how low-level visual processes interact with semantic representation.

All the experiments which used motion detection showed effects on d' and it was effects at later stages that were more variable. The results showed that when semantic access was contiguous with motion perception, congruent conditions increased d' (Experiments 6.4 & 6.5) or incongruent conditions decreased d' (Experiments 6.2). In one experiment where semantic access preceded motion stimuli, a general reduction in d' was observed for all directional words (Experiment 6.3). The TMS experiment (Chapter 7) showed a tendency for congruent conditions to impair d' relative to control (but not to incongruent) conditions. In contrast, only 2 out of these 5 experiments showed reliable changes in measures of decision threshold (Experiments 6.2 & 6.4). In one case the change in decision threshold was in a different condition to changes in d' (6.2) and in the other case the change was in the same condition (6.4). Finally, one experiment showed a change in reaction times not caused by congruence (6.2); one showed reaction times differences that reflected an effect of congruence not seen in other measures (6.5) and one showed an effect of congruence that was present in other dependent measures (6.4). Therefore, d' was consistently influenced by manipulating congruence between the semantic and visual motion stimuli where other dependent variables were not. D' represents the strength of the signal as detected by the participant, being directly proportional to the separation of distributions of signal and noise in a detector during a

2-alternative forced choice task. A decrease in d' indicates impaired sensory level stimulus processing, resulting in a diminished separation of signal and noise distributions hence, poorer perceptual sensitivity (Wickens, 2002). From motion detection tasks we know that d' measures the earliest stages of motion perception, since direct stimulation of the motion processing complex MT+ typically produces a reduction in d' (e.g. Silvanto, Lavie & Walsh, 2005). Given that d' taps the earliest stages of motion processing, we can conclude that low-level visual processes are reliably influenced by semantic content that is congruent or incongruent with visual stimuli. Evidence from the visual lexical decision experiment with irrelevant motion (Experiment 8.3) provides converging evidence for a close connection between low-level visual processes and semantic content. The presentation of ambiguous and salient motion manipulated low-level motion processing bottom-up (i.e. changes in external stimuli that affected processing). In line with previous findings (Tsushima et al, 2005), salient motion was suppressed and did not interfere with lexical decision; in contrast, ambiguous motion was not suppressed and it produced longer reaction times for lexical decision to incongruent words. In the original study the account was that general visual resources were consumed by the processing of the ambiguous motion signal, reducing performance in the central task. In contrast, our results showed interference only when the semantic content conflicted with the ambiguous motion. If the interference from ambiguous motion is the product of taxing visual resources, this suggests that semantic content was also placing some burden on visual processes, modifying the conditions in which interference occurred. I will now discuss the possible mechanisms that would produce these effects, in order to establish how closely our results support a weak or strong view of embodied semantics.

In the discussion below, I will not include the results from the TMS study (Experiment 7.1) as unforeseen methodological problems produced results that were inconclusive. Changes in presenting visual stimuli also meant that the observed congruency effects are not directly comparable to the other experiments. However, the finding that d' was the only dependent measure to show significant changes is taken as further support that low-level visual processes are implicated more heavily than higher order decision processes.

For the other experiments which used motion detection, when semantic access occurred at the same time as motion perception, the d' results showed congruent facilitation and

incongruent interference. The lexical decision experiment also showed incongruent interference (in reaction times). Strong embodied explanations propose this pattern is the product of the semantic representation of motion words engaging motion processing areas. In some sense, the semantic representation is similar to actual visual motion. This argument has been used to explain why the perception of congruent motion interferes with comprehension of motion sentences (Kaschak et al, 2005), since when the two are processed concurrently they compete for the same neural resources. Directionally selective neurones have to process both visual and semantic motion at the same time. In contrast to this, we find incongruent interference and congruent facilitation. Rather than competition for the same resources, this pattern suggests integration of the semantic and visual information *as if the two were the same*. In line with embodied theories the visual and semantic systems appear to make no distinction between motion information that comes from visual or linguistic stimuli. For this reason, I would like to conclude that there is a close and automatic connection between visual and semantic information, therefore some version of embodiment is needed to account for the results. Strong embodied theories that stipulate sensory-motor simulation as semantic representation can account for the results seen here, although the absence of congruent interference (e.g. Kaschak et al, 2005) would need to be explained. However, a weaker version of embodiment can also explain the results, stipulating that semantic and visual information are integrated outside the visual system rather than supported by the same substrate within it.

As noted in the discussion of Chapter 6 (section 6.6.1) the results can be explained within a framework of perceptual learning, where the visual system actively integrates information from stimuli that are present alongside the target event (Seitz & Watanabe, 2003; Watanabe, Nanez & Sasaki, 2001). This framework can be added to strong versions of embodiment and account for some of the task based differences in results. In this case, the motion words do directly engage motion processing areas and if there are predictive relationships between visual and linguistic stimuli the visual system can use this semantic-visual activation to improve perception. The comparison that strong embodied theories invite is that semantic activity is equivalent to presenting a motion pattern in the same or different direction to the target. The visual perception literature shows that a distracter motion pattern congruent with a subsequent target has been shown to facilitate or interfere with motion detection (as indicated by a lowered or raised threshold respectively). Congruent distracters produce interference or facilitation

depending on whether or not they are attended; supporting the idea that areas such as MT+ are routinely modulated by attention (Raymond, 2000). In other words, bottom-up manipulations of the visual system are modulated by top-down factors such as attention. The RDK stimuli we have used engage second-order motion processing, also known as 'object-based' motion processing since information is extracted from complex stimuli with multiple motion vectors and segmented into discrete object representations (ibid). There is an "emerging consensus... that object based perceptual and attentional mechanisms may integrate with motion processing at this level" (ibid, p.44). This is an important point since it means that our results, showing consistent low-level interactions between motion perception and comprehension may not be a simple product of direct engagement, but a more complex picture where motion processing integrates information from several sources (e.g. the visual stimuli and whatever else is attended). Semantic information may present another source of information that can be integrated. As noted above, the fact the visual and semantic information are integrated at low-levels and treated as isomorphic at some point in visual processing means that the connection between the two is close and automatic. However, the weaker versions of embodied theories can also account for this data, given what we know about the visual system. In some embodied accounts, sensory and motor areas are co-opted into semantic representation through *associations* between the lexical stimuli and the sensory-motor experience (Pulvermüller, 1999). Whilst this account still places a clear emphasis of the necessity of sensory and motor areas in semantics, this same principle can produce a semantic system that is actually independent of sensory and motor information, whilst still benefiting from a close connection to that information. Elsewhere, theories have recognised a connection between semantic and sensory-motor systems, without committing to strong embodiment (Vigliocco et al, 2004; Rogers et al, 2004). Here, semantic representations are derived from conceptual information, which is grounded in sensory and motor information (Vigliocco et al, 2004). Semantic representations are one-step removed from direct engagement; derived from direct experience but not necessarily recreating it. It is possible that semantic representations are themselves cues to particular events, allowing access to the relevant information (and therefore having close connections to systems engaged in direct experience) but not presenting that information per se (Rogers et al, 2004). For our stimuli, this means a permanent connection between the semantic representation of motion words and motion processing areas. Thus, when the two are activated in close temporal proximity they can influence each other via a bidirectional connection; as we have seen incongruent interference

occurs for both comprehension on perception and perception on comprehension. This relationship is one of mediation, rather than modulation, and is also entirely compatible with perceptual learning and other mechanisms at work during task execution. Indeed, one would expect it to be influenced by task strategies which could utilise this connection to a greater or lesser degree. This account is admittedly very speculative, but it is laid out here in order to make a point: embodiment does not imply simulation. There are many ways that a close and meaningful connection between semantic and sensory-motor information can be produced, rejecting the idea of a completely amodal, abstract semantic system²⁴.

Finally, I want to discuss the possible mechanisms that would produce the results seen for criterion (c and β). In two experiments (6.2 & 6.4) congruence reduced the decision threshold: participants were more likely to respond that a target even occurred. This is in contrast to their accuracy in detecting whether that even actually did occur, which was improved by congruence in only one of these experiments (6.4). This criterion shift was present when a robust semantic context was present (i.e. either through blocked word presentation or a sentence context) and when the onset of the RDK was relatively unpredictable; indicating that it was due to strategic processes. The criterion shift for sentence comprehension was weaker as compared to blocked, unsynchronised single word presentation; this is in line with the token linguistic context being more constrained (perhaps resulting in a weaker motion representation) and the RDK being more predictable. RDK onset was quite predictable when sentences were comprehended but it was still less so than experiments where RDK onset always occurred either during or after the presentation of a word (6.3 & 6.5). It is possible that when the visual system cannot predict the onset of the RDK, a congruent context increases the salience of the target event (i.e. the motion direction). Therefore, deciding whether a particular motion direction was observed is easier and criterion is lowered (see below for a further discussion of criterion effects). Further experiments that carefully manipulated the predictability of RDK presentation in concert with the

²⁴ This alternative account can also explain the non-specific interference from directional words found for Experiment 6.3. Participants heard a single word and then shortly afterwards they saw an RDK. Crucially, filler trials asked questions about the word so participants had to maintain semantic information whilst they made the motion detection judgement. We found that motion detection was worse for directional words as compared to control words. According to the weak embodied account laid out above, the maintenance of semantic information produces recurrent activity within the semantic system and therefore also in the connection between motion words and motion processing areas. The weak activation of this connection could then have impaired motion detection since it would consume resources that were trying to be allocated to motion detection.

strength of the linguistic context would be needed to verify whether criterion shifts are modulated by these variables. Manipulating the strength of the motion in sentential representations (e.g. “The rocket flew upwards” versus “The women rose from bed”) would be an interesting way to explore the constraints of token representations on semantic interpretation, particular if the same directional verb was used (e.g. “The rocket climbed into the sky” versus “The child climbed onto the chair”).

A final point is that as criterion was lowered in response to congruent conditions (indicating an easier decision); speeded response times in the same conditions could be a product of decision priming, rather than low-level sensory or motor processing. However, in the one experiment where we found a trend for facilitated response times and decision processes, we also found facilitation at the perceptual level. In addition, if stimulus-response compatibility is reduced, or removed, in experimental conditions, it is unlikely that speeded reaction times will represent decision priming (see Question 3 for a fuller discussion of this argument). Therefore, this issue will not be problematic as long as experiments take care to tap the interactions between semantic and sensory-motor content implicitly (or adopt methods where low-level processing can be measured). Happily, much of the existing work using reaction time measures does just that.

9.2.1. **Summary and Conclusions**

All the experiments that measured perceptual sensitivity (d') showed interactions between semantic content and motion detection. The unconscious and obligatory perception of motion (i.e. a signal at threshold that could not be suppressed) was also found to interfere with the recognition of incongruent motion words. Therefore low-level visual processes are implicated in these interactions. This connection is automatic and somewhat direct; this conclusion is best supported by the lexical decision experiment that showed interactions between irrelevant motion and comprehension when the visual information was not consciously perceived. The lexical decision experiment also showed that when perceptual information was suppressed, it did not influence semantic access; nor did it appear to impair semantic processing. This supports the weaker version of embodiment laid out above, where visual and semantic information is integrated (e.g. via a permanent connection). Strong embodiment predicts that the suppression of motion processing will impair the simulation of

semantic motion: the resources needed for simulation are curbed. We do not find evidence for this.

The results mean that (at least for some tasks) visual and semantic motion information are not separable, they do share a common representation. Again, the results have to be accounted for by some version of embodiment. Embodiment implicates sensory and motor systems in semantic representation; therefore what is known about how these areas work (e.g. perceptual learning and suppression) has to be incorporated into explanations. For that reason, task based differences in the results support a novel version of weak embodiment that can account for the data.

9.3. ***Question 3: Are visuo-spatial attention and motion perception affected in the same way by comprehension?***

In order to manipulate visuo-spatial attention, we used a shape classification task that had been used previously to show the effect of semantics on attention (Richardson et al, 2003; Bergen et al, in press). Even when location is manipulated implicitly, the allocation of visuo-spatial attention to locations in space involves quite a lot of top-down control (e.g. the identification of the target space, the conscious direction of attention to that space, oculo-motor control). In fact, attention is considered an archetypal top-down influence on perception itself (Raymond, 2000). The motion perception task initially involved judging the direction of dynamic motion patterns, but the majority of experiments used motion detection where a coherent motion signal had to be detected. In contrast to allocating visuo-spatial attention, the detection of coherent motion signals in central vision requires decisions about low-level perceptual signals. Therefore, it does not demand that visuo-spatial attention is adjusted and less top-down control is needed to complete the task. One requires the active identification of a location (and the object at that location) and the other requires the classification of an ambiguous perceptual event. The similarities and differences between the results using these different tasks will highlight how higher and lower order visual processes are affected by semantic information.

In order to answer this question we need to, as far as possible, directly compare results from those experiments which differ only in their visual manipulation: visuo-spatial (shape categorisation in different hemifields) or motion (motion detection). There are three pairs that fall in this category. First, blocked passive comprehension of single words whilst concurrently performing the visual task; second, active comprehension of single words followed by the visual task; third, active comprehension of sentences with the visual task at verb offset. It is possible to compare and contrast the patterns of facilitation and interference in different conditions of congruence. The one limitation is that the only dependent measure for the visuo-spatial task was reaction time, whereas the motion detection tasks have two extra measures from Signal Detection analysis (sensitivity and criterion). Where possible explanations will take this into account.

For the blocked single word comprehension and visuo-spatial target (5.1), reaction times were longer when blocks presented incongruent or mixed directional words. There are actually two different motion tasks that were combined with blocked single word comprehension: direction judgement (6.1a) and motion detection (6.2). For direction judgement, reaction times were faster in congruent and mixed conditions, as compared to incongruent conditions (the control condition had averages that were similar to those in the incongruent condition). For motion detection, reaction times were not changed by congruence, but perceptual sensitivity was reduced in incongruent conditions and criterion (c) was reduced in congruent conditions. The results for each task diverge from each other, however, there are two instances of incongruent interference and two instances of congruent facilitation. Incongruent conditions produced longer reaction times for the visuo-spatial task and lower d' values for motion detection. Congruent conditions produced shorter reaction times for the motion direction task and lower criterion values for motion detection. In other words, measures of signal processing pattern with the visuo-spatial task whereas measures of decision threshold pattern with the motion direction task. The task in the visuo-spatial experiment was shape categorisation, so response options (circle or square) were independent from the manipulation of visuo-spatial location (top or bottom). Therefore, biases to allocate attention to a particular location should affect visual processes involved in target identification (i.e. signal processing) rather than response preparation; as such, the reaction times results reflect processes at this level. There was no difference in error rates, but this may be because performance was close to ceiling (error rates were around 5-6%). It was only by manipulating visual stimuli close to threshold

levels that we were able to see small effects on sensitivity reflected in accuracy rates. However, in line with the interpretation that the reaction time changes in the visuo-spatial task reflect effects at low-levels, the motion detection task also showed incongruent interference for d' . In contrast, when judging the *direction* of motion (6.1a), the visual manipulation (up or down) is conflated with response preparation (up or down). Reaction times showed congruent facilitation. It looks as though in this task, the reaction times reflect changes in decision and response preparation. As noted when these pilot experiments were discussed, this data could reflect something like the Simon effect, where faster responses are made when stimulus-response variables are congruent. In line with this interpretation, the motion detection task shows congruent facilitation at the decision level; with the criterion set lower under congruent conditions. In terms of the congruence between stimuli and response, the motion detection task is actually half-way between the visuo-spatial and motion direction tasks. Participants have to respond if motion is present or absent, but when motion is present it is either up or down; participants could very well be preparing the response 'up or random'/'down or random'. In this case, there is some overlap between stimulus and response, producing congruent facilitation for decision thresholds. The effects on signal processing are revealed in d' , just as they are present in reaction times when stimulus-response compatibility is removed in the visuo-spatial task. These three experiments illustrate very well the importance of choosing tasks which remove this kind of compatibility between the critical manipulation (the stimulus) and the response; reaction times can reflect the processes of interest but these can also be swamped by more trivial effects at the decision/response stage. Despite the difference in the tasks, we see common patterns of incongruent interference at low-levels of visual processing and congruent facilitation for decision/response processes.

The second pair of experiments that can be compared is active comprehension of single words followed by the visual task. In this case there is no robust directional context and the effects of semantic content are arguably due to the maintenance of semantic information in working memory rather than the result of immediate semantic access. The visuo-spatial task showed no effects of congruence; for the motion detection task, perceptual sensitivity was reduced for directional words. The discussion of results for visuo-spatial location (Ch.5 and Question 1 above) concluded that the allocation of attention was only affected when a robust directional context was present; either specifying a concrete location (as in sentence comprehension) or consistently biasing

attention to a general location (downwards or upwards) so that a particular location becomes salient. In that case, it is not surprising that there were no effects of congruence. The visuo-spatial task was simply not sensitive to the transient semantic representation of motion. In contrast, the more sensitive motion detection task tapped not only the relevant visual stimulus (motion rather than location) but also used threshold stimuli that are sensitive to small effects. The results showed that motion detection was impaired following the comprehension of all directional words. Therefore, maintaining semantic motion information produced a general demand on the motion processing system, perhaps consuming general visual resources that were then no longer available for motion detection. In this case, the results reflect the difference in sensitivity between the two visual tasks to the processes of interest. Allocating attention to a region in space in order to identify a target is not influenced by transient semantic motion; a specified concrete location is the most effective way to influence attention. For motion detection (being closer in nature to the semantic information) even the briefest activation of semantic motion is sufficient to influence low-level motion processing.

