Non-Referring Concepts[*]

by
Sam Scott, B.Sc., M.C.S.

A thesis submitted to
the Faculty of Graduate Studies and Research
in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

Cognitive Science Program
Institute of Interdisciplinary Studies
Carleton University
Ottawa, Ontario

February 26, 2003

[*] Carleton University Cognitive Science Technical Report 2003-03
URL http://www.carleton.ca/iis/TechReports

The undersigned recommend to the
Faculty of Graduate Studies and Research
acceptance of the thesis

'NON-REFERRING CONCEPTS'

Submitted by SAMUEL SCOTT, M. C. S.

In partial fulfillment of the requirements for the degree of Doctor of Philosophy

Michel Gauthier

Chair of the Examining Board

Institute of Interdisciplinary Studies, Cognitive Science

Department

Thesis Supervisor

(JESSE PRINZ, UNC/Chapel Hill)

External Examiner (Ph.D.)

Carleton University
March 21, 2003

To Anne

Abstract

Non-referring concepts are mental representations of nonexistent things like dragons and time machines. Non-referring words are the words that express those concepts. In reference-based approaches to formal semantics, there is a well-known puzzle about how non-referring words and concepts get their meanings. The default solution is that they are special cases, with a different semantic structure from their referring counterparts. Despite over a century of debate in formal semantics, the issue of non-reference has, until now, been ignored in the psychological literature on concepts. But it is not obvious in advance what the psychological structure and processing of non-referring concepts will be like. Furthermore, experimental evidence about non-referring concepts can help resolve issues fundamental to both semantics and psychology, such as the nature of meaning, the nature of concepts, and the debate over representational externalism.

I used the methods of experimental psychology to provide the first empirical test of the claim that non-referring concepts are a special kind of concept. I found that non-referring concepts have a very similar structure to referring concepts. This finding suggests that many popular versions of reference-based semantics are flawed, perhaps fatally. I also found that non-referring concepts take measurably longer to process than referring concepts. I argue that the best way of reconciling the second result with the first is to make a distinction between the knowledge that is constitutive of a concept and the knowledge that is external to it. Non-referring concepts are processed more cautiously because we know that their referents do not exist – a fact that is about the world rather than the concept itself. This distinction is consistent with the apparently contradictory evidence that motivates "Theory Theory" accounts of conceptual structure on one hand, and similarity-based accounts such as prototype and exemplar theories on the other.

This dissertation concerns the following topic areas: cognitive psychology, cognitive science, concepts, empty names, externalism, internalism, mental representation, nonexistent objects, non-referring concepts, philosophy of language, philosophy of mind, prototypes, reference, semantics, similarity, and Theory Theory.

Acknowledgements

I am enormously indebted to my supervisor, Rob Stainton, for five years of unwavering support, encouragement, and guidance. The influence he has had on my intellectual maturation extends well beyond the scope of this dissertation.

I am also very grateful to Craig Leth-Steensen for detailed advice and assistance on the experimental portions of the dissertation. The statistical analyses in particular benefited tremendously from his expert guidance.

I am extremely thankful to Andy Brook, not only for his instruction, support and critical commentary, but also for being so central to the establishment of a rich and vibrant cognitive science program.

I would also like to thank Jo-Anne LeFevre for her keen interest in my work, her helpful comments, and her great assistance in the planning of further experimental work.

It was a real pleasure to have Jesse Prinz as my external examiner. His detailed questions and comments made the defense both challenging and enjoyable.

I thank Anne for being a great friend and loving partner, and also for many years of discussions about language, cognition, politics, and everything else. I've learned a lot from her.

I owe a great debt to the following instructors and supervisors, who helped guide me through my education in the cognitive sciences: Ken Barker, Martin Davies, Rob Gaizauskas, Helen Goodluck, Chris Herdman, Rob Holte, Miroslav Kubat, Stan Matwin, Franz Oppacher, Bruce Pappas, Shu Hui Wu, and Helmut Zobl.

My years as a PhD student were vastly improved by the intellectual and emotional support, as well as the friendship, of my fellow students and postdocs. Thanks to them, I'll be leaving with many warm memories of my time at Carleton.

I am extremely grateful to all my other fantastic friends, and in particular to Al, Fred, Gabi, Isabel, Jen, Julie, Martine, and Mélissa, who came through for me at a crucial moment.

My parents have always been fully supportive throughout my endless years of schooling, without once asking me why I was bothering. I love them for that.

Finally, I gratefully acknowledge the financial support of Carleton University and the National Science and Engineering Research Council.

Table of Contents

Chapter 1: Introduction

Vern: You think Mighty Mouse could beat up Superman?

Teddy: What? Are you cracked?

Vern: Why not? I saw the other day he was carrying five elephants in one hand!

Teddy: You don't know nothing. Mighty Mouse is a cartoon. Superman is a real guy. No way a cartoon could beat up a real guy.

Vern: Yeah. Maybe you're right. Would be a good fight though.

– *Stand by Me* (Columbia Tri Star, 1985).

Superman vs. Mighty Mouse

Personally, my money's on Mighty Mouse. All he needs is a chunk of Kryptonite and it's curtains for the so-called "Man of Steel". But that's beside the point. For my purposes, what's interesting about Vern and Teddy's debate is that in order to engage in it, they have to make liberal use of what might be called "non-referring" terms. The names "Superman" and "Mighty Mouse" are non-referring in the sense that they fail to refer to existing individuals in the way that ordinary names such as "David Bowie" seem to. They are non-referring terms because their alleged referents, Superman and Mighty Mouse, don't actually exist in the real world.

But Vern and Teddy are not just *talking* about Superman and Mighty Mouse. They must be *thinking* about them as well. According to the default assumptions of cognitive science, this means that they must be performing computations on their *concepts* of Superman and Mighty Mouse. So in addition to making use of non-referring terms, Vern and Teddy are also making use of their corresponding *non-referring*

concepts. This distinction between referring and non-referring terms and concepts turns out to be an important one, and the source of quite a bit of trouble for some of the most popular and successful approaches to the linguistic and philosophical study of meaning. This dissertation is a philosophically motivated psychological investigation into the structure and processing of those troublesome non-referring concepts. (For those of you who are still thinking about who would win the fight, perhaps Figure 1-1 will be useful.)