The final pair of experiments that can be compared are those which used active comprehension of sentences with the visual task at verb offset. Here, the dual-task method meant the visual stimulus was perceived during the remainder of the sentence. The visuo-spatial task showed that reaction times were faster for directional sentences overall, as compared to the control sentences. The motion detection task showed congruent facilitation at all levels; sensitivity was increased, criterion was lowered and reaction times were slightly faster. Therefore there is a general benefit for visuo-spatial attention and a specific (congruent) benefit for motion perception. This can also be explained by the difference between what each of these tasks actually taps. The semantic representation of the sentence is still under construction when the stimulus is perceived; there is incomplete information aside from the object being in motion along the vertical axis (there is no end point). Clear congruency effects in visuo-spatial tasks only emerge when a final destination for the motion is specified (i.e. the target appears after sentence offset, Experiment 5.3). The general benefit found when the target appeared at verb-offset might be due to some visual-semantic information being better than none, preparing the allocation of attention along a given axis. In the original experiment the target could appear at the top, bottom, left or right of the screen (Richardson et al, 2003); therefore axis was manipulated rather than a specific location.

In order to conclude that partial sentence comprehension primes attention along one axis, we would need an experiment that showed facilitation for targets in the top and bottom of the visual field, but not for those in the left or right, when targets are presented at verb-offset.

The fact that there was a representation of motion (but perhaps not location) is evidenced by the result from motion detection. When RDKs were presented at verb offset, sensitivity was increased for congruent conditions, as though the semantic information primed the detection of visual motion. Congruence also produced a lower criterion and faster reaction times. This is in line with experiments that show transient (localised to the verb) but facilitatory effects of perceiving rotational motion on reading times for verbs referring to rotation (Zwaan & Taylor, 2006). From these experiments it appears that the motion representation is short-lived, occurring only at the point at which it is lexically specified. If we assume the same is true for the sentences we used, the motion representation is active shortly after the verb (thereby influencing motion perception) with a fuller representation of the object's location appearing once it has been specified, at the end of the sentence (and thereby influencing visuo-spatial attention). In combination with one another, the results from these two experiments demonstrate the dynamic nature of semantic representations during comprehension and the visual processes they affect according to the content of those representations. This further supports the close connection between comprehension and visual processes, showing very subtle differences in which processes are affected. These different processes are engaged depending on the relevance of the semantic content to those processes; motion perception is influenced when motion is represented and visuo-spatial attention is affected when a visuo-spatial location is specified. This is quite striking in its specificity, supporting much other work which shows fine-grained influences of semantic content on visual perception (e.g. Zwaan, 2004). It is here where simulation offers the most intuitive account of the data; if the details of each semantic representation are simulated then the consequent processing of similar perceptual events is facilitated. The important question is whether simulation is the only explanation of these effects. Eye-movement studies show that semantic information is quickly indexed to objects in the environment and that attention to objects and locations is dynamically manipulated depending on content (Chambers et al, 2002; 2004). This indexing could be a product of simulation, with the simulation of congruent items allowing them to be quickly identified in the environment (as perceptual and semantic

information agree). However, this would run against previous experiments that showed congruent interference for allocating visuo-spatial attention (Richardson et al, 2003; Bergen et al, in press; Estes et al, submitted). Congruent interference is explained by the simulation of visual motion interfering with the identification of objects in the environment (as in Perky, 1910). Therefore, these two explanations from strong embodiment conflict just as the results presented here conflict with the previous demonstrations of congruent interference for visuo-spatial attention; this will be discussed further in section 9.5.1.

Weak versions of embodiment would explain the present results by the perception of particular visual events (i.e. objects, locations, motion) being primed by connections to semantic content, through strong associations rather than simulation. Distributed network theories of semantic content propose a spreading activation across features that constitute a semantic representation (e.g. Farah & McClelland, 1991; McRae, de Sa & Seidenberg, 1997; Rogers et al, 2004). The sensory and motor systems could be part of this distributed network because they are involved in conceptual representations from which semantic content is derived (Vigliocco et al, 2004). When sensory-motor systems are activated by task processes, e.g. object recognition or motor preparation, as well as by associations to semantic content, the observed effects of congruence could be the cross over between these two paths of activation; one coming from semantics and one from the task. The associations between semantic access and experience are non-arbitrary, so it is not correct to speak of semantic representation as abstract or symbolic. This explanation is still embodied, but semantic content acts more like a cue to a particular event than representing the event itself (e.g. Elman, 2004) and it relies on the integration of sensory-motor and semantic content. This account is a post-hoc alternative and has some problems dealing with the apparent detail of semantic representations, implicitly priming the perception of things such as object form (e.g. whether a bird has its wings outstretched for flight or tucked in for perching; Zwaan, Stanfield & Yaxley, 2002). There would have to be rich, contextually variable connections between semantic representations and object recognition processes. However, it is informative to look at how other theories of semantic representation might account for these effects without using the mechanism of simulation.

9.3.1. Summary and Conclusions

The barrier to comparing visuo-spatial attention and motion detection tasks was the difference in dependent measures. Future experiments which manipulate visuo-spatial attention in a similar fashion to motion detection, presenting stimuli that are on the threshold of detection (by manipulating their duration for example) could be used so that signal detection measures can be extracted. However, the comparison of results has been informative. First, it has shown how reaction times can reflect early visual processes as long as stimulus-response compatibility effects are removed. Second, it has shown that both tasks show incongruent facilitation when stimulus presentation is unsynchronized and congruent facilitation when presentation is synchronized (i.e. blocked word versus sentence comprehension). Therefore, both tasks are unable to suppress the interference when the visual and semantic information conflict, but both tasks are able to use that information to the benefit of perception when circumstances allow. Where the tasks diverge is in the precise aspect of semantic content to which they are sensitive. Visuo-spatial attention requires the specification of a particular location or a robust directional context that can progressively bias attention towards a location. Motion perception is apparently always sensitive to semantic motion, even when that information is merely being maintained in working memory. This provides further support for a close, automatic connection between semantic content and visual processing. Strong versions of embodied theory can account for these results through the simulation of visual motion, although the results do not support the congruent interference that is found in other studies (Richardson et al, 2003; Bergen et al, in press). The patterns of incongruent interference and congruent facilitation that are found for both visuo-spatial attention and motion perception support the routine *integration* of semantic content at different levels of visual perception. This is more in line with weaker versions of embodied theory, where a permanent connection provides a common base for semantic and visual motion, without stipulating that the visual system is engaged in simulation.

9.4. ***Question 4: For visual motion, are congruency effects in the same conditions for comprehension on perception and perception on comprehension?***

Several experiments demonstrated how the comprehension of motion words variably influenced low-level visual processes involved in motion perception (Chapter 6 & 7). However, only one experiment showed reliable effects of motion perception on the comprehension of motion words (Experiment 8.3); this experiment will first be summarised in some detail, along with any conclusions that can be drawn. The result will then be compared to the motion detection experiments in order to find any similarities.

Two visual lexical decision experiments looked at the influence of passively viewed motion on comprehension. A 200ms RDK was presented at word onset; in experimental trials the motion was congruent, incongruent or neutral in comparison to the meaning of the word. When the motion was ambiguous (i.e. at threshold), there were longer reaction times for incongruent conditions. When the motion was salient, no effects of congruency were observed. This paradigm was motivated by previous results which showed that irrelevant, salient motion was suppressed so it would not take resources from the central task. In contrast, ambiguous motion could not be suppressed (since it does not breach the threshold that initiates suppression); it therefore takes resources from the central task and impairs performance (Tsushima et al, 2006). We replicated the interference from ambiguous motion, supporting the suppression of irrelevant salient motion in order to maintain task performance. However, the general resource account put forward by Tsushima et al does not explain these results. If the processing of the ambiguous motion signal consumed general resources, there should be interference in all conditions relative to salient motion (admittedly this would be difficult to detect with our between subjects manipulation). Instead, the interference was semantically modulated, appearing only when the ambiguous motion conflicted with the semantic motion. Directionally selective motion processing slowed the verification of semantic content when the two were incongruent. We assume that for salient motion there is direct suppression of motion processing (MT+) from the frontal lobes (ibid). This result shows that obligatory²⁵ visual processing influences semantic access. This is very strong evidence for an automatic connection between the visual and

²⁵ Ambiguous signals that cannot be suppressed and are therefore not open to strategic effects or reliable

semantic systems. However, the suppression of salient motion did not affect semantic access (for example, slowing recognition to congruent or incongruent motion words); this indicates that visual motion processing can be suppressed without impairing the processing of motion words. This goes against predictions from strong embodiment.

The incongruent interference seen in the lexical decision experiment is in line with the general patterns found for the motion detection experiments; one experiment showed incongruent interference (6.2) and two showed congruent facilitation (6.4 & 6.5). As has already been discussed, perceptual learning may play an important role in the motion detection experiments; certainly in those which show congruent facilitation. However, the one experiment where the words could not be employed to benefit perception (as there was no predictive relationship between the motion words and the RDKs), we find incongruent interference for d' . In this experiment blocked single words were passively comprehended whilst participants detected motion. This experiment is not that easily compared to the lexical decision, except that both experiments present a single task (lexical decision or motion detection) with the other stimulus passively perceived. We can also argue that in both experiments the irrelevant stimulus was not being used for the benefit of task performance. For lexical decision, apart from the fact that ambiguous motion produced interference, motion was not informative for lexical decision (i.e. it did not signal lexicality) since it appeared for all stimuli²⁶. For motion detection, the onset of the words did not predict RDK onset so sensitivity could not improve. Thus, both experiments show interference during incongruent conditions and in both cases we know that the interference has its basis in low-level visual perception. In the lexical decision experiment it is the result of a bottom-up manipulation of visual perception on comprehension and in the motion detection experiment it is the result of a top-down effect of comprehension on sensitivity. Therefore, there are parallels in the way perception affects comprehension and vice-versa.

conscious perception.

²⁶ It is an open question whether manipulations of predictability between visual and semantic motion would change the patterns seen for the lexical decision experiment. For example, one direction of motion could be predictive of lexicality (providing an opportunity for learning), whilst still congruent, incongruent and neutral in comparison to word meaning.

9.4.1. **Summary and Conclusions**

Two experiments, each looking at a particular direction of influence between perception and comprehension, showed incongruent interference. Both experiments used a single task paradigm where the irrelevant stimulus was passively perceived. The results reinforce the conclusion that semantic and visual motion are treated alike; it is not just motion processing which shows interference from perception; perception similarly affects comprehension. It is important that both experiments tap low-level visual processes, reinforcing the inference that the link between semantic and visual motion is rooted in the earliest stages of visual motion processing. However, the lexical decision experiment also showed that when salient motion signals were suppressed, semantic access was not noticeably impaired. It appears that there are both top-down and bottom-up connections between semantic and visual motion that are automatically and directly activated; but they can also be modulated by controlled processes, such as suppression and perceptual learning. This supports a weaker version of embodiment that does not require simulation as the driving mechanism.

Unfortunately, no conclusions can be drawn about sentence comprehension and motion perception. However, the fact that results were found for single word presentation suggests the correct methodology should find comparable effects for sentences.

9.5. *Further Discussion*

9.5.1. **How do the present results compare with the available literature?**

9.5.1.1. *Visuo-spatial attention:*

The first set of experiments in this investigation were motivated by the finding that sentences referring to an object's motion (Richardson et al, 2003; Bergen et al, in press) or location (Estes et al, submitted) affected the allocation of visuo-spatial attention. In all of these studies, when the semantic and visual target direction/location were congruent, reaction times to categorise the target object were longer. Explanations of this effect have centred around 'image schemas', visual similarity and results from conscious visual imagery (Perky, 1910). The argument is that comprehension activates visual representations (similar to conscious visual imagery) which interfere with the identification of objects in the same location. The act of comprehension is therefore similar to the processes of producing visual imagery, and both (presumably) use the visual system. In contrast to these results, we found incongruent interference when the semantic and visual locations were incongruent (blocked single words and sentences) and congruent facilitation when the semantic and visual locations were congruent (sentences only). This pattern of results contrasts with the existing literature. The method that we used was a direct extension of that used in Richardson et al (2003) and Bergen et al (in press), therefore the differences may be due the kinds of sentences that were employed. Richardson et al used both concrete and abstract sentences and collapsed results across these categories. Bergen et al had a very small number of items (5 in each condition), but found effects only with concrete sentences. For the sentence stimuli, we had 30 items per condition all of which were designed to be as concrete as possible; admittedly the number of participants was smaller, but the effects were still reliable. As noted at the end of Chapter 5, an experiment which used all of the stimuli from all the experiments so far (within subjects) would help to clarify whether item selection is responsible for the different effects. However, our results do agree with the conclusion that language needs to refer to concrete events in order to effect visuo-spatial attention (Bergen et al, in press), since single words did not influence attention and the influence of blocked single words built up over time, suggesting that a strong context

was responsible (rather than the semantic content of the individual words per se). Elsewhere, single words have been shown to affect visuo-spatial attention if they refer to concrete spatial locations and relationships (Zwaan & Yaxley, 2003; Estes et al, in press). It is of interest that one of these studies also showed incongruent interference/congruent facilitation when pairs of words were presented in the correct or incorrect spatial relationship (e.g. 'attic' above or below 'basement') (Zwaan & Yaxley, 2003). These results pattern with our experiments, but the methodology is very different (looking at comprehension on judgements about word pairs). We directly adapted the method from Richardson et al (2003) so replications and extensions of the current experiments are needed to explore how very similar tasks can produce congruence effects in different conditions; as for interactions between motor action and sentence comprehension, the timing may be critical (Borreggine & Kaschak, 2006).

The work that shows congruent interference during the allocation of visuo-spatial attention has been accounted for with explanations based on visual imagery (Richardson et al, 2003; Estes et al, submitted). This explanation is less applicable to our results unless we allow that 'visual schemas' produced during semantic access interfere with visuo-spatial attention in both congruent and incongruent conditions, with some as yet unidentified variable accounting for the difference between data sets. Experiments that look at the influence of comprehension on eye-movements are a good place to try and tease apart the conflicting results. We know that eye-movements do follow the path of described events and return to the location of named objects, even if there are no visual stimuli to see (Spivey & Geng, 2001; Spivey et al, 2000; Richardson & Spivey, 2000). Some of this work has been explained as the externalisation of working memory; using sensory-motor markers to aid in recall and task execution (e.g. Wilson, 2002). I would like to argue that the embodied effects we see in our experiments (incongruent interference and some congruent facilitation) are due to attention, and possibly the eyes, going with the described event. The semantic system does influence the allocation of attention, when it represents locations to which attention can respond; the representation of the location could be embodied (i.e. coded within the visual/oculo-motor systems) without being an image (i.e. a visual simulation). For example, the representations of perceptual targets and motor plans for saccades have been shown to have a similar representation (Eckstein et al, 2007); if semantic information provides valid information about an object location this could also be coded as a saccade target, a plan for action. If language is routinely used to veridically direct attention, the eyes and attention will

tend to go where semantic content informs them to, and show interference (i.e. slower responses) when it has to work against this information.

Experiments are needed which monitor the latency and pattern of eye-movements in response to comprehension, with the critical comparison being whether latencies to targets in particular locations (e.g. up or down) are facilitated or impaired by directional sentences. If imagery disrupts the allocation of attention, congruent conditions should be slower relative to incongruent conditions. Alternatively, if attention follows semantic cues, incongruent conditions should be slower. It would also be particularly useful to explore how tracking an object with the eyes would affect the comprehension of directional sentences, in a reversal of the experiments presented here. Just as we found similar effects between comprehension and motion, it is interesting to explore the bidirectional relationship between visuo-spatial attention and comprehension. Finally, experiments could employ secondary visual tasks, e.g. the maintenance of imagery (e.g. a coloured shape) or a visuo-spatial location (e.g. left-top or right-bottom). If attention goes with the specification of a location, rather than being disrupted by an image, a visuo-spatial task should be more disruptive than an image. In contrast, if attention is disrupted by an image, the pattern should be reversed.

For the moment, I would like to argue that the 'visual imagery' interpretation of semantic representation (Richardson et al, 2003; Bergen et al, in press; Estes et al, submitted) is more complex than it appears; especially since no experiments of this kind have directly manipulated the degree of visual imagery (if similar processes are in operation, the effects of comprehension and imagery could be additive). One of the most well developed theories of embodied simulation states very clearly that simulation is a separate process to conscious visual imagery (Barsalou, 1999; see section 1.3.2.5.1), although the two may be tap the same apparatus. The reasons for this are sensible; first, if simulation underpins conceptual and semantic representation it cannot be conscious (since little of the conceptual and semantic processing that we do is conscious). Second, visual imagery experiments typically show that it takes seconds (rather than milliseconds) to produce and manipulate visual imagery. In comparison comprehension is very fast. Third, semantic representations are schematic, dynamic and (for Barsalou, 1999) decomposed; they are not static, detailed recreations of visual scenes. If semantic representation is to be placed on a continuum with visual imagery, there is much more that needs to be specified.

There is some evidence that a secondary visuo-spatial task interferes with narrative comprehension more than a verbal task (Fincher-Keifer, 2001). There are the numerous results which show that sentence comprehension facilitates the processing of congruent pictures, again supporting something like visual imagery during comprehension (e.g. Stanfield & Zwaan, 2001). The results from narrative comprehension also provide compelling evidence for image like representations or mental models that are produced during comprehension, for example, analogously representing space, distance and objects (e.g. Zwaan, 1999). However, the experiment usually referenced when explaining congruent interference during the allocation of visuo-spatial attention is one that employed conscious *effortful* visual imagery. In the Perky experiment (1910), participants actively imagined an object on the wall in front of them; they were subsequently slower to identify the appearance of a real visual stimulus in that location. I find it problematic to compare automatic effects from semantic access with conscious, effortful visual imagery. Future experiments will need to explore the continuum between semantic, imagistic and conscious imagery processes to refine this account. The long running debate about visual imagery has shown that similar neural substrates are active during actual visual perception and conscious visual imagery (e.g. Kosslyn, 1996); therefore, if the semantic representation of visual information also taps these same substrates, it is entirely possible that there are similar processes in operation. However, this question has not yet been extensively investigated.

As we have seen, visuo-spatial attention responds to a linguistic context, particularly when concrete locations are specified. Similarly, motion perception and the semantic representation of motion show a bidirectional influence on each other, suggesting a shared representation at some point in processing. These influences are automatic, dynamic and transient. They are also entirely dependent on what is specified in the semantic content, and influences on the perceptual system vary accordingly. Comprehension and semantic representation are closely connected to the systems that we use in everyday perception and action. This is no doubt an essential part of how semantics works; engaging the systems through which word forms can be associated to referents.

Chapter 6 and 7 detailed experiments which explored the influence of comprehending language referring to motion on the perception of visual motion. All experiments found that comprehension influenced perception at the level of perceptual sensitivity (d'), i.e. the stage at which visual signal processing occurs. There are no visual perception tasks in the literature that are directly comparable to the motion detection task used here; the existing studies all use reaction times as the dependent variable. In addition, only one study (to the best of my knowledge) has looked at comprehension on motion perception. There are two experiments looking at the perception of dynamic motion on comprehension, but these will be discussed below (when discussing similar experiments from this investigation). Zwaan et al (2004) presented sentences over headphones, with filler sentences followed by comprehension questions. On each trial the sentence was followed by two pictures, separated by a brief mask, and participants had to judge if the second picture was the same as the first. On critical trials the sentences described the motion of a ball towards or away from the comprehender; for example, "The pitcher hurled the softball at you" and "You hurled the softball at the pitcher". The pictures were of a similar ball to that described in the sentence (e.g. for the above sentences, a softball) and the second picture was slightly smaller or slightly larger than the first; thereby approximating the motion of the ball towards (larger) or away (smaller). When the motion of the ball in the sentence was congruent with the pictured motion, judgements were faster. There are two things to note about this method, first, comprehension precedes the motion stimuli; second, the motion is apparent. Although apparent motion (i.e. motion created by a succession of static stimuli) form the basis of all computer animation, in this experiment participants compare the first and second pictures, making salient the fact that the pictures are separate events. However, the manipulation of object size between the first and second picture (to produce apparent motion) is implicit; any influence of comprehension on perception is therefore automatic. The result showed congruent facilitation, supporting the hypothesis that comprehension of the motion sentences had produced a simulation of motion within the visual system; this system was then primed to process pictures congruent with that motion. We also found congruent facilitation of motion perception when the visual motion stimulus was presented during sentence comprehension. We found facilitation at every level of processing, from perceptual sensitivity (d') to criterion setting (c and β) and a weak effect on reaction times. Due to demonstrations elsewhere that motion

perception only influenced comprehension at the verb (Zwaan & Taylor, 2006), suggesting motion information was localised to this region, we did not look at motion detection following sentence comprehension. Therefore it is not possible to directly compare the two experiments. However, the results are in line with Zwaan et al (2004) and with much other work which shows facilitatory effects on congruent picture judgement following sentence comprehension (see section 2.1.2.3). Zwaan et al (2004) concluded that their experiment hinted at the dynamic nature of semantic representations (by using implicit apparent motion); the motion detection tasks we have used firmly support the dynamic representation of semantic motion. Our results also add to these previous findings since they did not employ a control condition; therefore the results could have been congruent facilitation or incongruent interference. The finding of congruent facilitation in our sentence comprehension experiment suggests that this may be the condition of interest. It is interesting that the sentence comprehension experiment was the only experiment in which all dependent measures converged on a pattern of congruent facilitation. The complete event described by sentences provides a stable semantic context which interacts with perception in a more uniform and pervasive manner. An interesting question is whether the speeded reaction time results we found were due to facilitation at the perceptual level (d'), the decision level (c and β) or both. From other experiments we know that effects at the decision level can diverge from effects at the perceptual level (experiment 6.2) and that reaction times can diverge from both SDT measures (all experiments except 6.4). One other experiment showed congruent facilitation at the decision level when blocked single words were presented (experiment 6.2), indicating that congruent contexts (when substantive enough) produce decision priming. It is not clear what mechanism produces shifts in response bias/criterion; signal detection theory allows the *separation of criterion and stimulus factors*, rather than providing a clear framework for the interpretation of shifts in criterion²⁷. One interpretation is that a downwards criterion shift indicates less sensory evidence is needed to make a decision (e.g. Sperling, 1984). Congruent contexts may make the target event appear more likely, producing a reduction in criterion values as expectations prime responses to the target event. In this case the congruent context would have to be sufficiently robust to influence implicit

²⁷ It has been noted that “An understanding of the psychological factors that are involved in bias shifts...is, if anything, farther away than an understanding of stimulus effects” (Wickens, 2002, p.75). See also

'expectations'; this is in line with reliable criterion shifts only appearing for blocked single words and sentences²⁸.

The important point for this work is that if sentence comprehension canonically produces a lower criterion in congruent conditions, the reaction time changes in other experiments could be a product of this rather than facilitated perceptual processing (i.e. effects at d'). As has been argued above, as long as the interactions between semantic and sensory-motor variables are tapped implicitly, it is unlikely that speeded reaction times are due to effects at the decision level. Although the influence of sentence comprehension on decision processes might offer a more trivial explanation of reaction time effects, they do provide further support for the automatic way in which semantic information is integrated into perceptual processing. When sentences are presented, the experimental manipulation of direction is never made explicit to participants and sentences are presented in a random order amongst fillers which disguise their content. It is unlikely that participants use conscious strategies and any congruency effects between semantics and perception are automatic. It is encouraging that our experiment also showed facilitation in perceptual processing: sentence comprehension could have a pervasive influence that affects both decision and perception in similar ways. Crucially, the origin of both of these influences is in the automatic connection between semantic and perceptual motion.

It is more difficult to integrate the results we found with single words into the existing literature as few other studies have used single word comprehension without a sentence context. In fact, there are no studies looking at motion perception and single word comprehension aside from the ones presented here. These results have already been integrated with findings from the perceptual learning literature, which show that practise and predictable stimulus relationships improve performance (Gold, Sekuler & Bennett, 2004; Seitz & Watanabe, 2003; Watanabe, Nanez & Sasaki, 2001). This was necessary in order to interpret the findings from experiment 6.5, in which congruent words synchronised to motion stimuli improved motion detection across the course of a block. This was in contrast to experiment 6.2, in which incongruent conditions impaired

²⁸ Models of response bias based on the decoding of sensory information to produce 'sensory likelihoods' have been developed elsewhere: "the subjective experience of sensory events arises from the representation of sensory likelihoods, and not directly from the responses of sensory neuron populations" Jazayeri & Movshon, 2007, p.915. Therefore congruent conditions could influence the decoding of sensory likelihoods towards the target event, producing a downwards shift in criterion.

motion detection (and the words were unsynchronised to motion stimuli). Since the results showed changes in perceptual sensitivity we know that semantic information influenced perception at low-levels. It has been argued that perceptual learning (i.e. practice effects) work primarily to increase the strength of the perceptual signal and operate at early stages of perception (e.g. before the binocular integration of visual information) (Gold, Sekuler & Bennett, 2004). Therefore, if our results are due to perceptual learning, this is likely a product of the low-level connection between semantic and perceptual motion. Just as perceptual learning works to increase the strength of the signal, semantic information can sensitise the visual system towards congruent signals. Most importantly, when learning cannot take place, semantic information cannot be ignored and instead is integrated to the detriment of perception (possibly because it is sensitising perception in the 'wrong' direction).

In sum, it looks like semantic and perceptual motion are habitually integrated at low-levels of visual perception; in addition, the representation of semantic motion is dynamic, influencing the perception of RDK stimuli. This argues against the interpretation that shifts in criterion might be the cause of faster reaction times for previous experiments involving sentence comprehension (see above). The majority of these experiments manipulated the critical variables implicitly, so that participants were not aware of relationships between semantic and perceptual factors. As discussed elsewhere (see Question 3 above), reaction times can reflect both perceptual and decision/response processes. When issues of response congruence are removed reaction times can reflect changes in perceptual processes. I would like to conclude that when our results are combined with the previous literature, the results support a low-level locus for effects. Criterion shifts may be the exception rather than the rule, occurring when the context is robust and the appearance of the stimulus is unpredictable so decision processes are more open to beneficial, contextual information.

9.5.1.3. *Motion perception on Comprehension*

There are two experiments which have demonstrated the influence of motion perception on comprehension (Kaschak et al, 2005; 2006). Participants judged the sensibility of sentences whilst viewing a motion stimulus presented for approximately 30 seconds. Reaction times showed that sentence comprehension was slower when the semantic and perceptual motion were in the same direction (Kaschak et al, 2005). Following this,

further experiments looked at the modality of sentence comprehension and motion perception (Kaschak et al, 2006). In Experiment 1, four second segments of auditory motion (white noise filtered to give the perception of motion) were presented whilst participants read sentences presented one word at a time via rapid serial visual presentation (RSVP). Participants heard each auditory percept 8 times whilst comprehending 10 sentences (therefore the percept was blocked in a similar fashion to Kaschak et al, 2005). Reaction times again showed slower responses during congruent conditions. In a second experiment the same auditory percept was used but sentences were also presented aurally; in this case reaction times were faster in congruent conditions (Experiment 2). This reversal of the congruence effect was replicated when both experiments were completed within subjects (Experiment 3). These results were explained by the overlap between semantic and perceptual stimuli and the assumption of limited resources. When the two are processed sequentially, congruent facilitation is produced; the same resources are used for both and priming occurs between the semantic simulation and the perceptual processing of a similar percept. In contrast, when the two are processed simultaneously, incongruent facilitation is produced; resources are consumed by one task and are therefore unavailable to the other. Presenting the two in different modalities is equivalent to simultaneous processing and it produces congruent interference. Based on the assumption that comprehension is attentionally demanding, presenting the two in the same modality is equivalent to consecutive processing. Comprehension means that “participants were temporarily unable to process the ongoing auditory stimuli. On this view, the fact that the sentence and stimulus were not processed simultaneously means that there was no competition for resources. Instead, there was a priming effect that arose when participants processed a stimulus depicting motion in one direction immediately before processing a sentence that described motion in the same direction.” (Kaschak et al, 2006, p.742).

One experiment from this investigation found an effect of motion perception on comprehension (experiment 8.3). A visual lexical decision task showed that ambiguous motion slowed reaction times when incongruent with the target word. This experiment was motivated by a study which showed the suppression of irrelevant, salient motion stimuli whereas ambiguous motion did not initiate the same suppression (Tsushima et al, 2006). In line with this result we found that salient motion did not influence lexical decision times. In Kaschak et al’s experiments the motion stimulus (visual or auditory) was salient and irrelevant to the task, so it is possible that similar suppressive

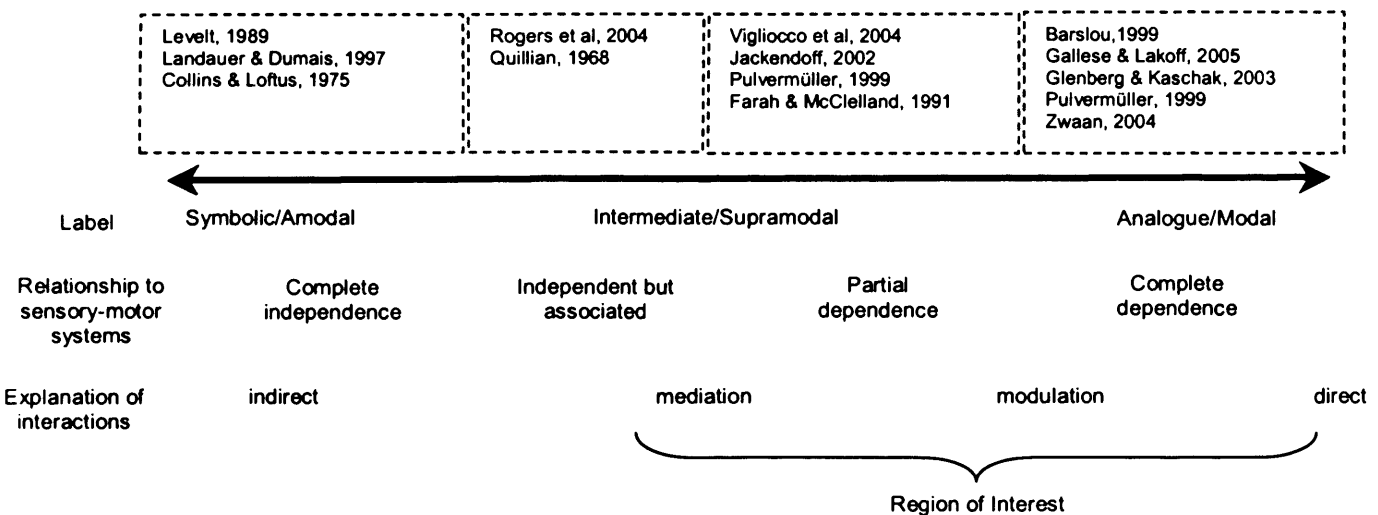
mechanisms were in operation (at least for the visual stimuli). In that case, the congruent interference may be a product of the salient stimulus being suppressed and it is this that causes problems for semantic representation (i.e. the semantic system cannot use these suppressed resources). When the two are presented aurally incongruent interference arises because the auditory stimulus is not suppressed, being processed along with the other auditory information (i.e. the sentence). Note that this explanation depends on suppressive mechanisms that benefit task execution in vision but not audition; this remains to be established. However, by employing this explanation, the results fit with what we find for the lexical decision experiment, stimuli that are obligatorily processed interfere with semantic information when they are incongruent with it. The lexical decision experiment did not show error rates or reaction times to support the inference that suppression caused problems for semantics, but the stimuli (words and motion) were presented for a short time. In Kaschak et al the blocking of motion stimuli means that these suppressive mechanisms may have been well established and constant during comprehension; thus it produced congruency effects. The benefit of this highly speculative interpretation is that it reconciles the results between all the experiments. It is difficult to apply the explanation from Kaschak et al to the lexical decision results since we find incongruent interference. The explanation that consecutive processing produces priming between perceptual and semantic information rests on interpreting their results as congruent facilitation, rather than incongruent interference. If we had found congruent facilitation it would be relatively simple to say that the brief motion stimulus primed semantic access in congruent conditions (although it was timed to *coincide* with semantic access it was present from word onset). However, since we find incongruent interference it is not clear how their explanation would fit this data.

Experiments that use visual presentation (e.g. a more sophisticated self-paced reading task and/or visual lexical decision) whilst manipulating the salience and blocking of visual motion are required to pull apart the different possibilities. In addition, since Tsushima et al (2006) implicate specific cortical areas, imaging experiments that look at motion processing (MT+), semantic processing (superior temporal areas) and the mechanisms of suppression (Dorsolateral PreFrontal Cortex) would also help to inform whether suppression is playing an important role when stimuli are presented concurrently.

9.5.2. What conclusions can be drawn about the relationship between semantic and visual motion, what implications does this have for theories of semantic representation as a whole and embodied theories in particular?

Sensory and motor information was implicated in theories of semantics early on (Quillian, 1968) and there is gradient from those theories which propose semantics is amodal (Levelt, 1989), amodal but connected to modal information (Rogers et al, 2004; Jackendoff, 2002), supramodal (Vigliocco et al, 2004) and finally necessarily modal (Barsalou, 1999; Gallese & Lakoff, 2005; Pulvermüller, 1999; Zwaan, 2004; Glenberg & Kaschak, 2003) (see Figure 9-1). The different theoretical assumptions about modality specific information produce different predictions: Abstract, amodal theories predict that there should be no interaction between semantic access and sensory-motor systems, since the two use distinct and separated domains. Intermediate theories propose that interactions can occur, depending on the task and situation, but that semantic representation is still possible without access to sensory-motor information. Embodied theories predict that sensory-motor information is a necessary part of semantic representation, so interactions between the two should be consistent and prevalent.

9-1: Schematic of where theories lie along the continuum from amodal to modal
 *Tyler & Moss; McRae, de Sa & Seidenberg and Smith, Shoben & Rips not included as no clear assertions are made



The key findings from this investigation are as follows:

- (1) There are bidirectional interactions between low-level visual processes and semantic content (lexical and sentential).
- (2) Interactions are automatic, arising whenever linguistic and visual stimuli are presented in close temporal contiguity
- (3) Interactions are subject to processes within the visual system such as perceptual learning and suppression.
- (4) The precise content of semantic representations dictates which visual processes are implicated.