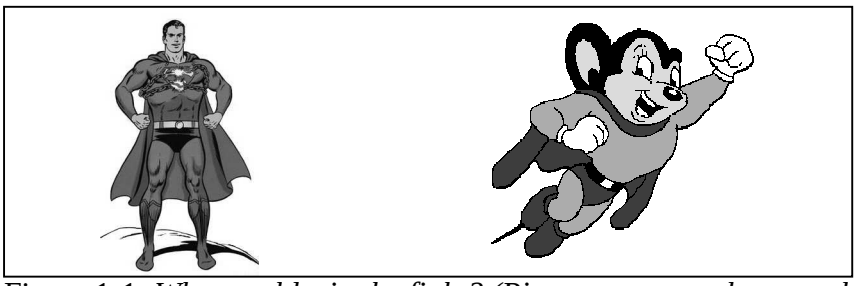


Figure 1-1: *Who would win the fight? (Pictures may not be to scale.)*

The dialogue between Vern and Teddy is pretty silly. But trying to account for their ability to have such a dialogue is not silly at all. In fact, a very large portion of natural language utterances contain non-referring terms, and a correspondingly large portion of human thoughts must therefore involve non-referring concepts. A large proportion (possibly the majority) of all published writing is fiction, and much of what human beings do for personal entertainment involves non-written fiction as well. But even when we're not reading novels, telling stories, watching TV, singing songs, daydreaming, or whatever, our ordinary talk and our day to day thoughts are still full of references to things that don't, couldn't, or at least might not, exist. Whenever we hypothesize, plan for the future, express hopes and fears, or try to deceive each other, we risk making use of non-referring terms and concepts. Reference to the nonexistent is not

just some curious little side product of human nature. It is a central feature of human thought and language.

In this dissertation, I will use the techniques of cognitive psychology to answer two main questions. The first concerns the psychological structure of non-referring concepts: *What is that structure like, and more specifically, do non-referring concepts have a different psychological structure from that of ordinary referring concepts?* The second concerns non-structural psychological differences (e.g., differences in deployment or processing) between referring and non-referring concepts: *Are there any other significant psychological differences between non-referring concepts and their referring counterparts?* These questions are presented in Figure 1-2 below for easy reference.

The Main Questions:

1. What is the psychological structure of non-referring concepts, and does it differ in any way from that of ordinary referring concepts?

2. Are there any other significant psychological differences between referring and non-referring concepts?

Figure 1-2: *The two main questions addressed in this dissertation.*

Related questions concerning the semantic, rather than strictly psychological, properties of non-referring terms have been studied for many years in linguistics and philosophy. In fact non-reference poses huge, possibly insurmountable, problems for one of the major groups of semantic theories – the type that seeks to ground meaning on truth conditions and reference to things in the real world. As things stand now, the default view, at least within the philosophical camp, has to be that non-referring terms, and by extension, non-referring concepts, constitute a *semantic special case*. They either come

by their meanings differently, have a different psychological structure, or are processed differently than referring terms and concepts. But despite this long tradition of concern with the problems of non-reference, and despite a number of strong theoretical claims that have been made about non-referring terms and concepts, there have not so far been any empirical inquiries into their psychological structure. So this dissertation is also an attempt to connect a very old, but also very current, linguistic and philosophical debate to the more recent traditions of empirical research into the psychological structure of concepts.

What Is A Concept?

There are two different meanings of the word "concept", both of which are still in common use. One is found more often in philosophical writing, the other is more common in psychology. Since this dissertation attempts to speak to both philosophy and psychology, the danger of equivocation is too great not to say something right away about how I intend to use the word "concept". The word "concept" throughout this dissertation will mean roughly the same thing as "sub-propositional mental representation", corresponding to the standard use of the term in the psychological (or, to be more precise, cognitive) literature on concepts (Laurence & Margolis, 1999). By this definition, concepts are things in the head. They are the mental entities upon which we perform computations in order to categorize, make inferences, and decide what to do next. For the rest of this section, I'll abbreviate this meaning of the word "concept" using the term "$\text{concept}_{\text{M}}$". The other meaning of the word "concept", which is more common in philosophical writings, is that a concept is an abstract entity, located outside of the head. For now, I'll abbreviate this meaning using the term "$\text{concept}_{\text{A}}$".

To give a sense of the importance of the distinction, on this reading a concept_A such as PENGUIN[1] is some kind of abstract entity that you can grasp, understand, or represent in your mind, but it is not something that is itself actually *in* your mind. The concept_A PENGUIN is external to your mind and would exist as a concept even if there were no penguins, or people to think about them. Following this definition, it makes sense to suppose that you can misrepresent the concept_A PENGUIN and thereby mis-categorize or make incorrect inferences about some of the penguins and non-penguins you encounter. That is, someone's concept_M PENGUIN can misrepresent the correct concept_A PENGUIN. Or, if you prefer, perhaps their concept_M could be taken to represent a different concept_A altogether, one which may be similar to but distinct from, the "correct" concept_A PENGUIN.

The distinction between concept_A and concept_M arises in principle in the case of non-referring concepts, though there is less danger of confounding them. When it comes to non-referring concepts, the idea of representation and misrepresentation of some objectively true concept is a little harder to swallow. Going back to Vern and Teddy, on this reading Vern might have a concept_M SUPERMAN_1 that he expresses with the word "Superman", and it may be structured differently from Teddy's concept_M SUPERMAN_2 which he also expresses with the word "Superman". In fact, maybe this is even the source of their initial disagreement. If so, then they are simply using the same word to express slightly different concepts_M, and there is nothing more to say about it. There is no reason, at least in this particular case, to think of the argument as concerning the representation or misrepresentation of some "correct" concept_A SUPERMAN. There is no reason to suppose that one of the two boys must be wrong – they are discussing a question about which

there is probably no fact of the matter. The question of who would win a fight between Mighty Mouse and Superman is probably just a bad question.

Having contrasted these two senses of "concept", let me say explicitly that this dissertation is intended to be an internalist inquiry, in Chomsky's (2000) sense, into the psychological structure of non-referring concepts_{M}. It isn't directly concerned with non-referring concepts_{A}. I believe it is important for any cognitive scientist investigating concepts or other mental entities to consciously take an internalist perspective, and to acknowledge it up front.

When we are dealing with non-referring concepts_{M} it's pretty clear that there is often simply no fact of the matter about them being right or wrong. But from an internalist perspective this may well be true of referring concepts_{M} as well. If your INSECT concept_{M} includes spiders and mine doesn't, then we have different INSECT concepts_{M}. From the point of view of an entomologist, I'm right and you're wrong, but that is neither here nor there for the internalist cognitive scientist. That doesn't mean, of course, that concepts_{A} are not important. In fact, concepts_{M} can't be studied directly at all without the normative influence of concepts_{A}. If I ask an experimental participant to describe Godzilla and she claims he is a small, mouse-like animal, I would want to exclude her response. The most natural justification for that would be to argue that she got the concept_{A} wrong. The internalist perspective is that even though concepts_{A} cannot be done away with, the focus should still be on concepts_{M}. In any case, from this point on I will drop the annoying subscripts and "concept" should always be read as "concept_{M}".

Why Are Non-Referring Concepts Interesting?

This dissertation examines the psychological structure and processing of naturally occurring non-referring concepts. By "naturally occurring", I mean those concepts that are formed over a long period of time, usually starting in childhood, through natural interactions with the physical and social environment. Naturally occurring concepts, like FURNITURE, MONSTER, TREE, DRAGON, and so on, are already integrated into the conceptual repertoire that adult experimental participants bring with them to the artificial environment of the lab. They have not been created for experimental purposes, and in that sense they are naturally occurring. To date, no experimental work has been done regarding these natural non-referring concepts, and thus the questions addressed in this dissertation should be of considerable interest to linguists, philosophers, and psychologists alike.

Why Linguists and Philosophers Should Be Interested

From a linguistic and philosophical point of view, the experimental work in this dissertation will shed light on a very long-standing debate in natural language semantics. Syntacticians, semanticists, and philosophers of mind and language, particularly those who favor so-called "reference based" or "truth conditional" semantic theories, have long recognized a distinctive class of expressions called "referring expressions". The exact definition varies, but basically the idea is that proper names like "Tony Soprano", natural kind terms like "vampire", and definite descriptions like "the vampire that bit Tony Soprano" get their meanings by picking out, or referring to, objects or sets of objects in the world. When placed in sentences, these expressions contribute their referents (the objects or sets of objects they refer to) to the meaning of the whole. So in the sentence

"Tony Soprano bit the vampire", the semantic contribution of the expression "Tony Soprano" is Tony Soprano, the man himself, and the contribution of "the vampire" is some particular, contextually determined, vampire.

Of course you can already see where this is going. What if there is no referent for a so-called referring expression? How can "Tony Soprano" contribute Tony Soprano to the meaning of a sentence when there is no Tony Soprano? (Sure, there's an actor who plays Tony Soprano on TV, and he does exist, but he's not Tony Soprano, the mob boss, from North Caldwell. He's James Gandolfini, the actor, from Westwood.) There are a number of proposals on the table to resolve this issue and save reference-based semantics, many of which are pretty clearly unacceptable right from the start (for a recent collection, see Everett & Hofweber, 2000). For instance, it's been proposed that non-referring expressions literally have no referent at all (e.g., Everett, 2000). But that doesn't seem right. Surely "Tony Soprano" isn't meaningless – I know exactly what I mean when I say it (and I don't mean James Gandolfini). It has also been proposed that non-referring expressions refer to mysterious, non-material entities variously referred to as "nonexistent objects" (e.g., Parsons, 1980), "abstract objects" (e.g., Zalta, 2000), "elements of a fictional plenitude" (e.g., Deutsch, 2000), and so on. But that doesn't seem like a good way to go either, unless we can find some sort of independent confirmation for these elusive, non-corporeal entities.

The problems of non-reference and the proposed solutions will be discussed in much more detail in Chapter 2, but the bottom line for now is that although many linguists and philosophers remain committed to reference-based semantic theories, no one has yet solved the big semantic problems posed by non-referring terms and the non-

referring concepts that they express. In the absence of such a solution, the default view remains that non-referring terms must constitute a semantic special case. In contrast with the norm, they must get their meanings, and make their semantic contributions to sentences, in some way other than through reference. And if semantics has anything to do with cognition, as we ought to expect it should, then this implies that non-referring concepts must be some kind of a special case too. The questions posed by this dissertation are thus directly relevant to linguistics and the philosophy of language, and the evidence collected could potentially help to resolve some very old and vexing questions about language (and about thought).

Why Psychologists Should Be Interested

From a psychological point of view, it is not obvious in advance what the structure of natural non-referring concepts will turn out to be like either. On the one hand, it is a well-established result that concrete object concepts, like CLOTHING and BIRD, are structured, at least partly, around principles of similarity. An item is generally considered by participants to be a better or worse example of a category to the extent that it is more or less similar to other members of that category (e.g., Rosch & Mervis, 1975). On the other hand, it is known that other types of concepts are not structured around similarity. For instance ad hoc concepts, so-called because they are thought to be generated on the fly rather than permanently stored in memory, are not structured by similarity. The concept THINGS TO SAVE FROM A BURNING HOUSE is a classic example of this kind of concept (Barsalou, 1983). It is also known that some (but not all) abstract concepts fail to exhibit a similarity-based structure. The concept TYPE OF JOB is one such non-similarity-based abstract concept (Hampton, 1979). Finally, there is also a widespread intuition, though

perhaps no proof, that some mathematical or logical concepts, such as PRIME NUMBER, are structured around strict definitions (i.e., around lists of necessary and sufficient conditions) rather than around "fuzzier" relations like similarity (e.g., Armstrong, Gleitman & Gleitman, 1983).

The psychological evidence for the claims made above will be reviewed in detail in Chapter 3. For now, the point is simply that there is no way to predict, prior to experimental inquiry, whether non-referring object concepts (SUPERHERO, UFO, WEREWOLF, SMURF, and so on) will be structured more like referring object concepts, ad hoc concepts, abstract concepts, or logical concepts. It seems quite reasonable, since the things that these concepts fail to refer to would have been concrete objects, that non-referring object concepts will be structured around similarity in the same way that concepts that succeed in referring to concrete objects are. But on the other hand, there has to be *some* difference between referring and non-referring object concepts. At the very least, something in our heads has to be telling us which ones refer and which ones don't, and for that matter, which ones could or might refer, otherwise we'd be permanently out of touch with reality. The question is whether this constitutes a very minimal difference or a big, structural difference.

For instance, one could argue that non-referring object concepts will turn out to be more like abstract concepts. After all, they are abstract, at least in the sense that they lack a concrete object referent, so perhaps their psychological structure will turn out to be more like that of abstract concepts. On the other hand, maybe non-referring concepts are so rare or variable in our thought that they do not have permanent, stable representations at all. In that case, we might expect them to behave more like ad hoc concepts. Finally,

maybe non-referring concepts are structured like definitions. A mermaid, for instance, is basically a woman with a large fish tail instead of legs, so maybe the word "mermaid" actually corresponds to a concept that is itself composed of other concepts, like WOMAN, FISH, TAIL, and so on, put together into some kind of definitional list of necessary and sufficient conditions. One way to decide between these alternative possibilities is to experiment.

So the psychological structure of non-referring concepts is interesting in its own right, but it is also interesting for what it could tell us about human conceptual structure in general. Referring concepts do seem to be tied more closely to the real world than non-referring concepts, and that means that it's possible that the structure of a referring concept like ANIMAL is not as free to vary as the structure of a non-referring concept like MONSTER. Suppose that human psychology, which no doubt allows for some flexibility in conceptual representation, nevertheless influences the structure of our concepts by pushing it towards some sort of psychologically "optimal" form. If so, it's possible that this optimal form might sometimes not perfectly match the most appropriate structure for usefully categorizing the external world. That opens up the possibility that a concept like ANIMAL might be further from the ideal psychological structure than a concept like MONSTER.[2] We are in some sense freer to define, as individuals and as a society, what things are properly called "monsters", than we are to define what things are properly called "animals". If this is right, then an investigation into the structure of non-referring concepts might also reveal some insights about the optimal, natural, or default psychological structure for all human concepts.