Taken together these key findings argue strongly against amodal theories and the results need some form of embodiment (i.e. stable connections between semantic and sensory-motor systems) to explain them. The implication for all theories of semantics is that they need to include non-arbitrary connections between the semantic representation of sensory or motor referents and the systems used to experience those referents. Semantic content affects all levels of task performance but this is not because comprehension has a global influence on cognition (i.e. it does not operate indirectly). Interactions between low-level visual processes and semantic content are always present whilst decision and response processes are variably affected depending on the task. Our experiments show that there is an automatic bidirectional connection between the semantic representation of motion and the perception/representation of visual motion. This suggests that connections between low-level sensory-motor processes and semantics are relatively direct and independent from strategic/decision effects. The connections are dynamic, rich and flexible: the content of semantic representations, for example the difference between the representation of motion and the representation of a concrete location, dictate which visual processes are implicated in interactions. Some form of embodiment is necessary to explain these results.

The strong prediction from embodied theories is *direct engagement*: semantic content *necessarily* and *directly* recruits the sensory and motor systems used during experience. Sensory and motor systems are essential for semantic representation and they are engaged during semantic access without being mediated by other cognitive processes. Interactions between the two are evidence of sensory or motor activity being modulated as they simulate semantic content. A weaker prediction is *indirect engagement*: semantic content requires close contact to sensory and motor systems but it is *not*

dependent on those systems. Sensory and motor systems are implicated because of stable associative relationships between semantic representation and experience and they are engaged during semantic access in a task dependent manner. Interactions between the two are evidence of sensory or motor activity being mediated by semantic content. Although interactions between the two were always present, we did see task based differences in the interactions between semantic and visual motion. In particular, processes that operate on low-level visual perception (such as perceptual learning and suppression) dictated whether semantic information could be used to the benefit of perception (experiment 6.5) and whether visual information interfered with semantic comprehension (experiment 8.3).

The strong embodied interpretation is that the semantic representation uses directionally selective motion processes. When these are engaged in processing incongruent motion it takes longer for semantic content to be simulated 'correctly' or for visual motion to be perceived as easily. Perceptual learning is possible because the modulation of motion processes primes visual perception, just as a congruent visual stimulus would. Suppression of the visual system should modulate semantic representation and our experiment may not been sensitive to this (it is highly speculative whether it has been demonstrated elsewhere, Kaschak et al, 2005). A weak embodied interpretation is that semantic content activates links to directionally selective motion processes. The links allow semantic content to refer to visual experience and mean that to some extent semantic and visual information are analogous. These links are bidirectional so semantic and visual areas mediate the activity of the other (e.g. Rowe et al, 2005). This mediation is automatic and direct so it disrupts processing when the motion is incongruent. However, when the activity precedes stimulus onset it can prime perception. Suppression of activity in the visual system should not influence semantic processing although it should hinder the effective connection between them.

For strong embodiment, semantic and visual processes compete for the same resources. For weak embodiment, semantic and visual information can share a common representation through their stable connection and it is the integration between the two that produces congruency effects. The experiments which showed incongruent interference support the idea that semantic and visual information are obligatorily connected, but we did not find congruent interference (except when incongruent conditions also produces some interference so the two cannot be separated, e.g.

experiments 6.3 and 7.1). Therefore, the results are not compatible with the argument that congruent conditions produce competition for identical resources (Kaschak et al, 2005). In addition, the dependency of congruency effects on task based manipulations (e.g. perceptual learning, suppression, the linguistic context and contiguity) is more in line with a weak version of embodiment that would predict such variation. Taken together, the results support the mediation of visual processing by semantic information, rather than direct engagement. However, the weak prediction that congruent interference is the exception rather than the rule conflicts with the literature for both visuo-spatial attention (Richardson et al, 2003; Bergen et al, in press; Estes et al, submitted) and motion perception (Kaschak et al, 2005; 2006).

9.5.3. **Concluding Statement**

We have presented data demonstrating that (i) automatic interactions operate between semantic representations and low-levels of sensory processing and that (ii) different visual processes are influenced by specific semantic content. Sensory and motor information is involved in semantic representation, but this does not require simulation or a strong version of embodied theory: the results we have presented can be explained by both strong and weak versions of embodiment. Figure 9-1 shows the region of interest that needs to be explored in order to separate these alternatives. In order to progress, embodied theories need to take into account what is known about sensory and motor processes; these systems have been intensively studied elsewhere and there is a wealth of information that can inform semantic theory. The manner in which sensory systems are affected by perceptual learning, suppression and other task based processes is a fruitful way to explore their interactions with semantics. This investigation has been a first step in that direction.

Appendix 1a: Instructions for the Rebus Norming Procedure

Instructions

In this task you will be given sentences in the form:

'circle' *verb* 'square'

or

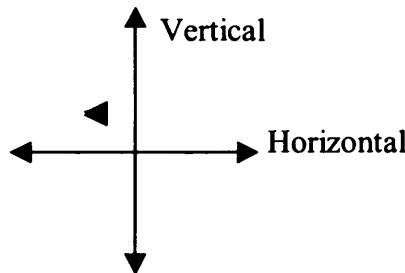
'circle' *verb*

We want you to match each of these sentences with one of 8 picture representations of the event that the sentence describes. The pictures all take the form of a circle and square (or just a circle) and an arrow that represents the event.

For example:



There are **two axes** along which the action can take place: vertical and horizontal. There are **two directions** along the axes.



Please select the pictures that you feel **best represent** the event described in the sentence, with your order of preference.

Use 1 for the most preferred picture, 2 for the next preferred and so on.

Please select a **minimum of one** picture for each sentence.

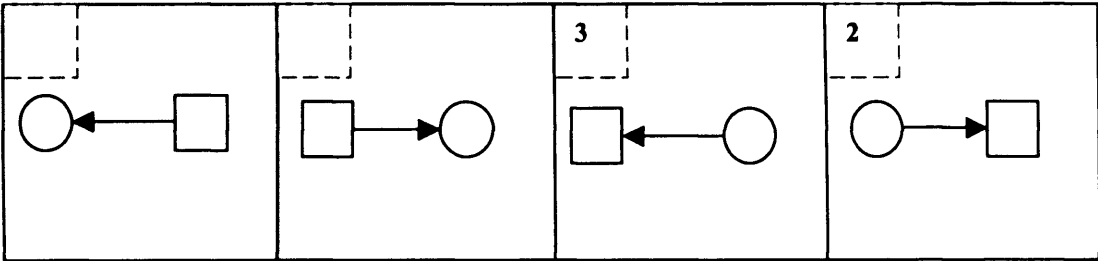
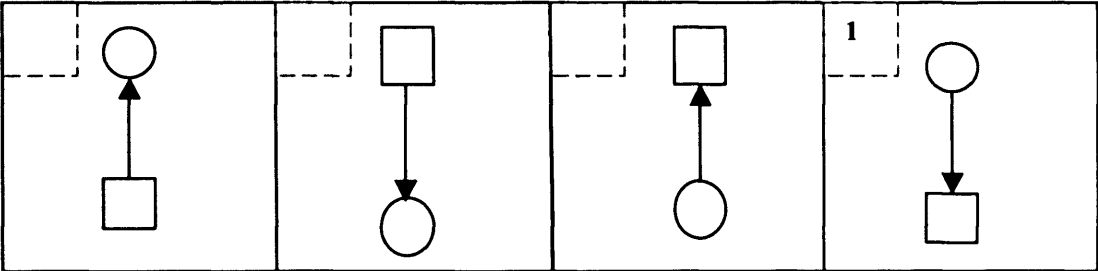
Write your choices in the box in the top left of each picture (please see examples overleaf)

Please **look at all the pictures** before you make each of your choices.

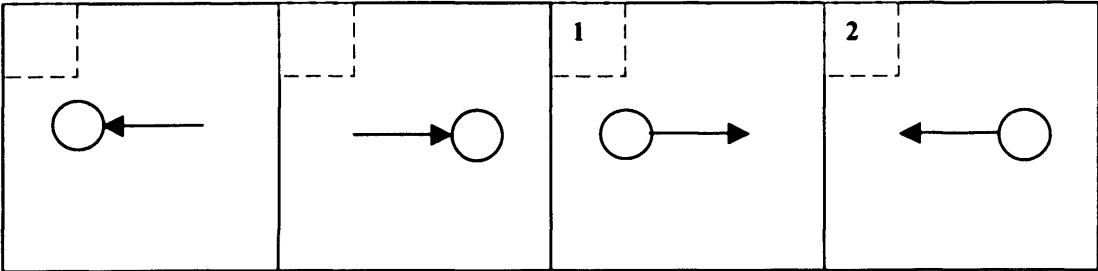
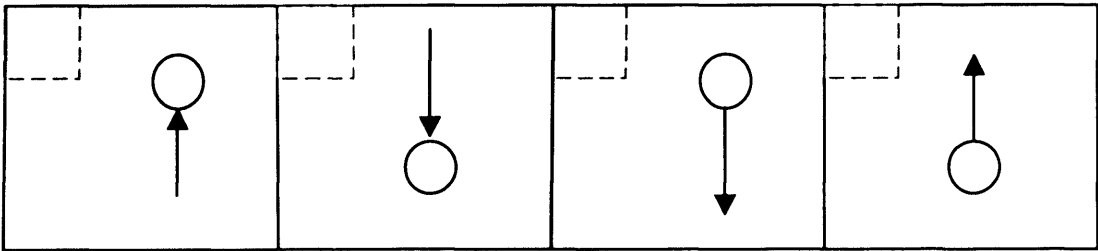
Please take as much time as you like for each sentence.

If you do not understand the verb then please write so on the page and move on to the next sentence. Your responses will be kept anonymous and confidential. If you have any questions please do not hesitate to ask the experimenter, you are free to leave at any point without penalty.

'circle' writes 'square'



'circle' strolls



Appendix 1b: List of verbs presented for norming in intransitive and transitive rebus sentences

Verb
ascend
bounce
climb
decrease
depart
depreciate
descend
diminish
drop
hang
hope
increase
inhale
move
sink
slide
stop

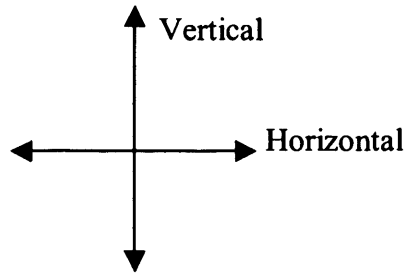
Appendix 1c: Instructions for the Single Word Norming Procedure

Instructions

In this task you will be given single words: e.g. *stroll*

We want you to match each of these words with any of 4 arrows that represents the event that the word describes.

There are **two axes** along which the event can take place: vertical and horizontal. There are **two directions** along each axes.



Please select the arrows that you feel **best represent** the event described by the word, with **your order of preference**.

Use 1 for the most preferred arrow, 2 for the next preferred and so on.

Please select a **minimum of one** picture for each word.

Select as many arrows as you think are appropriate (maximum of 4). You do not have to choose all four.









Write your choices in the box in the top left of each picture (please see examples below)

Please **look at all the arrows** before you make each of your choices.

Please take as much time as you like for each word.

If you do not understand the word then please write so on the word and move on to the next one. Your responses will be kept anonymous and confidential. If you have any questions please do not hesitate to ask the experimenter, you are free to leave at any point without penalty.

Examples:

stroll	<input type="checkbox"/>	<input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>
				
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
write	3 <input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	<input type="checkbox"/>
				
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Appendix 1d: Nonsense words used in lexical decision experiments

Non-word	Non-word	Non-word
ashed	ninked	treared
blirps	pauve	trieled
bloans	pell	trourts
blupped	phimes	twald
boids	phlep	tweighed
braifs	phoasts	twirque
caitch	phresk	twoob
chawled	phrimpse	twosts
cheel	phrirl	vacts
chirms	phriths	varred
chirts	phruth	veed
cleiced	phurp	vorl
cleige	phutched	whoats
cliel	plauce	yaide
clurped	plird	yoans
coost	plom	yoards
demps	plymed	yutt
dilge	prouns	
dim	quirts	
drimes	rhelts	
droafed	scrot	
fralled	sculps	
freeped	shrals	
fryst	sket	
gheeds	skiest	
ghirch	smoached	
glead	smuck	
glends	snawl	
glooks	sparch	
gonds	splabe	
grarked	splirmed	
groid	spriece	
himes	sproal	
janced	sprork	
kafed	stelk	
kelm	stimb	
kessed	swaive	
knades	swerd	
lailed	swoarse	
leezed	theint	
loff	threbbs	
loked	thun	
lorn	trafe	

Appendix 1e: Nonsense sentences used in sensibility judgments

Nonsense Sentence	Nonsense Sentence
the actor exited the tenants	the puppy retrieved bus-stops
the architect created a spoonful	the queen approved of the eyelash
the artist sculpted a weather	the restaurant provided the blush
the author wrote another moon	the rope fastened to the dream
the baby grabbed the smoke	the runner competed in the cake
the bailiff evicted stage left	the ship endured a child
the bank invested in the grin	the show amused the light bulb
the box-office held on the hillside	the siren warned us the gift
the boy received the elbows	the stereo ejected the giraffe
the bride married the bet	the surgeon prepared the lawnmower
the captain punished the candle	the tanks withdrew from the brat
the chef baked a phone	the teacher confiscated audience
the chemicals reacted with knee	the team explored in a glove
the clerk organised pig	the tutor instructed the bush
the committee suggested a duck	the war destroyed the ant
the council acquired the cloud	the woman accepted of danger
the couple adopted the storm	the woman decorated in mice
the court compensated the bicycle	
the doctor injected the traffic	
the family inherited in the garden	
the fireman breathed in the toy	
the girl learned from the eye	
the grandmother sat in the idea	
the group celebrated the comma	
the group enjoyed the ear	
the house rested the tickets	
the hypnotist mesmerized the magazine	
the instructions explained all month	
the kidnappers threatened the bark	
the laughter infected the jam-jar	
the library lent the breakfast	
the lorry blocked the insulin	
the make-up transformed the canyon	
the man argued with the door	
the mother spoke the traitor	
the noise alarmed to his health	
the officer suspended the bash	
the party drank the dog	
the patient coughed the boxer	
the photograph disappointed the carve	
the pond stagnated a fortune	
the program installed the chicken	
the pupil borrowed the elephants	

Filler Word	Associated Target	Unassociated Target
add	maths	bank
agree	compatible	wood
amaze	wonder	change
analyse	inspect	flower
argue	debate	wash
ask	question	change
bake	oven	advise
balloon	gas	remove
bathe	wash	plan
blossom	flower	remember
bring	carry	say
carve	wood	memory
celebrate	birthday	recommend
compute	calculate	oven
confuse	puzzle	chair
connect	join	describe
consider	think	boot
construct	build	borrow
convict	crime	create
delete	erase	adventure
deliver	package	question
destroy	damage	secure
disappoint	let-down	breathe
embarrass	blush	maths
enjoy	fun	give
explain	describe	shrink
explore	adventure	god
fasten	secure	damage
forget	memory	carry
frighten	afraid	wonder
illustrate	draw	bark
infect	germ	knowledge
inspect	examine	join
invent	create	insult
invest	bank	fun
learn	knowledge	let-down
lend	borrow	blush
marry	wedding	fake
obey	command	examine
offend	insult	afraid
organise	plan	erase
pretend	fake	discipline
provide	give	inspect

Filler Word	Associated Target	Unassociated Target
react	respond	minus
receive	gift	germ
record	album	key
reduce	shrink	birthday
refuse	deny	build
remind	remember	puzzle
separate	apart	lesson
sit	chair	gift
subtract	minus	funny
suggest	advise	deny
teach	lesson	crime
tell	say	calculate
transform	change	essay
unlock	key	debate
withdraw	remove	compatible
worship	god	apart
write	essay	wedding

Appendix 1f: Filler words and targets used for active comprehension questions.