Finally, the study of non-referring concepts might eventually be of interest to developmental psychologists. It's been pointed out that children's knowledge of the external world, particularly the knowledge embodied in concepts like LIVING THING and ARTIFACT, develops as they get older (Carey, 1985; Keil, 1989). Young children also have a slightly different take in the distinction between what is real and what is not. Of course, young children often believe that terms like "Santa Claus" and "The Tooth Fairy" actually refer, but they also seem to move effortlessly back and forth between fantasy and reality, sometimes appearing to take on the information that they get from fiction as though it was describing actual objects and events. This is not to say that they do not maintain as sharp a distinction as adults between what is real and what is not (see Woolley, 1997), but there is some evidence of a developmental difference. For instance, young children are more willing than older children to group real and imaginary objects together into common categories (Morison & Gardner, 1978). What happens to their conceptual structure, and their concepts and theories of the world, as they mature and begin to accept the sharper, adult distinctions between reality and fantasy? How do their concepts of the real and the unreal change in order to accommodate their changing beliefs? These are questions that have received little attention in the developmental literature. This dissertation, which exclusively concerns the structure of adult concepts, could serve as a basis for future work on these interesting questions about childhood conceptual development.

Sneak Preview

I think the study of language and the mind should be construed as a naturalistic and, in many cases, purely internalist pursuit. I am sympathetic to the opinion, that Chomsky

(2000) and Jackendoff (1992, in press) seem to share, that we should be quite cautious about applying the lessons learned from the externalist study of reference to the internalist study of concepts and semantic representations. I plan to back up this opinion by presenting experimental evidence that bears on the main questions in Figure 1-2 concerning the psychological differences, structural and otherwise, between referring and non-referring concepts. In fact, it's going to turn out that the structural differences aren't as great as some of the linguistic and philosophical theories seem to predict, and I will argue that this casts a shadow of doubt over their approach and their main claims. On the other hand, there does appear to be some evidence for differences in the way referring and non-referring concepts are processed and deployed.

I will answer the main questions of Figure 1-2 by replicating some classic categorization experiments from the psychological literature on naturally occurring concepts, but this time using non-referring terms as the stimuli. In particular, I will attempt to replicate some of the basic results from an experimental paradigm pioneered by Eleanor Rosch and her colleagues in the 1970s. Rosch's work showed for the first time that naturally occurring concrete object concepts (FISH, WEAPON, and so on) do not seem to be structured around strict definitions – an account that had been the prevailing view beforehand. Rather, it seems that these concepts are structured, at least partly, around some kind of psychological similarity, sometimes called "family resemblance", or "prototype structure". Rosch found evidence for prototype structure using simple property listing and instance rating tasks, and performed a series of speeded categorization studies to demonstrate that the prototype structure has a measurable effect on semantic processing as well. The logic of my experimental work is that if this basic set

of results can be replicated for non-referring concepts, then we can take that as evidence that non-referring concepts do not constitute a special case, and therefore the default position of reference-based semantics is untenable. But if the results cannot be replicated, or if there turn out to be significant structural differences between referring and non-referring concepts, then maybe there's some truth to the default position of reference-based semantics after all. As it will turn out, though, many of Rosch's basic results *can* be extended to non-referring concepts, and the evidence therefore points to the first conclusion.

I will discuss and motivate the general problem posed by non-referring terms and concepts in Chapter 2, trying to show that within the confines of traditional referential semantics, the problem is such that it can't be solved without appealing to special structures or processing mechanisms for non-referring concepts. I will then lay out a couple of these "special structure/mechanism" solutions that are currently on offer. Then in Chapter 3, I will review the relevant empirical work on concepts, and in Chapter 4, I will present the results of my own empirical work on non-referring concepts. The results of this empirical work suggest that non-referring concepts are probably not special cases, and thus gives reason to doubt the correctness of reference-based semantics. Finally in Chapter 5, I will present some positive ideas about the psychological structure of concepts that can account for what we now know about both referring and non-referring concepts.

Two Ways To Read This Dissertation

One way to read this dissertation is simply as an empirical investigation of some claims about the structure of language and thought made by certain theories in linguistics and

philosophy, and about the psychological structure of the sub-class of non-referring concepts. But there's another way to read this dissertation as well. For the last 15 years or so, linguists like Chomsky (2000) and Jackendoff (1992, in press) have been engaged in a philosophical debate aimed at banishing certain varieties of the philosophy of mind and language from the heartland of cognitive science. As I read them, they believe that much of the traditional philosophical preoccupation with external content, objective truth, and intentionality, has no place in a properly naturalistic study of human language and thought. The claim is that the mind, or parts of the mind such as the language faculty, should be studied like any other natural objects – for what they *are* rather than for what their *function* is, or for what the various bits *mean*. To put it bluntly, there can be a science of psychological mechanisms and linguistic representations, but there can be no science of intentionality or aboutness. The debate over internalism is largely beyond the scope of this dissertation, but it seems to me that the approach I am taking can provide some evidence that is important to it. That's the other way to read this dissertation. I worry that Jackendoff and Chomsky go too far in their criticisms of externalism, but I think that their basic orientation is correct, and that they're putting their finger on an important issue about how we should go about doing cognitive science, namely whether it is always necessary, or even appropriate, to mix externalist and internalist projects. The general issue of mixing philosophical projects will come up repeatedly in the next chapter.

Chapter 2: Reference and Reality

> I'm not normally a praying man, but if you're up there please save me, Superman!
>
> – Homer Simpson (*The Simpsons*, "Lost our Lisa")

Semantics and Psychology

This dissertation asks two *psychological* questions about non-referring concepts. To paraphrase from Figure 1-2, those questions are: What is the psychological structure of non-referring concepts, and what psychological differences are there between referring and non-referring concepts? For many years, linguists and philosophers have struggled with related *semantic* questions about non-referring terms: What is the semantic structure of non-referring terms, and what semantic differences are there between referring and non-referring terms? This chapter is about the reasons these semantic questions have come up in linguistics and philosophy, what sorts of answers have been suggested, and what predictions can be derived from them about the psychological structure of non-referring concepts. Along the way, we'll also encounter the recurring theme that when you try to tackle two projects at once, as several of the philosophers reviewed in this chapter are guilty of doing, you can often end up doing a disservice to both.

First, a very basic question needs to be addressed concerning the relationship between semantics and psychology. How closely related are the questions semanticists have asked about linguistic expressions to the psychological questions I am asking about concepts? If you believe, as some do, that language is properly conceived of as a social or normative phenomenon, external to the minds or brains of the individuals who produce and understand it, then it may make sense to say that the answers to the semantic

questions can be cut off from the answers to the psychological questions. From this perspective, it might seem reasonable to entertain the possibility, for instance, that non-referring terms are a semantic special case but the corresponding non-referring concepts are not a psychological special case, or vice versa.

If, on the other hand, you believe that language is at least partly a psychological phenomenon, internal to the minds or brains of individuals, and that the nature of language, both external and internal, derives at least in part from the nature of human psychology, then you might be more inclined to expect that the answers to the semantic questions should agree with the answers to the psychological questions. From this perspective, psychological evidence showing that, for instance, non-referring concepts have the same structure as referring concepts, should be taken very seriously when developing a semantic theory for non-referring terms, possibly even being used to rule out certain types of solutions. That is the perspective from which this dissertation is written, and it is in this spirit that I now turn to look at what semanticists have had to say about non-referring terms.