Appendix 1g: Filler sentences and questions used in active comprehension experiments

Filler Sentence	Question with negative response
The motorists abandoned their cars	the motorists abandon their lorries
The woman accepted the gift	the woman return the gift
The couple adopted a child	the couple adopt a dog
The noise alarmed the dog	the noise alarm the cattle
The king banished the traitor	the king banish the prince
The lorry blocked the traffic	the lorry block the drive-way
The firemen breathed in the smoke	the firemen breathe in the chemicals
The teacher confiscated the magazine	the teacher confiscate the bubble-gum
The patient coughed all night	the doctor cough all night
The girl craved chocolate cake	the girl hate chocolate cake
The prisoner denied the charges	the prisoner accept the charges
The man deserted his wife	the man desert his job
The party drank to his health	the party drink until dawn
The ship endured the storm	the ship sink in the storm
The bailiff evicted the tenants	the bailiff collect the money
The actor exited stage left	the actor exit stage right
The student feared failure	the student fear spiders
The army fled the battle	the army charge into battle
The baby grabbed the toy	the baby grab the food
The box-office held the tickets	the box-office loose the tickets
The family inherited a fortune	the family inherit nothing
The doctor injected the insulin	the nurse inject the insulin
The hypnotist mesmerised his audience	the hypnotist mesmerise himself
The sportsmen obeyed the rules	the sportsmen disobey the rules
The drunk man offended his friends	the drunk man offend the lady
The boy promised to be good	the boy promise to brush his teeth
The jeering provoked the boxer	the jeering embarrass the boxer
The man regretted his behaviour	the man regret his haircut
The house rested on the hillside	the house rest on the river-bank
The criminal robbed the bank	the criminal rob the jewelers
The ghost scared the children	the ghost scare the vampire
The supermarket sold groceries	the supermarket sell clothes
The mother spoke to her child	the mother speak to her friend
The pond stagnated in the garden	the pond stagnate in the field
The cat stayed indoors in winter	the cat stay by the fire
The politician thanked his supporters	the politician thank his party
The fish tugged at the line	the fish swim into the net
The siren warned us of danger	the child warn us of danger
The film delighted the critics	the film disappoint the critics
The house depreciated in value	the house go up in value
The situation deteriorated quickly	the man's health deteriorate quickly
The crowd honored the athlete	the crowd cheer for the athlete
The men mined for diamonds	the men mine for gold
The student rejoiced at her results	the student cry at her results
The fans worshipped the pop-star	the fans hassle the pop-star

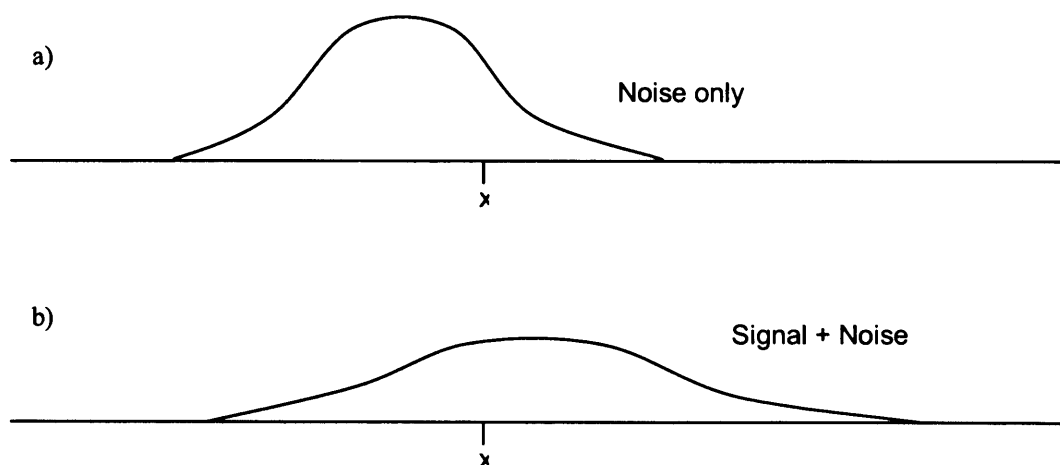
Appendix 2: Signal Detection Theory

This section provides a brief summary of signal detection theory and the calculations used in the present investigation. The information for this section was primarily taken from Wickens (2002), Stanislaw & Todorov (1999) and Gold, Sekuler & Bennett (2004). Graphs are adapted from Wickens (2002) unless otherwise stated.

Signal detection theory is used to explain how observers detect an event. It is based on the assumption that psychological (internal) responses are variable, so that a particular stimulus event has some probability of producing a particular internal response. This is in contrast to an ideal observer, for whom identical signal events will produce identical internal responses.

The probabilistic nature of detection is due to the fact that there is both internal and external noise, therefore a pure signal is never perceived. It is the internal noise that makes our observer non-optimal, since it means that identical external information will produce different internal responses. Thus, a particular signal event produces a distribution of internal responses as do non-signal (noise) events. The internal state of the observer is characterised by two distributions, one for the signal events and one for the noise events. Figure App2-1 presents the graphical illustration of signal and noise distributions.

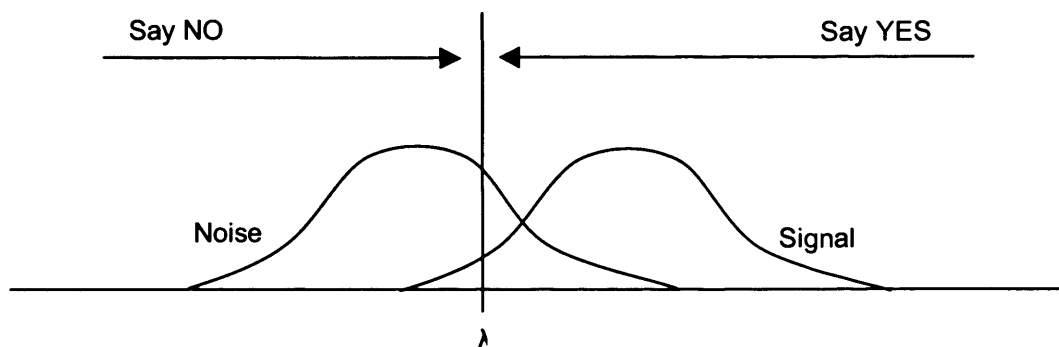
Figure App2-2: Probability distributions for (a) a noise event and (b) a signal event. The horizontal axis represents the internal evidence for a particular event, the x is an observation that could have come from either the signal or noise distributions.



In order to detect a particular signal event, the observer compares it to an internal criterion. If the event information exceeds the criterion, it is judged to be a signal. If the signal information falls below the criterion, it is judged to be noise (i.e. a non-signal event). The internal criterion is the level of evidence which acts as the decision threshold around which decisions are made. When the signal and noise distributions are both represented on the same axis (even though they are separate) along with some criterion value, the typical signal detection graph is produced, see Figure App2-2. Three assumptions are made in order to extract useful measures from this model: first, the signal and noise distributions are Gaussian (normal); second, the noise distribution is given a $\mu = 0$ and a $\sigma = 1$; third, the signal and noise distributions have equal variance. It is beyond the scope of this summary to detail why these assumptions are made, but they are necessary for useful parameter estimates (measures) to be made.

Figure App2-2: Signal and noise distributions from Figure App2-1 on a single axis, with a criterion at λ

If a particular event exceeds the criterion the observer will judge that it was a signal and response YES, if the event falls below the criterion the observer will respond NO, that the event was not a signal. In this graph the two distributions have the same variance; the equal-variance SDT model assumes equal variance in order to calculate measures of sensitivity and criterion (Wickens, 2002).



The strength of the signal information is given by the separation between the signal and noise distributions, being dependent on their position and distributions. The observer's sensitivity, which depends on the internal strength of the signal, is given by the ratio of the signal (the distance between the two distributions) to the noise (the standard deviation of the distributions). A stronger signal will mean that the signal distribution is shifted to the right, decreasing the overlap between the two distributions. The measure

of sensitivity, d' , describes the relationship of the signal and noise distributions to each other. If the value is close to 0, then the distributions overlap a great deal and there is little separation between the two, when the value is large the two distributions are well separated and the signal can easily be detected. The overlap between the distributions is independent from the level of the criterion (λ) which can be set at any point along the axis and reflects the observer's strategy. The higher the criterion, the more conservative the decision and the more likely the observer is to respond NO (since fewer events are judged to be signal); the lower the criterion, the more lenient the decision and the more likely the observer is to respond YES (as more events are judged to be signal). This is the real elegance of SDT, separating decision processes (criterion setting) from signal processing (sensitivity). There is one measure of sensitivity (d') and two measures of criterion (c and β).

Figure App2-3: The measures d' , c and Beta as they relate to the signal and noise distributions and the criterion λ

(Taken and adapted from Stanislaw & Todorov, 1999)

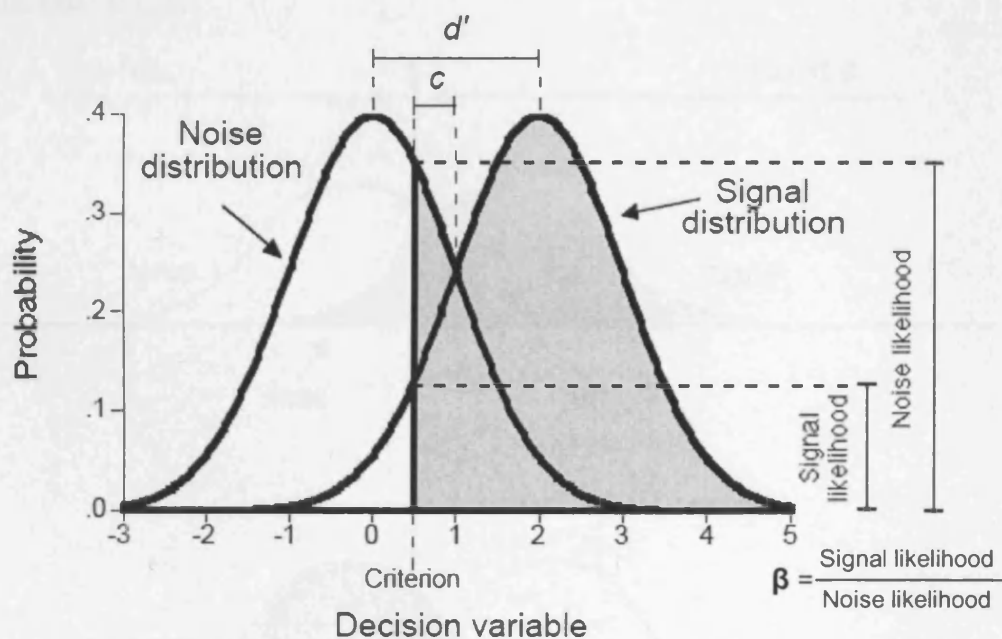


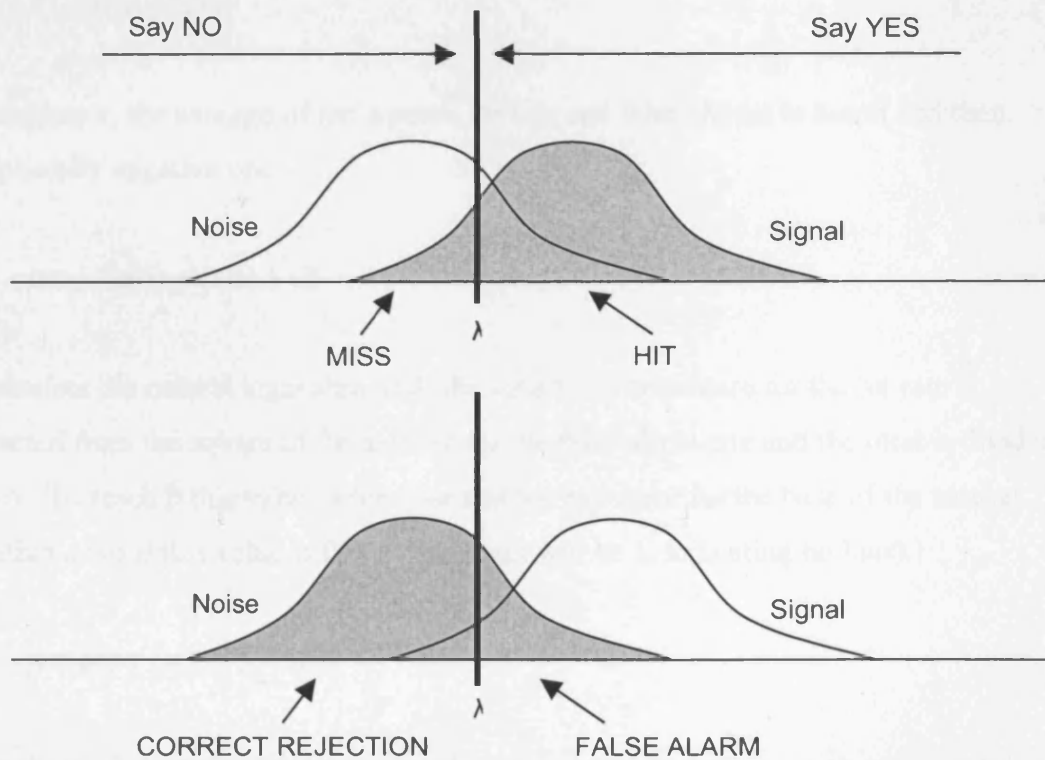
Figure App2-3 presents the three measures and their relation to the signal and noise distributions. As a measure of sensitivity, d' represents the distance between the centre of the signal and noise distributions, the larger the value, the greater the strength of the signal relative to the internal noise. Measures of criterion take two forms: c represents the distance of the criterion from the neutral point where the signal and noise distributions cross over (so there is an equal probability of both); values of c greater than 0 mean the criterion is above the neutral point and there is a bias to respond NO,

values of c less than 0 mean the criterion is below the neutral point and there is a bias to respond YES. β is the ratio of the signal likelihood (the probability of a hit given some λ) to the noise likelihood (the probability of a false alarm given some λ); values less than 1 indicate a bias to respond YES since the criterion is set lower, values more than 1 indicate a bias to respond NO since the criterion is set higher. Note that β is affected by changes in d' (i.e. the overlap between the distributions) whereas c is not.

Table App2-1: Response Types in a 2-alternative forced choice detection task

Event	Response	
	YES	NO
SIGNAL	HIT	MISS
NOISE	FALSE ALARM	CORRECT REJECTION

Figure App2-4: Four types of response as they relate to the signal and noise distributions and criterion



In order to calculate measures of sensitivity (d') and criterion λ (c and β), two response types are used. These are summarised in Table App2-1 and Figure App2-3; in a task

where the signal is either present or absent and the response to be made is YES (signal present) or NO (signal absent), four response conditions are possible. A HIT occurs when the signal is present and the response is YES; a FALSE ALARM (FA) occurs when the signal is absent and the response is YES; a MISS occurs when the signal is present and the response is NO and a CORRECT REJECTION occurs when the signal is absent and the response is NO. The frequencies of HIT and FA responses are converted to percentages and then z-scores so that they correspond to the areas in a Gaussian distribution (like the shaded areas in Figure App2-4). The HIT rate, $z(h)$, then gives the percentage of the signal distribution that falls above the criterion λ and the FA rate, $z(fa)$, gives the percentage of the noise distribution that falls above the criterion λ . Since these values can be interpreted using Gaussian distributions, the centre of the signal and noise distributions can be inferred.

To calculate d' , the z-score of the false alarm rate is subtracted from the z-score of the hit rate:

$$d' = z(h) - z(fa)$$

To calculate c , the average of the z-score for hits and false alarms is found and then multiplied by negative one:

$$c = - (z(h) + z(fa)) / 2$$

To calculate the natural logarithm of β , the square of the z-score for the hit rate is subtracted from the square of the z-score for the false alarm rate and the total is divided by two. To reach β this value is then used as the exponent for the base of the natural logarithm e (so if this value is 0 then the result will be 1, indicating no bias).

$$\beta = \exp ((z(fa)^2 - z(h)^2) / 2)$$

These are the calculations that were used in the present investigation. It is beyond the scope of this summary to detail how the calculations are derived, but it is hoped that enough information was provided so that the measures can be sensibly interpreted. For detailed discussion of these calculations see Stanislaw & Todorov (1999) and Wickens (2002).

Blocked perception of motion on single word and sentence comprehension

These two experiments adapted the procedure used by Kaschak et al (2005).

Participants were presented with single words for lexical decision (a) or sentences for sensibility judgements (b). They judged the linguistic stimuli in blocks during which they viewed a 30 second animation of black and white lines moving upwards or downwards. In line with the results from Kaschak et al (2005) the hypothesis is that when the semantic and visual motion are congruent, participants will be slower to judge the words or sentences; this is in comparison to the condition in which semantic and visual motion are incongruent, and should also be the case in comparison to the control condition. This should be the case for sentences (as for Kashcak et al, 2005) and should also extend to single words (based on the assumption that semantic representations are accessed during auditory lexical decision and that they are analogous to the semantic representation of sentences).

There was a two list design in Expeirment 1b as the experiment was completed at the Science Museum, therefore it needed to be short enough for visitors to the museum to complete without inconvenience to their visit.

Participants

a) 30 Native English speakers took part (22 female), the mean age was 21.8 years (SD = 4.6 years).

b) 66 Native English speakers took part (33 female), 35 completed List 1 and 31 completed List 2. The mean age was 35.6 years (standard deviation = 10.4 years).

Method

Design

a) A 3x2 repeated measures design was used. The independent variables were Word Category (Control, Up, Down) and Motion Direction (upwards, downwards). The dependent variable was the reaction time (ms) to complete and errors made in the lexical decision. There were a total of 12 blocks, half with the lines moving upwards and half with the lines moving downwards; line motion was alternated between successive blocks. Each of the blocks contained 18 trials composed of 9 experimental and 9 nonsense words. The order of presentation of the words within blocks was fully randomised.

b) A 2x3x2 mixed design was used. The independent variables were Experimental List (1,2), manipulated between subjects, Sentence Category (Control, Up, Down) and Motion Direction (upwards, downwards) manipulated within subjects. The dependent variable was the reaction time (ms) and errors for the sentence judgement. There were a total of 6 blocks, alternating between upwards and downwards motion. Each of the blocks contained 9 trials composed of 7 experimental and 2 nonsense sentences. The order of presentation of the sentences within blocks was fully randomised.