The Puzzles of Non-Reference

Why do semanticists care about non-referring terms? The concern comes from a very common and sensible intuition that has given rise to some of the best and most successful semantic traditions – the intuition that words stand for, or refer to, things in the world. This referentialist intuition best accounts for the meaning of noun phrases and proper names that pick out concrete objects. What could be more obvious and true that, for example, a phrase like “the mouse” generally refers to some particular, contextually determined, mouse, and that a proper name like “Theodore” generally refers to some

particular individual, maybe somebody's pet mouse. This leads to a very natural interpretation of the role of such words in simple sentences, like (1) and (2) below.

(1) The mouse ran up the clock.

(2) Theodore ate some bad seeds.

These sentences say things about the individuals referred to by the noun phrases: (1) says that some particular, contextually determined, mouse ran up some particular, contextually determined, clock, and (2) says that an individual named "Theodore" ate some bad seeds. On this semantic picture, the phrases "the mouse", "the clock", "Theodore", and "some bad seeds" all not only refer to things, but contribute those things directly to the semantic interpretation of the sentences in which they appear. That is, (2) says, of Theodore, that he ate some bad seeds. It attributes a property to that particular individual, Theodore – the property of having eaten some bad seeds. Whether the statement made in the sentence is true or not is then determined by whether or not the thing named has the attributed property.

To flesh out this naïve intuition about reference into a full theory (that is, a reference-based semantic theory) requires quite a bit more development. For instance, it may require the notion of truth values as the referents of sentences, or functions as the referents of many of the non-object denoting constituents of language (verbs, adjectives, and logical connectives in particular).[3] But the fundamental idea, which is still as popular as ever, remains that the noun phrases and names that appear in a given sentence contribute their referents to the semantic interpretation of that sentence, and the other elements of the sentence attribute properties, or say things about, those referents. It's all very sensible and intuitive, until the trouble starts.

Consider the sentences below concerning Stuart Little, a well-known talking mouse character who has been the subject of a number of books and movies, and Papa Smurf, the spiritual and intellectual leader of the little blue people known as "smurfs" in the children's cartoon of the same name (see in Figure 2-1 if you're unfamiliar with these characters).

(3) Stuart Little is a mouse.

(4) Stuart Little is Papa Smurf.

(5) Stuart Little is Stuart Little.

(6) Stuart Little does not exist.

Note that (3) to (6) are not meaningless. They are all readily understandable, and they are all easily judged to be true or false. Stuart little is a mouse (3), and not a smurf (4). He would seem, like every genuine object, to be identical with himself (5), but he doesn't really exist (6).[4] The immediate problem that these sentences pose for reference-based semantics is that they all contain the non-referring name "Stuart Little". "Stuart Little", according to the intuitions of reference-based semantics, picks out an individual, and sentences containing that phrase say things about that individual. But in some intuitive sense, there is no such individual in this case – Stuart Little isn't a "real" thing. So what is it that (3) to (6) are supposedly saying something about? In fact (6) gives the puzzle the flavor of a real paradox. If "Stuart Little" contributes some referent to the semantics, then how could it ever be true of that referent, of that thing being contributed directly to the semantics, that it does not exist? To whom or what, exactly, is the property of not existing being ascribed?

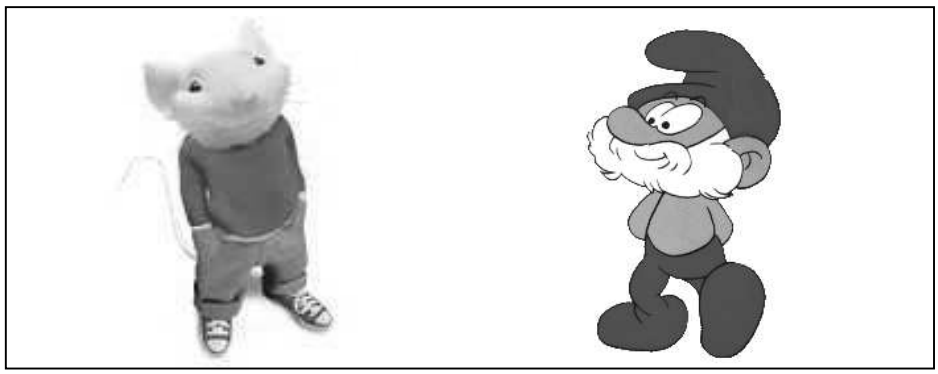


Figure 2-1: *Stuart Little (left) and Papa Smurf.*

Simple Solutions

If you've never encountered these sorts of puzzles before, or even if you have, you may be getting uncomfortable at this point. There are two observations that many people instinctively offer as the basis of a simple solution to the puzzles of non-reference. The first is that objects like Stuart Little are representations. For example, you might argue that Stuart Little is an animated character. Therefore the phrase "Stuart Little" does refer – it refers to the animated character. So sentences containing the phrase "Stuart Little" are saying things about that animated character. This is a very sensible thing to say, and there's a reading upon which it is also true. Stuart Little *is* an animated character, after all. But unfortunately, this "representational solution" does not work as a solution to the technical semantic problem at hand.

Suppose someone makes a videotape of me typing my dissertation, and then shows it to you, points at the animated character on the screen and says, "Sam is typing his dissertation." You would not take from this statement that "Sam" is the name for that animated character, the two-dimensional pattern of pixels. You would assume that the character on the video screen is a representation of another object, namely me. In fact, the person's statement wouldn't make sense at all unless you did. Animated characters are

not people. Rather, they are patterns of color on a screen, or electromagnetic pulses on a tape, or perhaps something within us (some kind of perceptual illusion), or whatever you like. But whatever they are, they're definitely not people, and they don't have the ability to type dissertations. So the only sensible interpretation of the statement, "Sam is typing his dissertation," even if you have never seen me in person and have no proof that I exist, is that it is a statement about me, and not a statement about the animated character on the screen that represents me.

The same argument can be made for Stuart Little. The animated character you see on the screen and the drawings you see in the books are representations of Stuart Little. They are not Stuart Little himself. To see that this distinction (between representations of Stuart and Stuart himself) has semantic consequences, consider your intuitions about (3) and (6). If the name "Stuart Little" contributes an animated character or a drawing as its referent to the semantic interpretation of those sentences, then they should both be false. An animated character could not actually *be* a mouse, though it might be used to represent one. So (3) must be false on this account. But the usual interpretation of (3) is that it is true. Stuart Little *is* a mouse. As for (6), it seems that animated characters, however you define what they really are, really do exist. So (6) should be false as well. But the usual interpretation of (6), the interpretation that would usually be intended by someone who uttered it, is that it is also true. Stuart Little *doesn't* really exist. So the representational solution gives us the wrong interpretation for those sentences, and must be rejected.