Materials

a) Item Set One Single Words was used. 54 nonsense words were selected from the set created for this purpose. The items were sorted into two lists, so that each word appeared twice, once with upwards line motion and once with downwards line motion. First, items were quasi-randomly sorted into 6 lists (one for each presentation block), so that each contained 9 nonsense words and 9 experimental words split equally between the three groups (Control, Down, Up). These lists comprised the first presentation of each item. The items were then quasi-randomly resorted into 6 other lists of the same composition; these comprised the second presentation of each sentence. Thus, across the two presentations items were randomly re-sorted so that the experimental items appeared with different accompanying items in the two presentation blocks they

appeared in. This was done to further randomise presentation order and prevent expectations arising due to consistent block items.

b) Item Set One Sentences was used. Two sentences were randomly removed from each condition (Control, Up, Down) to further shorten the length of the lists. Therefore there were 14 sentences in each category. 12 nonsense sentences were selected from the set created for this purpose. The experimental and nonsense items were randomly assigned to two lists so that each contained 7 Up, 7 Down, 7 Control and 6 nonsense sentences. Lists were organised so that each sentence appeared twice, once with upwards and once with downwards motion. Within each list, items were quasi-randomly sorted into three lists, so that each contained 2 nonsense sentences and 7 experimental sentence split as equally as possible between the three groups (3 from two groups and 2 from one group). These lists comprised the first presentation of each sentence. The items were then quasi-randomly resorted into three other lists of the same composition, these comprised the second presentation of each sentence.

Graphic Display

The screen displayed black and white lines moving up or down the screen. This was accomplished by presented a series of 8 bitmap files successively to create an animation of line motion. The display filled the 640 x 480 pixel screen. Black horizontal bars subtended 0.17° of visual angle separated by horizontal white bars subtending 0.34° ; each image was presented for 68ms and the lines moved at approximately $7.5^\circ/s$. There was a red fixation cross constantly presented in the centre of the screen (courier font, 18pt, bold).

Apparatus

The experiment was run using PC computers (Intel Pentium II) using E-Prime 1.0 Software (Schneider, Eschman, & Zuccolotto, 2002). A standard CRT monitor was used with a refresh rate of 60 Hz. A pair of Beyer Dynamic DT 100 headphones (2 x400 Ω) were used. Participants responded using standard 'qwertyuiop' keyboards. Chin rests were also used to minimise head movements, these were placed at a distance of 57cm from the computer screen.

Procedure

a) Participants were instructed to judge whether or not the words they heard had a meaning and were real words in English. They were instructed to press the smiley face if the word had a meaning and the frowning face if it did not. Stickers with a smiling and frowning face were placed on the L and A keys and balanced across participants. They were asked to focus on the red fixation cross in the centre of the screen whilst they were judging the words and to rest their eyes between blocks.

Participants completed 10 practice trials, 5 of which were non-words. In each block the motion animation was constant for 32.4 seconds. Individual words were played with an SOA of 1.8 seconds. At the end of each block, a grey screen was displayed for 10 seconds and participants were instructed to rest their eyes.

b) The procedure was identical to 8.1a except that instructions were modified for judging the sensibility of sentences. Participants were instructed to press the smiley face if the sentence made sense and the frowny face if it did not. They completed 10 practice trials, 4 of which were nonsense sentences. In each block sentences were played with an SOA of 3.7 seconds and the motion animation lasted 33.7 seconds.

Analysis

a) For each participant, trials on which the word had been judged incorrectly were removed and analysed separately as errors. Reaction times were measured from word onset so the length of each word was subtracted from the reaction time from to get the reaction time relative to word offset (giving an RT of 0 for a response at word offset). Reaction times below -500 milliseconds and above 1000 milliseconds were excluded, removing 2.67% of the data.

b) Trials on which the sentence had been judged incorrectly were removed and analysed separately as errors. Reaction times were calculated in the same manner as for Experiment a. Following this, reaction times below -300 milliseconds and above 1200 milliseconds were excluded, removing 2.1% of the data overall (2.8% from List 1 and 1.3% from List 2).

Experimental conditions were collapsed into Control, Match and Mismatch conditions and analysed with 8.1a) 3x2 repeated measures ANOVA by Subjects and 3x2 mixed

effects ANOVA by Items; 8.1b) A 2x3x2 mixed effects ANOVA by Subjects. Follow up 3x2 repeated measures ANOVAs were used to compare conditions within each List. A by items analysis was not performed as within each list there were only 7 items per condition. We judged this was too few to see any reliable effects.

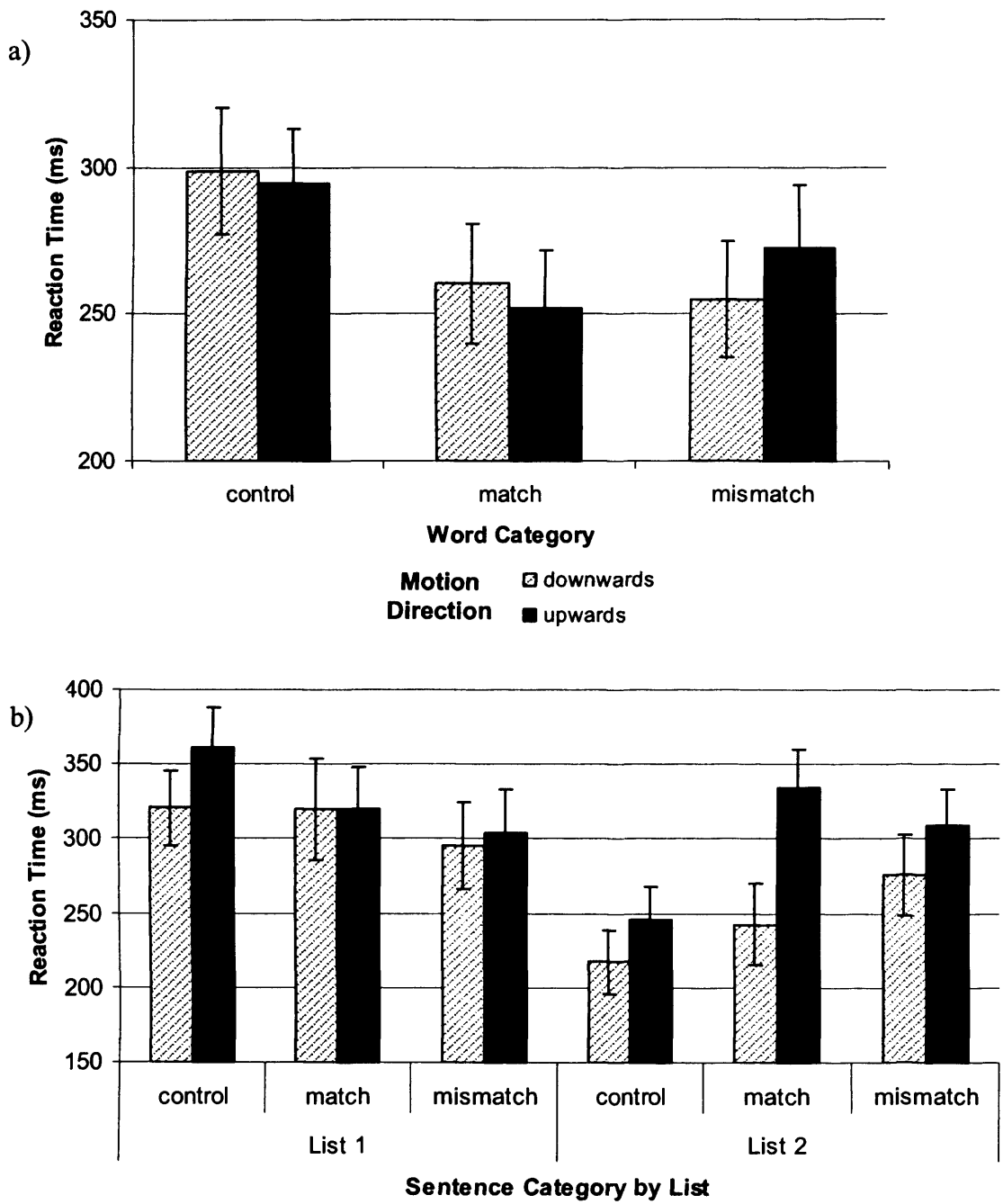
Results

Reaction Times

a) There was a main effect of Word Category ($F(2,58) = 13.411$, $p = 0.000$; $F(2,51) < 1.1$), driven by significant differences between the Control condition and the Match ($F(1,29) = 20.638$, $p = 0.000$; $F(1,34) < 1.6$) and Mismatch ($F(1,29) = 13.913$, $p = 0.001$; $F(1,34) = 1.2$) conditions. The Control condition was 40ms slower than the Match condition and 33ms slower than the Mismatch condition. See Table App-3 and Figure App-3a.

b) Across both lists, there was a significant main effect of Motion direction ($F(1,64) = 12.995$, $p = 0.001$), with Downwards motion 33.61ms faster than Upwards motion. There was also a marginal interaction between Motion Direction and List ($F(1,64) = 3.575$, $p = 0.063$, $p = 0.053$). In List 1 the main effect of motion was not significant ($F < 1.2$) whereas in List 2 the main effect of motion was significant ($F(1,29) = 21.787$, $p = 0.000$). There was also a significant interaction between Sentence Category and List ($F(2,128) = 12.979$, $p = 0.000$). For List 1, the main effect of Sentence Category was significant ($F(2,68) = 3.972$, $p = 0.025$) with the Control reaction times 29ms longer than Match ($F(1,34) = 4.560$, $p = 0.034$) and 33ms longer than Mismatch ($F(1,34) = 7.585$, $p = 0.008$) conditions. For List 2 the main effect of Sentence Category was also significant ($F(2,58) = 7.716$, $p = 0.001$), however here the Control reaction times were 56ms shorter than Match ($F(1,30) = 13.862$, $p = 0.000$) and 57ms shorter than Mismatch ($F(1,30) = 15.316$, $p = 0.000$) conditions.

Figure App-3: Reaction Times to judge stimuli for (a) Lexical Decision and (b) Sentence sensibility (by List)



Finally, the 3 way interaction between List, Sentence Category and Motion Direction was significant ($F(2,128) = 5.566$, $p = 0.080$). There was a significant interaction between Sentence Category and Motion direction for List 2 ($F(2,58) = 5.507$, $p = 0.160$) but not for List 1 ($F < 1.9$). For List 2, there were interactions when the Match condition was compared to the Mismatch ($F(1,30) = 6.255$, $p = 0.173$) and Control ($F(1,30) = 11.441$, $p = 0.276$) conditions. For the Match condition in List 2 Upwards

motion was slower than Downwards motion by 93.4 ms, but for the Mismatch and Control conditions Downwards motion was faster than Upwards motion by 26.8ms 31.9ms respectively.²⁹ Figure App-3b illustrates the different patterns in reaction times across List 1 and 2 for each Sentence Category and Motion Direction (see also Table App-3).

Errors

a) Friedman tests compared the errors for each condition. There was a trend by Subjects for more errors in the Control condition (mean = 2.14) compared to the Match (mean = 1.52) and Mismatch (mean = 1.76) conditions ($X^2_1 = 5.396$, $df = 2$, $p = 0.067$; $X^2_2 < 1$; by Items Mann Whitney Test, all $z < 1$).

b) No significant differences were found.

Table App-3: Mean response times for Lexical Decision in Experiment 8.1a and 8.2a

(Standard deviation in brackets)

Experiment	Condition					
	Control		Match		Mismatch	
	Downwards	Upwards	Downwards	Upwards	Downwards	Upwards
Exp 8.1a						
Blocked line motion	298.67 (117.88)	294.17 (103.04)	260.27 (111.11)	251.85 (110.35)	255.06 (108.49)	272.34 (117.23)
<i>Mean</i>	296.42 (109.79)*		256.06 (109.87)		263.70 (112.32)	
Exp 8.2a						
Trial dot motion	417.80 (94.09)	399.29 (113.41)	394.59 (86.81)	353.54 (103.67)	376.91 (119.34)	404.08 (103.29)
<i>Mean</i>	408.54 (103.76)*		374.07 (97.06)		390.50 (111.53)	

*Significantly different to other two conditions, $p < 0.05$

²⁹ When each List was analysed in a 3x2 ANOVA with Word Category (Control, Up, Down) and Motion Direction (Downwards, Upwards), List 1 showed a main effect of Word Category ($F(2,68) = 3.972$, $peta = 0.105$), with reaction times for Control sentences a mean of 31ms longer than Down and Up sentences. List 2 also showed a main effect of Word Category ($F(2,60) = 15.686$, $peta = 0.343$), with all Word Categories significantly different from each other. Shortest reaction times were present for Control (231.545), then Down (275.932) sentences with longest times for Up sentences (305.353). When collapsed into Match and Mismatch conditions, this produces the interaction seen above for List 2.

Discussion

Reaction times for the lexical decision showed significant differences between the Control words (which do not refer to motion) and directional words (congruent or incongruent to the direction of visual motion). Participants took more time (and had a tendency to make more errors) on Control words as compared to Down and Up words, regardless of the visual motion direction. It is possible that motion words were generally easier to comprehend in the presence of visual motion, whereas non-directional words were not. Under this explanation, the visual motion primes semantic access for motion words because motion processing is already active so accessing semantic motion information is easier (if the two are grounded in the same system). However, if this is the case then it is not clear why congruence between semantic and visual motion would not make a further difference to reaction times. For example, in line with Kaschak et al (2005), we could expect congruent conditions to have longer reaction times as semantic and visual motion compete for the same directionally selective resources. There are a number of reasons that could contribute to this lack of congruency effects. First, it is possible that uncontrolled for lexical variables separate the Control from the Down and Up words. These groups were matched on frequency and length only. If the Control words are more familiar, more concrete or have smaller phonological neighbourhoods, reaction times to make lexical decision would be faster as compared to the Down and Up words. Second, the directional words were comprised of both concrete (e.g. rise) and abstract (e.g. hope) items. The more abstract items may increase the variance in response times and may not even be influenced by visual motion; therefore any effects of congruence are too weak to show through in reaction times.

Similar patterns were found for sentence judgements, with Control sentences behaving differently to the directional sentences regardless of congruence; however, the two Lists showed this pattern in opposite directions. For List 1, the Control sentences took longer to judge than Down or Up sentences. For List 2, Control sentences were judged more quickly than Down sentences, which were also judged more quickly than Up sentences (therefore showing a general difference between the Down and Up sentences as well as the difference to Control sentences). No effect of congruence was found so the results from Kaschak et al (2005) have not been replicated and the hypothesis is not supported. As for the single words, uncontrolled lexical/sentential differences between Sentence

Categories could have produced the different reaction times (e.g. concrete and abstract events and familiarity with subject, verb and object combinations). This is particularly likely as the two Lists showed opposite patterns between the Control and directional sentences. For example, in List 1 Control sentences may have been harder to comprehend than directional sentences (because of the particular combination of subject, verb and object) whereas the opposite may have been true for List 2. In addition, in List 2 Down sentences were responded to significantly faster than Up sentences, lending more weight to the argument that Control, Down and Up sentences differed in ways that affected their ease of comprehension. Thus, the difference in reaction time patterns between Lists means that no inferences can be made about the influence of motion direction on comprehension. In addition, it makes similar inferences unreliable for the single word data. Aside from the lack of congruency effects, Experiment 8.1b did show a significant main effect of motion direction, with responses faster for downwards motion as compared to upwards motion. This is in line with results from Chapter 6 and 7 in which differences between downwards and upwards motion are seen, such that downwards motion is more easy to perceive. Here, we see that upwards motion was more interfering across the experiment than downwards motion, further supporting the increased difficulty in viewing/perceiving upwards motion. In fact, some participants did report that the upwards motion made them feel slightly disorientated. A second problem with blocked perception of visual motion is the Motion After Effect (MAE) in which prolonged viewing of one direction of visual motion (30 seconds and over) leads to illusory motion in the opposite direction when the visual stimulus is removed. As noted by Kaschak et al (2005), black and white line motion animations were of a long enough duration and high enough contrast to produce the MAE. Although we did not find any effect of congruence, so inferences about the influence of visual or MAE motion on comprehension are void, it is possible that the MAE discouraged participants to fixate properly. For both experiments it is possible that participants did not fixate on the motion animation (either by averting their gaze or closing their eyes), thereby negating the key manipulation. Participants did place their chin on a chin rest to control head movements and encourage fixation; in addition enforced breaks of 10 seconds were introduced so that fixation was maintained during each block. However, it is still possible that participants did not consistently fixate. This is especially true in the less controlled environment of the Science Museum, where participants could not be closely monitored.

These experiments failed to replicate and extend the original findings. This is most likely due to uncontrolled differences between linguistic items in each condition and prolonged viewing of strong visual motion that may have discouraged fixation. Therefore, Experiments 8.1a/b in Chapter 8 used different sets of linguistic stimuli (see Chapter 4) and presented motion on a trial by trial basis, rather than in blocks.

- Aziz-Zadeh, L., Wilson, S. M., Rizzolatti, G., & Iacoboni, M. (2006). Congruent Embodied Representations for Visually Presented Actions and Linguistic Phrases Describing Actions. *Current Biology*, 16, 1818-1823.
- Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX Lexical Database* [CD-ROM]
- Baddeley, A. (1999). Knowledge (Ch.13) in *Human Memory: Theory and Practice* (Revised Edition) (pp. 229-255). Hove, UK.: Psychology Press.
- Barsalou, L. W. (1999). Perceptual symbol systems. *Brain and Behavioural Sciences*, 22, 577-660.
- Barsalou, L. W., Kyle Simmons, W., Barbey, A. K., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences*, 7 (2), 84-91.
- Beckers, G., & Zeki, S. (1995). The consequences of inactivating areas V1 and V5 on visual motion perception. *Brain*, 118, 49-60.
- Beilock, S. L., & Holt, L. E. (2007). Embodied Preference Judgements: Can Likeability Be Driven by the Motor System? *Psychological Science*, 18(1), 51-57.
- Bergen, B. K., Lindsay, S., Matlock, T., & Narayanan, S. (2007). Spatial and linguistic aspects of visual imagery in sentence comprehension. *Cognitive Science* 31 (5), 733-764
- Blake, R., Sekuler, R., & Grossman, E. (2004). Motion processing in the human visual cortex. In J. H. Kaas & C. E. Collins (Eds.), *The primate visual system* (pp. 311-344). Boca Raton, Florida: CRC Press.
- Blaser, E., Sperling, G., & Lu, Z.-L. (1999). Measuring the amplification of attention. *Proceedings of the National Academy of Sciences*, 96, 11681-11686.
- Bock, K., & Griffin, Z. M. (2000). Producing words: How mind meets mouth. In L. Wheeldon (Ed.), *Aspects of language production* (pp. 7-47). Hove, UK.: Psychology Press.
- Borghi, A. M., Glenberg, A. M., & Kaschak, M. P. (2004). Putting words in perspective. *Memory & Cognition*, 32(6), 863-873.
- Boroditsky, L. (2000) Metaphoric structuring: understanding time through spatial metaphors. *Cognition* 75, 1-28.
- Boroditsky, L. (2001) Does Language Shape Thought?: Mandarin and English Speakers' Conceptions of Time. *Cognitive Psychology* 43, 1-22.
- Borreggine, K. L., & Kaschak, M. P. (2006). The Action-Sentence Compatibility Effect: It's All in the Timing. *Cognitive Science*, 30, 1097-1112.
- Bosbach, S., Prinz, W., & Kerzel, D. (2004). A Simon Effect With Stationary Moving Stimuli. *Journal of Experimental Psychology: Human Perception and Performance*, 30(1), 39-55.
- Boulenger, V., Roy, A. C., Paulignan, Y., Deprez, V., Jeannerod, M., & Nazir, T. A. (2006). Cross-talk between Language Processes and Overt Motor Behaviour in the First 200msec of Processing. *Journal of Cognitive Neuroscience*, 18(10), 1607-1615.
- Braddick, O. (1973). The masking of apparent motion in random-dot patterns. *Vision Research*, 13 (2), 355-369.
- Braddick, O. (1980). Low-level and high-level processes in apparent motion. *Philosophical Transactions of the Royal Society of London B Biological Sciences*, 290 (1038), 137-151.
- Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1993). Response of neurons in macaque MT to stochastic motion signals. *Visual Neuroscience*, 10(6), 1157-1169.
- Buccino, G., Riggio, L., Melli, G., Binkofski, F., Gallese, V., & Rizzolatti, G. (2005). Listening to action-related sentences modulates the activity of the motor system: A combined TMS and behavioural study. *Cognitive Brain Research*, 24, 355-363.
- Chambers, C. G., Tanenhaus, M. K., Eberhard, K. M., Filip, H., & Carlson, G. N. (2002). Circumscribing Referential Domains during Real-Time Language Comprehension. *Journal of Memory and Language*, 47, 30-49.
- Chambers, C. G., Tanenhaus, M. K., & Madden, C. J. (2004). Actions and Affordances in Syntactic Ambiguity Resolution. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 30(3), 687-696.
- Chao, L. L., Haxby, J. V., & Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature Neuroscience*, 2(913-919).

- Chao, L. L., & Martin, A. (2000). Representation of Manipulable Man-Made Objects in the Dorsal Stream. *NeuroImage*, 12, 478-484.
- Chater, N., & Vitanyi, P. (2003). Simplicity: a unifying principle in cognitive science? *Trends in Cognitive Sciences*, 7(1), 19-22.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, USA.: MIT Press.
- Clark, A. (1997). *Being There: Putting Brain, Body and World Together Again* (paperback (1998) ed.). London, UK.: MIT Press.
- Clark, A. (1999). An embodied cognitive science? *Trends in Cognitive Sciences*, 3(9), 345.
- Clark, A., Bechtel, W., & Graham, G. (1998). Embodied, situated, and distributed cognition. In *A companion to cognitive science* (pp. 506). Malden, MA.: Blackwell.
- Collins, A. M., & Loftus, E. F. (1975). A Spreading-Activation Theory of Semantic Processing. *Psychological Review*, 82(6), 407-428.
- Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology*, 33A, 497.
- Connell, L. (2006). Representing object colour in language comprehension. *Cognition*, 102(3), 476-485.
- Dahan, D., & Tanenhaus, M. K. (2005). Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. *Psychonomic bulletin and review*, 12(3), 453-459.
- Damasio, H., Grabowski, T. J., Tranel, D., Ponto, L. L. B., Hichwa, R. D., & Damasio, A. R. (2001). Neural Correlates of Naming Actions and of Naming Spatial Relations. *NeuroImage*, 13, 1053-1064.
- Damasio, H., Tranel, D., Grabowski, T. J., Adolphs, R., & Damasio, A. R. (2004). Neural Systems behind word and concept retrieval. *Cognition*, 92, 179-229.
- Eckstein, M.P., Beutter, B.R., Pham, B.T., Shimozaki, S.S. & Stone, L.S. (2007) Similar neural representations of the target for saccades and perception during search. *Journal of Neuroscience*, 27(15), 3935-3936.
- Elman, J. L. (2004). An alternative view of the mental lexicon. *Trends in Cognitive Sciences*, 8(7), 301-306.
- Eriksen, B.A. & Eriksen, C.W. (1974) Effects of noise letters upon the identification of a target letter in a non-search task. *Perception and Psychophysics* 16, 143-149.
- Estes, Z., Verges, M., & Barsalou, L. W. (in press) Head up, foot down: Object words orient attention to their typical location. *Psychological Science*.
- Farah, M. J., & McClelland, J. L. (1991). A Computational Model of Semantic Impairment: Modality Specificity and Emergent Category Specificity. *Journal of Experimental Psychology: General*, 120(4), 339-357.
- Fincher-Kiefer, R. (2001). Perceptual components of situation models. *Memory & Cognition*, 29(2), 336-343.
- Fischer, M. H., & Zwaan, R. (in press). Embodied Language - A Review of the Role of the Motor System In Language Comprehension. *Quarterly Journal of Experimental Psychology*.
- Fodor, J. A. (1983). *The Modularity of Mind: An Essay on Faculty Psychology*. London: MIT Press.
- Fodor, J. A. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. London: MIT Press.
- Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments & Computers*, 35, 116.
- Frith, C. (2007). *Making up the mind: How the Brain Creates our Mental World*. Oxford, UK.: Blackwell.
- Gallese, V., & Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology*, 22(3/4), 455.
- Garrett, M. F. (1975) The Analysis of Sentence Production. In G. H. Gower (Ed) *The Psychology of Learning and Motivation, Volume 9*. Academic Press, New York.
- Gerlach, C., Law, I., & Paulson, O. B. (2002). When Action Turns to Words. Activation of Motor-Based Knowledge during Categorisation of Manipulable Objects. *Journal of Cognitive Neuroscience*, 14(8), 1230-1239.

- Gibson, J. J. (1986). *The Ecological Approach to Visual Perception*. Hillsdale, NJ, USA.: Lawrence Erlbaum Associates.
- Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *Proceedings of the National Academy of Sciences*, 103(2), 489-494.
- Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (in press 2007). Support for lateralization of the Whorf effect beyond the realm of color discrimination. *Brain and Language*.
- Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic bulletin and review*, 3(9), 558-565.
- Glenberg, A. M., & Kaschak, M. P. (2003). The body's contribution to language. In B.H.Ross (Ed.), *The Psychology of Learning and Motivation* (Vol. 43, pp. 93-126). San Diego, CA: Academic Press.
- Glenberg, A. M., & Robertson, D. A. (2000). Symbol Grounding and Meaning: A Comparison of High-Dimensional and Embodied Theories of Meaning. *Journal of Memory and Language*, 43, 379-401.
- Gold, J. M., Sekuler, A. B., & Bennett, P. J. (2004). Characterising perceptual learning with external noise. *Cognitive Science*, 28, 167-207.
- Goldinger, S. D. (1996). Auditory Lexical Decision. *Language and Cognitive Processes*, 11(6), 559-567.
- Goldstone, R. L., & Barsalou, L. W. (1998). Reuniting perception and conception. *Cognition*, 65, 231.
- Grabowski, T. J., Damasio, H., & Damasio, A. R. (1998). Premotor and prefrontal correlates of category-related lexical retrieval. *NeuroImage*, 7, 232-243.
- Halgren, E., Dhond, P. R., Christensen, N., Van Petten, C., Marinkovic, K., Lewine, J. D., et al. (2002). N400-like MEG Responses Modulated by Semantic Context, Word Frequency, and Lexical Class in Sentences. *NeuroImage*, 17(9), 1101-1116.
- Harley, T.A. (1995) *The Psychology of Language: From Data to Theory*. Erlbaum (UK) Taylor & Francis.
- Harnad, S. (1990). The Symbol Grounding Problem. *Physica*, D 42, 335.
- Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic Representation of Actions Words in Human Motor and Premotor Cortex. *Neuron*, 41(301-307).
- Hawkins, H. L., Hillyard, S. A., Luck, S. J., Mouloua, M., Downing, C. J., & Woodward, D. P. (1990). Visual Attention Modulates Signal Detectability. *Journal of Experimental Psychology: Human Perception and Performance*, 16(4), 802-811.
- Henderson, J. M., & Ferreira, F. (Eds.). (2004). *The Interface of Language, Vision, and Action: Eye Movements and the Visual World*. London, UK: Routledge.
- Hesslow, G. (2002). Conscious thought as simulation of behaviour and perception. *Trends in Cognitive Sciences*, 6, 242.
- Holcomb, P. J. (1993). Semantic priming and stimulus degradation: Implications for the role of the N400 in language processing. *Psychophysiology*, 30, 47-61.
- Horton, W. S., & Rapp, D. N. (2003). Out of sight, out of mind: Occlusion and the accessibility of information in narrative comprehension. *Psychonomic bulletin and review*, 10(1), 104-110.
- Humphreys, G. W., Price, C. J., & Riddoch, M. J. (1999). From objects to names: A cognitive neuroscience approach. *Psychological Research*, 62, 118-130.
- Inman, D. (2000). *Parrallel Distributed Processing: An Introduction*. Retrieved January 23, 2007, from <http://www.scism.sbu.ac.uk/inmandw/tutorials/pdp/pdpintro.html>
- Isenberg, N., Silbersweig, D., Engelien, A., Emmerich, S., Malavade, B., Beattie, B., et al. (1999). Linguistic threat activates the human amygdala. *Proceedings of the National Academy of Sciences*, 96, 10456-10459.
- Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution* (Large format paperback ed.): Oxford University Press.
- Jazayeri, M., & Movshon, J. A. (2007). *A new perceptual illusion reveals mechanisms of sensory decoding*. *Nature*, 446, 912-915.

- Jeannerod, M. (2001). *Neural Simulation of Action: A Unifying Mechanism for Motor Cognition*. *NeuroImage*, 14, S103-S109.
- Johnson-Laird, P. (1993). *The computer and the mind: An introduction to Cognitive Science*. Glasgow: Fontana Press.
- Kan, I. P., Barsalou, L. W., Solomon, K. O., Minor, J. K., & Thompson-Schill, S. L. (2003). Role of mental imagery in a property verification task: fMRI evidence for perceptual representations of conceptual knowledge. *Cognitive Neuropsychology*, 20, 525-540.
- Kaschak, M. P., Madden, C. J., Theriault, D. J., Yaxley, R. H., Aveyard, M., Blanchard, A. A., et al. (2005). Perception of motion affects language processing. *Cognition*, 94, B79.
- Kaschak, M. P., Zwaan, R., Aveyard, M., & Yaxley, R. H. (2006). Perception of Auditory Motion Affects Language Processing. *Cognitive Science*, 30(4), 733-744.
- Kastner, S., De Weerd, P., Pinsk, M. A., Elizondo, M. I., Desimone, R., & Ungerleider, L. G. (2001). Modulation of Sensory Suppression: Implications for Receptive Field Sizes in the Human Visual Cortex. *The Journal of Neurophysiology*, 86(3), 1398-1411.
- Kaup, B., Ludtke, J., & Zwaan, R. (2006). Processing negated sentences with contradictory predicates: Is a door that is not open mentally closed? *Journal of Pragmatics*, 38, 1033-1050.
- Kaup, B., & Zwaan, R. (2003). Effects of Negation and Situational Presence on the Accessibility of Text Information. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 29(3), 439-446.
- Kellenbach, M. L., Wijers, A. A., & Mulder, G. (2000). Visual semantic features are activated during the processing of concrete words: event-related potential evidence for perceptual semantic priming. *Cognitive Brain Research*, 10, 67-75.
- Kemmerer, D. (2005). The spatial and temporal meanings of English prepositions can be independently impaired. *Neuropsychologica*, 43(5), 797-806.
- Kitajo, K., Nozaki, D., Ward, L. M., & Yamamoto, T. (2003). Behavioural Stochastic Resonance within the Human Brain. *Physical Review Letters*, 90(21).
- Klatzky, R. L., Pellegrino, J. W., McCloskey, B. P., & Doherty, S. (1989). Can you squeeze a tomato? The role of motor representations in semantic sensibility judgements. *Journal of Memory and Language*, 28, 56-77.
- Kolers, P. A. (1983). Perception and Representation. *Annual Review of Psychology*, 34, 129-166.
- Kosslyn, S. M. (1996 (new edition)). *Image and Brain: The Resolution of the Imagery Debate*. London, England: MIT Press.
- Laeng, B., & Teodorescu, D.-S. (2002). Eye scanpaths during visual imagery re-enact those of perception of the same visual scene. *Cognitive Science*, 26, 207-231.
- Lakoff, G. (1987). *Women, Fire and Dangerous Things: What categories reveal about the mind*. Chicago, IL.: University of Chicago Press.
- Landauer, T. K., & Dumais, S. T. (1997). A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction and Representation of Knowledge. *Psychological Review*, 104(2), 211-240.
- Leibowitz, H. W., Johnson, C. A., & Isabelle, E. (1972). Peripheral Motion Detection and Refractive Error. *Science*, 177(4055), 1207.
- Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. London, England.: MIT Press.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1-75.
- Levinson, E. & Sekuler, R. (1975) Inhibition and disinhibition of direction-specific mechanisms in human vision. *Nature*, 254 (5502), 692-694.
- Levinson, E. & Sekuler, R. (1976) Adaptation alters perceived direction of motion. *Vision Research*, 16(7), 779-781.
- Levinson, E. & Sekuler, R. (1980) A two-dimensional analysis of direction-specific adaptation. *Vision Research*, 20(2), 103-107.