The other simple solution that is often proposed is that things like Stuart Little are just ideas in our heads. So (3) to (6) are actually saying something about the idea of

Stuart Little, rather than about Stuart Little himself. Again, this seems a sensible thing to say, and may even be true on some interpretations. In fact, Steven Schiffer has seriously proposed a solution to non-referring phrases that tries to take this "conceptualist" intuition into account (Schiffer, 2000). Unfortunately, this move also fails as a solution to the semantic problem, for the very same reason that the representational solution failed. Ideas, like representations, are just the wrong sorts of things to put into the semantics to give us the proper interpretations of those sentences. If (3) is saying, about a particular idea, that it is actually a mouse, then it must be false. No idea can actually be a mouse. And if (6) is saying, about a particular idea, that it does not actually exist, then it must be false too. The idea of Stuart Little definitely exists. It's Stuart Little himself who does not.

There are a number of other obvious solutions that jump to mind, all of which have fatal problems as well. One such possibility is that non-referring phrases are simply empty – they refer to nothing, or perhaps to the empty set. Anthony Everett (2000) recently explored a solution along these lines. But on its own, this suggestion does not really help, since it makes all non-referring expressions refer to the same thing – namely, nothing, or the empty set, or whatever. So it is difficult to explain our intuitions about sentences like (4). If "Stuart Little" and "Papa Smurf" both contribute the same thing to the sentence (or contribute the same "nothing"), then the statement that Stuart Little is Papa Smurf must be true.[5] Another possible move is to simply deny truth values to sentences containing non-referring expressions. This was the solution preferred by Frege (1892). Frege thought that sentences containing non-referring expressions could be neither true nor false, but should be tagged *fictitious*. But not only is this solution not very

useful in accounting for our intuitions about (3) to (5), but it actually seems to give the wrong answer for (6). In what sense is a statement that denies the existence of something like Stuart Little a "fictitious" statement? Surely, it's a true statement.

The problems posed to reference-based semantic theories by the existence of non-referring terms are hard problems and they resist simple or dismissive solutions. The fact that reference-based semantic theories are still, in many ways, the default approach for working semanticists reflects the strength and power of the underlying intuition, rather than the fact that the problems of non-reference are unimportant or have been solved to everyone's satisfaction. The problems of non-reference, particularly those posed by so-called "empty names", are still the subject of ongoing debate a full century after they were first discussed by Frege and Russell, the fathers of modern semantics. (For recent discussion of non-referring names, see the dissertation by Caplan, 2002; or the collection edited by Everett & Hofweber, 2000). But in the absence of an agreed upon solution, the default position must remain that non-referring terms are simply a semantic special case, requiring a special semantic solution.

The purpose of what follows is to highlight some of the more notable non-simple solutions to the problems of non-reference. The next section, "Extreme Solutions," discusses the ideas of Alexius Meinong and Bertrand Russell, who exemplify two diametrically opposed solutions to the problems of non-reference. I argue that both the solutions they suggested were derived from an inappropriate intermingling of semantics and metaphysics, and that decoupling the two projects opens up a third, internalist possibility, which I call "Psychological Meinongianism". Then in the section on "Special Case Solutions", I discuss the default position within reference-based semantics, focusing

in particular on suggestions from Kenneth Taylor, Jerry Fodor, and Fred Dretske as examples of special case solutions that have an explicitly psychological flavor. Fodor and Dretske's solution is another example of a possibly inappropriate intermingling of philosophical projects, this time semantics and metasemantics.

Extreme Solutions

To summarize what we've seen so far, there's a really promising semantic theory based on reference as a fundamental element of linguistic (and mental) content. Unfortunately, on this view, phrases like "Papa Smurf" and "Stuart Little" look like they ought to have referents, that is they look and feel a lot like terms that do have referents, such as "Papa Doc Duvalier" and "Martha Stewart". But "Papa Smurf" and "Stuart Little" don't have referents, and the simple solution that they refer to representations or ideas doesn't work at all. Nevertheless, that doesn't seem to stop us from stating meaningful things that are apparently about Papa Smurf and Stuart Little. Worse yet, it doesn't seem to stop us from stating clearly true and false things about them. What do we do now?

This section looks at two historical suggestions for solving the semantic issues of non-reference using extreme measures that most linguists, philosophers, and psychologists have subsequently ruled out. At one extreme is the suggestion, from Bertrand Russell, that referring and non-referring terms should be removed from the logical interpretations of the sentences. If you accept this, it turns out that you have to accept that even the simplest sentences often mean a lot more than they appear to. At the other extreme is the view, most often attributed to the German psychologist and philosopher Alexius Meinong, that non-referring terms really do refer, albeit to nonexistent objects.[6] If you accept this view, you have to accept the bizarre and

seemingly contradictory view that the world is full of all sorts of things that don't really exist. In fact, both of these extreme solutions derived, at least in part, from Russell and Meinong's intuition that language and thought must mirror reality, leading them to allow their semantics and metaphysics to inappropriately constrain one another. Recognizing and subsequently letting go of this motivation opens up the possibility for a third, less extreme, solution.

Russell's Semantic Solution[7]

Russell's best known exploration of non-referring terms is in "On Denoting" (Russell, 1905), where he introduced the notion of a quantificational paraphrase to explain the semantics of sentences containing overt definite descriptions (like "the talking mouse") and apparent names (like "Stuart Little"). Russell's goal was to assign truth values to sentences containing apparently non-referring definite descriptions, such as his classic example in (7) below.

(7) The present King of France is bald.

Russell was not satisfied with the suggestion of Frege's, mentioned above, that truth values should be denied to sentences containing non-referring terms – that they are all *fictitious* rather than being true or false. Russell argued that, on the contrary, (7) is "plainly false" since there is no present King of France (Russell, 1905, p. 108).

Russell identified two possible ways to proceed: Either provide a referent for phrases like "the present King of France" (the solution he attributed to Meinong) or construct a semantics in which such phrases do not refer at all. Russell favored the latter approach, and the solution he hit upon, put in modern terminology, was essentially to

treat the definite article as a special quantifier phrase. Thus (7), according to Russell, means the same thing as the following paraphrase:

(8) There exists exactly one present King of France, and all present Kings of France are bald.

Put in quantificational notation, the paraphrase looks something like this:

(9) $\exists_1 x \mid \text{PresentKingOfFrance}(x) \wedge \forall y [\ \text{PresentKingOfFrance}(y) \rightarrow \text{bald}(y)\ ]$

Since (7) is now interpreted as implying the existence of the present King of France, the whole evaluates to the truth value that Russell wanted. (7) now comes out as false.

Of course the argument about definite descriptions does not apply directly to proper names like “Papa Smurf”, but in the same paper, Russell extended the theory to include them by suggesting that most proper names are actually “disguised definite descriptions”. So for example, the name “Papa Smurf”, for Russell, would have actually been shorthand for a definite description such as “the oldest and wisest smurf”, and Russell’s quantificational paraphrase could then be applied to this description. So according to Russell, most statements of any substance containing non-referring terms will turn out to be false because, to put it crudely, they assert the existence of their referents. For example, (10) below paraphrases as in (11), and comes out false since there are no smurfs at all, let alone an oldest and wisest one.

(10) Papa Smurf wears a red hat.

(11) a. The oldest and wisest smurf wears a red hat.

b. There exists exactly one oldest and wisest smurf, and all oldest and wisest smurfs wear red hats.