- Mahon, B. Z., & Caramazza, A. (2005). The orchestration of the sensory-motor systems: Clues from neuropsychology. *Cognitive Neuropsychology*, 22(3/4), 480-494.
- Malikovic, A., Amunts, K., Schleicher, A., Mohlberg, H., Eickhoff, S.B., Wilms, M., Palomero-Gallagher, N., Armstrong, E. & Zilles, K. (2007) Cytoarchitectonic Analysis of the Human Extrastriate Cortex in the Region of V5/MT+: A Probabilistic, Stereotaxic Map of Area hOc5. *Cerebral Cortex*, 17(3), 562-574.
- Markman, A. B., & Brendl, C. M. (2005). Constraining Theories of Embodied Cognition. *Psychological Science*, 16(1), 6-10.
- Markman, A. B., & Dietrich, E. (2000). Extending the classical view of representation. *Trends in Cognitive Sciences*, 4(12), 470-475.
- Marques, J. F. (2006). Specialization and semantic organisation: Evidence for multiple semantics linked to sensory modalities. *Memory & Cognition*, 34(1), 60-67.
- Martin, A., & Chao, L. L. (2001). Semantic memory and the brain: structure and processes. *Current Opinion in Neurobiology*, 11, 194-201.
- Martin, A., Haxby, J. V., Lalonde, F. M., Wiggs, C. L., & Ungerleider, L. G. (1995). Discrete Cortical Regions Associated with Knowledge of Color and Knowledge of Action. *Science*, 270(5233), 102-105.
- Martin, A., Wiggs, C. L., Ungerleider, L. G., & Haxby, J. V. (1996). Neural correlates of category-specific knowledge. *Nature*, 379, 649-652.
- Matlock, T., & Richardson, D. C. (2004). Do eye movements go with fictive motion? *Proceedings of the 26th Annual Conference of the Cognitive Science Society*, NJ, USA.
- McClelland, J. L., & Rogers, T. T. (2003). The Parallel Distributed Processing Approach to Semantic Cognition. *Nature Reviews Neuroscience*, 4, 310-322.
- McCloskey, B. P., Klatzky, R. L., & Pellegrino, J. W. (1992). Rubbing Your Stomach While Tapping Your Fingers: Interference Between Motor Planning and Semantic Judgements. *Journal of Experimental Psychology: Human Perception and Performance*, 18(4), 948-961.
- McNamara, T. P., & Miller, D. L. (1989). Attributes of Theories of Meaning. *Psychological Bulletin*, 106(3), 355-376.
- McRae, K., de Sa, V. R., & Seidenberg, M. S. (1997). On the Nature and Scope of Featural Representations of Word Meaning. *Journal of Experimental Psychology: General*, 126(2), 99-130.
- Meteyard, L., Bahrami, B. & Vigliocco, G. (2007) Motion detection and motion verbs: Language affects low-level visual perception. *Psychological Science*, 18 (11), 927-1021
- Moss, F., Ward, L. M., & Sannita, W. G. (2004). Stochastic resonance and sensory information processing: a tutorial and review of application. *Clinical Neurophysiology*, 115, 267-281.
- Murphy, G. L. (2002). *The Big Book of Concepts*. London, England.: MIT Press.
- Myung, J., Blumstein, S. E., & Sedivy, J. C. (2005). Playing on the typewriter, typing on the piano: manipulating knowledge of objects. *Cognition*, 98, 223-243.
- Neininger, B., & Pulvermuller, F. (2001). The Right Hemisphere's Role in Action Word Processing: a Double Case Study. *Neurocase*, 7, 303-317.
- Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998, 1998). *The University of South Florida word association, rhyme, and word fragment norms*. Retrieved 01/2005 from <http://www.usf.edu/FreeAssociation/>
- Newell, A. (1980). Physical Symbol Systems. *Cognitive Science*, 4, 135-183.
- Newell, A., & Simon, H. A. (1972). *Human Problem Solving*. Englewood Cliffs, New Jersey: Prentice-Hall Inc.
- Niedenthal, P. M., Barsalou, L. W., Winkielman, P., Krauth-Gruber, S., & Ric, F. (2005). Embodiment in Attitudes, Social Perception, and Emotion. *Personality and Social Psychology Review*, 9(3), 184-211.
- Ohtani, Y., & Ejima, Y. (1997). Anisotropy for direction discrimination in a two-frame apparent motion display. *Vision Research*, 37 (6), 765.

- Oliveri, M., Finocchiaro, C., Shapiro, K., Gangitano, M., Caramazza, A., & Pascual-Leone, A. (2004). All Talk and No Action: A Transcranial Magnetic Stimulation Study of Motor Cortex Activation during Action Word Production. *Journal of Cognitive Neuroscience*, 16(3), 374-381.
- O'Shea, J., Muggleton, N.G., Cowey, A., & Walsh, V. (2004) Timing of Target Discrimination in Human Frontal Eye Fields. *Journal of Cognitive Neuroscience*, 16, 1060-1066.
- Pecher, D., Zeelenberg, R., & Barsalou, L. W. (2003). Verifying Different-Modality Properties for Concepts Produces Switching Costs. *Psychological Science*, 14(2), 119-124.
- Pecher, D., Zeelenberg, R., & Barsalou, L. W. (2004). Sensorimotor simulations underlie conceptual representations: Modality-specific effects of prior activation. *Psychonomic bulletin and review*, 11, 164-167.
- Pecher, D., Zeelenberg, R., & Raaijmakers, J. G. W. (1998). Does Pizza Prime Coin? Perceptual Priming in Lexical Decision and Pronunciation. *Journal of Memory and Language*, 38, 401-418.
- Perky, C. W. (1910). An experimental study of imagination. *The American Journal of Psychology*, 21(3), 422-452.
- Phillips, J., Noppeney, U., Humphreys, G. W., & Price, C. J. (2002). Can segregation within the semantic system account for category-specific deficits? *Brain*, 125, 2067-2080.
- Pisoni, D., Hernandez, L., & colleagues. (2007). Speech & Hearing Lab Neighborhood Database. Retrieved 01/2005 from <http://128.252.27.56/Neighborhood/Home.asp>
- Posner, M.I. (1980) Orienting of attention. *Quarterly Journal of Experimental Psychology* 32, 3-25.
- Previc, F. H. (1990). Functional specialization in the upper and lower visual fields in humans: Its ecological origins and neurophysiological implications. *Behavioral and Brain Sciences*, 13, 519-575.
- Pulvermüller, F. (1999). Words in the brain's language. *Behavioral and Brain Sciences*, 22, 253-336.
- Pulvermüller, F. (2001). Brain reflections of words and their meaning. *Trends in Cognitive Sciences*, 5(12), 517-524.
- Pulvermüller, F. (2002). A brain perspective on language mechanisms: from discrete neuronal ensembles to serial order. *Progress in Neurobiology*, 67, 85-111.
- Pulvermüller, F., Harle, M., & Hummel, F. (2000). Neuropsychological distinction of verb categories. *NeuroReport*, 11(12), 2789.
- Pulvermüller, F., Harle, M., & Hummel, F. (2001). Walking or Talking? Behavioral and Neurophysiological Correlates of Action Verb Processing. *Brain and Language*, 78, 143-168.
- Pulvermüller, F., & Hauk, O. (2005). Category-specific Conceptual Processing of Color and Form in Left Frontotemporal Cortex. *Cerebral Cortex*, 16, 1193-1201.
- Pulvermüller, F., Hauk, O., Nikulin, V. V., & Ilmoniemi, R. J. (2005). Functional links between motor and language systems. *European Journal of Neuroscience*, 21, 793-797.
- Pulvermüller, F., Mohr, B., & Schleicher, H. (1999). Semantic of lexico-syntactic factors: what determines word-class specific activity in the human brain? *Neuroscience Letters*, 275, 81-84.
- Pylyshyn, Z. W. (1985). *Computation and Cognition: Toward a Foundation for Cognitive Science* (2nd ed.). London, England.: MIT Press.
- Quillian, M. R. (1968). Semantic Memory. In M. Minsky (Ed.), *Semantic Information Processing* (pp. 216-271). London, England.: MIT Press.
- Rastle, K., Harrington, J., & Coltheart, M. (2006). 358,534 nonwords: The ARC Nonword Database. *Quarterly Journal of Experimental Psychology*, 55A, 1339.
- Raymond, J. E. (2000). Attentional modulation of visual motion perception. *Trends in Cognitive Sciences*, 4(2), 42-50.
- Rees, G., Friston, K., & Koch, C. (2000). A direct quantitative relationship between the functional properties of human and macaque V5. *Nature Neuroscience*, 3(7), 716-723.
- Richardson, D. C., & Spivey, M. J. (2000). Representation, space and Hollywood Squares: looking at things that aren't there anymore. *Cognition*, 76, 269-295.

- Richardson, D. C., Spivey, M. J., Barsalou, L. W., & McRae, K. (2003). Spatial representations activated during real-time comprehension of verbs. *Cognitive Science*, 27, 767.
- Richardson, D. C., Spivey, M. J., & Cheung, J. (2001). Motor Representations in *Memory And Mental Models: Embodiment in Cognition*, NJ, USA.
- Richardson, D. C., Spivey, M. J., Edelman, S., & Naples, A. J. (2001). "Language is Spatial": Experimental Evidence for Image Schemas of Concrete and Abstract Verbs, *Proceedings of the 23rd annual meeting of the cognitive science society*. Mahwah, NJ: Erlbaum
- Rinck, M., & Bower, G. H. (2000). Temporal and spatial distance in situation models. *Memory & Cognition*, 28(8), 1310-1320.
- Rinck, M., Hahnel, A., Bower, G. H., & Glowalla, U. (1997). The metrics of spatial situation models. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 23, 622-637.
- Rips, L. J., Shoben, E. J. & Smith, E. E. (1973) Semantic Distance and the Verification of Semantic Relations. *Journal of Verbal Learning and Verbal Behavior*, 12, 1-20.
- Rogers, T. T., Lambon Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., et al. (2004). Structure and Deterioration of Semantic Memory: A Neuropsychological and Computational Investigation. *Psychological Review*, 111(1), 205-235.
- Rowe, J.B., Stephan, K.E., Friston, K., Frackowiak, R.S.J. & Passingham, R.E. (2004) The Prefrontal Cortex shows Context-specific Changes in Effective Connectivity to Motor or Visual Cortex during the Selection of Action or Colour. *Cerebral Cortex*, 15, 85-95.
- Rumelhart, D. E., McClelland, J. L., & Group, t. P. R. (1987). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition Volume 1: Foundations* (3rd ed.). London, England.: Bradford Book, MIT Press.
- Sack, A.T., Kohler, A., Linden, D.E.J., Goebel, R. & Muckli, L. (2006) The temporal characteristics of motion processing in hMT/V5+: Combining fMRI and neuronavigated TMS. *NeuroImage*, 29 (4), 1326-1335.
- Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime Reference Guide*. Pittsburgh, PA: Psychology Software Tools.
- Scorolli, C., & Borghi, A. M. (2007). Sentence comprehension and action: Effector specific modulation of the motor system. *Brain Research*, 1130, 119-124.
- Searle, J. (1980). Minds, Brains and Programs. *Behavioral and Brain Sciences*, 3, 417-457.
- Seitz, A. R., & Watanabe, T. (2003). Is subliminal learning really passive? *Nature*, 422(6927), 36-36.
- Seitz, A. R., & Watanabe, T. (2003). Is subliminal learning really passive? *Nature*, 422(6927), 36-36.
- Setola, P., & Reilly, R. G. (2005). Words in the brain's language: An experimental investigation. *Brain and Language*, 94(3), 251-259.
- Siakaluk, P. D., Pexman, P. M., Aguilera, L., Owen, W. J., & Sears, C. R. (in press). Evidence for the activation of sensorimotor information during visual word recognition: The body-object interaction effect. *Cognition*, doi:10.1016/j.cognition.2006.12.011.
- Silvanto, J., Lavie, N., & Walsh, V. (2005). Double dissociation of V1 and V5/MT activity in visual awareness. *Cerebral Cortex*, 15(11), 1736.
- Simon, H. A. (1979). Information Processing Models of Cognition. *Annual Review of Psychology*, 30, 363-396.
- Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and Process in Semantic Memory: A featural model for semantic decisions. *Psychological Review*, 81(3), 214-241.
- Solomon, K. O., & Barsalou, L. W. (2004). Perceptual Simulation in Property Verification. *Memory & Cognition*, 32, 244-259.
- Spence C., Nicholls M. E. R. & Driver J. (2001) The cost of expecting events in the wrong sensory modality. *Perception & Psychophysics*, 63(2), 330-336.
- Spivey, M. J., & Geng, J. J. (2001). Oculomotor mechanisms activated by imagery and memory: eye movements to absent objects. *Psychological Research*, 65(235-241).

- Spivey, M. J., Tyler, M. J., Richardson, D. C., & Young, E. E. (2000). Eye Movements during comprehension of Spoken Scene Descriptions, *Proceedings of the 23rd annual meeting of the cognitive science society*. Mahwah, NJ: Erlbaum
- Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science*, 12(2), 153-156.
- Stanislaw, H. & Todorov, N. (1999) Calculation of signal detection measures. *Behaviour Research Methods, Instruments & Computers*, 31(1), 137-149.
- Stillings, N. A., Feinstein, M. H., Garfield, J. L., Rissland, E. L., Rosenbaum, D. A., Weisler, S. F., et al. (1995). Cognitive Psychology: The Architecture of the Mind in *Cognitive Science An Introduction*. London, UK.: MIT Press.
- Suppes, P., Pavel, M., & Falmagne, J.-C. (1994). Representations and Models in Psychology. *Annual Review of Psychology*, 45, 517-544.
- Svensson, H., Lindblom, J., Ziemke, T., Zlatev, Ziemke, & Frank. (2004). Making sense of embodied cognition: Simulation Theories of Shared Neural Mechanisms for Sensorimotor and Cognitive Processes. in *Body, Language and Mind. Vol 1: Embodiment*.
- Talbot, J. B., Witton, C., McLean, M., Hansen, P. C., Rees, A., Green, G. G. R., et al. (2000). Dynamic sensory sensitivity and children's word decoding skills. *Proceedings of the National Academy of Sciences*, 97 (6), 2952.
- Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Scifo, P., et al. (2005). Listening to Action-related Sentences Activates Fronto-parietal Motor Circuits. *Journal of Cognitive Neuroscience*, 17(2), 273-281.
- Thomas, N. J. T. (1999). Are Theories of Imagery Theories of Imagination? An Active Perception Approach to Conscious Mental Content. *Cognitive Science*, 23(2), 207-245.
- Thompson-Schill, S. L., Bedney, M., & Goldberg, R. F. (2005). The frontal lobes and the regulation of mental activity. *Current Opinion in Neurobiology*, 15, 219-224.
- Tseng, M. J., & Bergen, B. K. (2005). Lexical Processing Drives Motor Simulation. *Proceedings of the 27th annual meeting of the cognitive science society*. Mahwah, NJ: Erlbaum
- Tsushima, Y., Sasaki, Y., & Watanabe, T. (2006). Greater Disruption Due to Failure of Inhibitory Control on an Ambiguous Distractor. *Science*, 314, 1786 - 1788.
- Tucker, M., & Ellis, R. (1998). On the relation between seen objects and components of potential actions. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 830.
- Tyler, L.K. & Moss, H.E. (2001) Towards a distributed account of conceptual knowledge. *Trends in Cognitive Science*, 5(6), 244-252.
- Vigliocco, G., & Vinson, D. P. (in press). Semantic Representation in G. Gaskell (Ed.), *Handbook of Psycholinguistics*. Oxford: Oxford University Press.
- Vigliocco, G., Vinson, D. P., Lewis, W., & Garrett, M. F. (2004). Representing the meaning of object and action words: The featural and unitary semantic space hypothesis. *Cognitive Psychology*, 48, 422-488.
- Vigliocco, G., Warren, J., Arcuili, J., Siri, S., Scott, S., & Wise, R. (2005). The role of semantics and grammatical class in the neural representation of words. *Cerebral Cortex*.
- Vinson, D. P., & Vigliocco, G. (2002). A semantic analysis of grammatical class impairments: semantic representations of object nouns, action nouns and action verbs. *Journal of Neurolinguistics*, 15(3), 317-351.
- Vitevich, M. S., & Luce, P. A. (2004). A Web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, & Computers*, 36(3), 481-487.
- Vitevich, M. S., Luce, P. A., Pisoni, D. B., & Auer, E. T. (1999). Phonotactics, Neighbourhood Activation and Lexical Access for Spoken Words. *Brain and Language*, 68, 306-311.
- Vogt, P. (2002). The physical symbol grounding problem. *Cognitive Systems Research*, 3, 429-457.
- Vuilleumier, P., Henson, R. N., Driver, J., & Dolan, R. J. (2002). Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming. *Nature Neuroscience*, 5(5), 491-499.
- Wakefield, J. C. (2003). The Chinese Room Argument Reconsidered: Essentialism, Indeterminacy, and Strong AI. *Minds and Machines*, 13, 285-319.

- Wallentin, M., Ellegaard, T., Ostergaard, S., Ostergaard, L., & Roepstorff, A. (2005). Motion verb sentences activate left posterior middle temporal cortex despite static context. *NeuroReport*, 16(6), 649-652.
- Watanabe, T., Nanez, J. E., & Sasaki, Y. (2001). Perceptual learning without perception. *Nature*, 413(6858), 844-848.
- Wickens, T. D. (2002). *Elementary Signal Detection Theory*. New York: Oxford University Press.
- Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin and Review*, 9 (4), 625.
- Yantis, S. (2005). How visual salience wins the battle for awareness. *Nature Neuroscience*, 8(8), 975-977.
- Zeki, S., Watson, J. D. G., Lueck, C. J., Friston, K. J., Kennard, C., & Frackowiak, R. S. J. (1991). A direct demonstration of functional specialisation in human visual cortex. *Journal of Neuroscience*, 11, 641-649.
- Ziegler, J. C., Muneaux, M., & Grainger, J. (2003). Neighbourhood effects in auditory word recognition: Phonological competition and orthographic facilitation. *Journal of Memory and Language*, 779-793.
- Ziemke, T. (2003). What's that Thing Called Embodiment, *Proceedings of the 25th Annual meeting of the Cognitive Science Society* NJ, USA.
- Zwaan, R. (1999). Situation Models: The Mental Leap Into Imagined Worlds. *Current Directions in Psychological Science*, 8(1), 15-18.
- Zwaan, R., & B.H.Ross. (2004). The immersed experiencer: toward an embodied theory of language comprehension in *The Psychology of Learning and Motivation* (Vol. 44, pp. 35). San Diego, CA.: Academic Press.
- Zwaan, R., Madden, C. J., Yaxley, R. H., & Aveyard, M. (2004). Moving words: Dynamic mental representations in language comprehension. *Cognitive Science*, 28, 611-619.
- Zwaan, R., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123, 162-185.
- Zwaan, R., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science*, 13(2), 168-171.
- Zwaan, R., & Taylor, L. J. (2006). Seeing, Acting, Understanding: Motor Resonance in Language Comprehension. *Journal of Experimental Psychology: General*, 135(1), 1-11.
- Zwaan, R., & Yaxley, R. H. (2003). Spatial iconicity affects semantic relatedness judgements. *Psychonomic Bulletin and Review*, 10(4), 954-958.