There are some caveats here, however. Negating false sentences containing non-referring terms can sometimes lead to true sentences. Consider the negative statement

about Stuart Little in (12) below, and the semantic paraphrase, with the non-referring proper name unpacked into a definite description, in (13) below:

(12) Stuart Little is not a smurf.

(13) The talking mouse who wears Converse sneakers is not a smurf.

The negation in this statement can be read with either wide scope (14a) or narrow scope (14b).

(14) a. **not [** There exists exactly one talking mouse who wears converse sneakers and all talking mice who wear converse sneakers are smurfs. **]**

b. There exists exactly one talking mouse who wears converse sneakers and all talking mice who wear converse sneakers are **non-**smurfs.

According to the wide scope reading (14a), the negated statement is true, while according to the narrow scope reading (14b) it is still false since it asserts the existence of a talking mouse. So some negative statements containing non-referring terms can be true, but in those cases they are only true because the entire sentence is negated.

Although Russell's theory of disguised definite descriptions still has some adherents (e.g., Faderman, 2000) there are a number of very serious objections that have been raised against Russell's proposal. Below are three objections that are particularly pertinent here because they attack Russell's semantic theory for its psychological implausibility.

1. Russell's paraphrasing introduces new assertions. By introducing quantificational paraphrasing into the semantic interpretations of sentences, Russell has brought in new assertions that were not explicitly present in the original sentences, and may not have been intended by the speakers. For instance, most people's intuitions allow for the fact that one can quite legitimately say things like:

(15) Papa Smurf is very old, and

(16) Stuart Little wears Converse sneakers,

and consider these sentences true without also believing in the existence of fictional characters. Yet on a Russellian interpretation, it is not just the case that all the statements are false – they are false because they *assert the existence* of Papa Smurf and Stuart Little even though this is not made explicit in the original utterance. If Russell's semantic theory is given a strong psychological reading, one in which semantic structure directly reflects psychological structure, it implies that one cannot think the singular, or directly referential, thought corresponding to (15), that Papa Smurf is very old. Rather one can only think the quantificational thought that there exists exactly one smurf who is the oldest and wisest smurf, and he is very old. And that just doesn't sound very plausible.

2. Russell assumed what he set out to prove. Part of Russell's motivation was that statements like (7) are "plainly false." But this obviously was not so plain to other thinkers at the time, such as Frege, who was quite comfortable assigning the value fictitious. In fact, it is quite likely that most people today, posed with "true or false" questions about entities like the present King of France, would probably be quite uncomfortable with simply answering, "That's false." In the end, they may pick false because it is the lesser of two evils, but most would probably prefer to respond, "That person doesn't exist," and avoid the issue of assigning a truth value altogether. Furthermore, statements like (15) and (16) concerning established characters from fiction also come out false on Russell's reading, contrary to most speakers' intuitions. Russell never argued for *why* most statements involving non-referring terms must be false, and it could be claimed that they are only obviously false if Russell's system has already been accepted, and native speakers' intuitions are ignored.

3. Russell's disguised definite descriptions look a lot like definitions. This objection must mostly be left as a promissory note at this point, but the idea that ordinary words of any kind correspond to definitional concepts has taken a sustained beating in the last 30 years or so. The psychological evidence for this will be reviewed in Chapter 3, but suffice it to say that these days, very few researchers believe that many concepts can be decomposed into exhaustive summary definitions, such as "a serf is a person in bondage or servitude". Russell's disguised definite descriptions look an awful lot like such summary definitions (recall my attempt, "Stuart Little is the talking mouse who wears Converse sneakers") and this raises the suspicion that a lot of the psychological evidence against definitions will apply to treating names as definite descriptions as well. More on this in Chapter 3.

Meinong's Metaphysical Solution

Russell offered two rather extreme ways out of the dilemmas of non-reference: Either find a referent, or get rid of the reference. Russell's own work focused on getting rid of the reference, driven as he was by what he called a "feeling of reality" (Russell, 1919, p. 213). But in the end, his solution has to be rejected, and has been rejected by most other theorists, in part because it is psychologically implausible – violating both speaker intuitions and what we know about the structure of concepts. At the same time as Russell was developing his own semantic ideas, the other extreme solution, that non-referring terms actually do have referents, was being actively defended by Meinong.

Meinong's starting point was the same as Russell's, namely the intuition that every sentence and every thought seems to have an object towards which it is directed. He also seemed to agree with Russell that a correct metaphysics should, in the end, be

compatible with, even predictable from, a correct semantics. But while Russell prioritized metaphysics over semantics, preferring to make the latter counter-intuitive than to make the former implausible, Meinong prioritized semantics over metaphysics, in particular proposing that objects that do not exist can still serve as the "objects of thought". That is, nonexistent objects can be referred to and can have properties, despite the fact that they don't exist (Meinong, 1904). The metaphysical conclusions that ought to be drawn from this are controversial, and exegesis of Meinong is not my goal. According to Russell's interpretation of Meinong's view, which I will adopt both for the sake of convenience and because I think it makes the most sense, Stuart Little and Papa Smurf are nonexistent objects with some kind of being, and that's why we can refer to them. As objects, they also have properties, and that's why we can say true and false things about them. [8]

The idea that words can stand in a "refers to" relation with things that don't exist might seem a little crazy. Certainly it has struck most philosophers since Russell that way. But Meinong wasn't crazy, and a number of modern philosophers, who also aren't crazy, have recently tried to follow in his footsteps (e.g., Caplan, 2002; Deutsch, 2000; Parsons, 1980).[9,10] In fact, it is arguably true that Meinong's approach was a much better fit to psychology than Russell's was. The odd metaphysics simply came from Meinong's attempt to rescue the basic psychological insight that every thought has an object towards which it is directed from the obvious rejoinder that the objects to which thoughts seem to be directed often don't exist. Although the idea that there are nonexistent objects, at least that they have some kind of being, is almost certainly false, it is not quite as bizarre as it is often taken to be. Positing unseen, undetectable, or simply difficult to fathom phenomena is not an unusual thing for a science to do. Astronomers suggest that the

universe is full of undetectable dark matter, mathematicians suggest that numbers are abstract objects that have fixed and definite properties, and physicists suggest that space itself is curved, and time passes at a different rate depending on where you are. These are all rather bizarre, yet perfectly acceptable ideas in the service of scientific explanation. So the idea of nonexistent objects shouldn't be ruled out simply on the basis of it being bizarre. Nevertheless, I think it can be ruled out along with Russellian solutions, because the motivations for suggesting both are misguided.

The intuition behind reference-based semantics, once again, is that words (and concepts) pick out objects in the world and that sentences (and thoughts) must be about those objects. When this intuition is followed through to its logical conclusions, a number of difficulties and contradictions emerge. In particular, semantic interpretations no longer line up very neatly with what appears to be true of the world. That is, we appear to be able to truthfully attribute properties to "things" that don't exist. But this semantic fact in itself is not necessarily a cause for metaphysical alarm – that is, we don't need to opt for either of the Russellian or Meinongian extremes. If we instead accept that the semantics of thought and language is just not a perfect fit to what is true about the world, then the more interesting challenge is to explain how we can think and speak more or less intelligibly, and do useful things with thought and speech, given that the system contains such "imperfections". To feel that we have to accept either the Russellian conclusion that our referential intuitions are wrong and have to be removed from the semantics, or the Meinongian conclusion that our metaphysical intuitions are wrong and we now have to accept the existence of nonexistent objects, may be based on the unsupported assumption that ordinary language and thought must mirror reality in a straightforward way.

Taking the Sting out of Meinong

If we drop the assumption that language and thought mirror reality, the way is cleared for the possibility that human psychology is at least partially committed to a Meinongian metaphysics, even though this metaphysics may not be the most accurate way to describe the world. That is, we can accept that there is some truth to the intuition that we can refer to things that don't exist, but deny the further conclusion that this tells us something metaphysical. It can be "true", in the sense that we judge it to be true, that Stuart Little wears Converse sneakers, but at the same time literally true that Stuart Little does not exist and does not have any other kind of being either. We can remain agnostic on the metaphysical questions, and focus instead on the internalist questions about how the psychology of reference works, and how it is that we can come to "refer to" and make "true" statements about nonexistent objects without permanently losing our grip on reality. In fact, this may be very close to what Caplan (2002, Chapter 2) argues is the proper reading of Meinong's original position, namely that nonexistent objects, though they can be referred to, don't exist and don't have any other kind of being either.[11] Call this the "Psychological Meinongian" position.

Taking a step back and viewing semantics from an internalist perspective, the intuition from reference-based semantics that words "contribute their referents" to sentences cannot literally be true. The name "Noam Chomsky", while it might literally put Noam Chomsky, the man himself, into the semantics of the sentences in which it occurs, cannot literally put Noam Chomsky into someone's head. To put it another way, Noam Chomsky himself may be required to make the semantic interpretations of sentences about him come out correctly, but he could not be required, at least in such a

literal way, in any internalist psychological explanation of the thoughts underlying statements about him. Similarly, Papa Smurf himself might seem to be required to explain the semantics of sentences including the name “Papa Smurf”, leading to all the familiar problems, but Papa Smurf himself is not required in any internalist explanation of the psychology underlying those sentences. From an internalist perspective, reference looks the same regardless of whether the thing being referred to exists or not.

The Psychological Meinongian position avoids Russell’s unmotivated semantic solutions, preserves Meinong’s original prioritization of psychology over metaphysics, but stops short of making any bizarre-sounding metaphysical predictions. Psychological Meinongianism does not offer a solution to the semantic problems that pushed Russell and Meinong into extreme positions, but it does perhaps provide a way of making progress on a related project by making an empirically testable prediction. Regarding the two central questions posed in this dissertation, the Psychological Meinongian position would predict that the answer to both is the same: Non-referring concepts have exactly the same psychological structure as their referring counterparts (whatever that turns out to be), and do not differ significantly in any other way either.

Special Case Solutions

The Psychological Meinongian position is at odds with the default position on non-referring terms in the philosophy of language. The default position is that non-referring terms come by their meaning in a different way from referring terms. On a strong psychological reading, the default position makes the opposite prediction to the Psychological Meinongian position, implying that non-referring concepts must have a

different structure from referring concepts. That is, they must be special cases. This section discusses two different pseudo-Russellian versions of the special case solution.

There is an obvious modification to Russell's theory of disguised definite descriptions that leads directly to the default position. So before turning to the default position, a brief word is in order about Russell's own position. Russell actually argued that, at least in the case of proper names, there are those that directly refer and those that are actually disguised definite descriptions, and therefore are quantificational rather than referential (Russell, 1905). But his own "special case" solution actually applies to any name referring to something with which the speaker was not directly acquainted. So on Russell's suggestion, only an extremely small subset of apparently referring names would actually refer, and this subset would be different for every speaker. Really, it was the directly referring names that were the unusual special cases. But this willingness to draw the line between those names that refer and those that are disguised definite descriptions opens the door to a pseudo-Russellian special case solution in which most names directly refer, but non-referring names are special cases which unpack semantically into definitions or definite descriptions. This section reviews two separate proposals that take on, at least in part, this pseudo-Russellian idea. Kenneth Taylor proposes that our intuitions regarding the truth values of statements involving non-referring terms arise during a special stage in the pragmatic processing of non-referring terms, in which definite descriptions are substituted on the fly for the corresponding terms. His account is meant to apply to proper names only. Fodor and Dretske, on the other hand, have proposed (somewhat apologetically in Fodor's case) that while referring terms correspond to simple, unstructured, word-like concepts, non-referring terms must

correspond to complex, structured, definition-like concepts. Their account extends beyond names to include natural kind terms as well.

The attempts to deal with non-referring terms reviewed in the previous sections have two features in common: 1) They are attempts to show the way to a unified semantics that will account for the meanings of both referring and non-referring terms in the same way, and 2) the solutions are pitched at the level of semantics, and are not made explicitly psychological. The two solutions reviewed here are different in both respects. They are both versions of the default position, the "special case" analysis of non-referring terms, and they are both explicitly psychological, making direct predictions about the structure and processing of non-referring terms and concepts. This makes them particularly relevant and interesting to the questions posed in this dissertation.

A Pseudo-Russellian Pragmatics

Kenneth Taylor's (2000) version of the pseudo-Russellian special case solution concerns what happens during the pragmatic processing of a sentence. Pragmatics is the study of how the full speaker meaning and the contextual implications of an utterance are computed by the listener. These speaker meanings and contextual implications are generally considered to be separate and distinct from the literal meaning of the expression in the language which, on one standard view, is the proper domain of semantics. Whereas the relationship between semantics and psychology is not always direct, the relationship between pragmatics and psychology must be. So the fact that Taylor's solution concerns pragmatics implies that it should be read as a psychological theory, making psychological predictions.

To see why pragmatics is more directly concerned with psychological processes, consider (17) below.

(17) She'll trip over that.

Even without any contextual information, (17) does have some meaning. It means, roughly, that some contextually salient female individual will trip over some demonstrated object. We can make inferences based on the assumption that the utterance is true, such as someone will trip over something, that person will not trip over herself, and so on. This type of context invariant meaning is referred to as "expression meaning" and is usually taken to be within the domain of formal semantics. But we cannot know the full meaning of some utterance of (17) without the contextually-based filling in of the pronouns "she" and "that". This filling in process is one of the most basic pragmatic processes, and contributes to what is often referred to as "literal utterance meaning". Utterance meaning for a given sentence changes from context to context, and often cannot be accounted for by formal semantics in the same way that expression meaning can. Since filling in is contextually constrained, the process through which it is done must at least partly be an inferential one, involving much more general psychological processing on the part of the listener than mere semantic decoding.

The full import of what a speaker means to communicate by uttering (17) might go quite a bit further than both the expression meaning and the literal utterance meaning. For example, the speaker, in uttering (17) might be suggesting that you pick up your bag, or that your idea to put a raised floor in the living room of granny's house is a bad one. These sorts of implications, known as "implicatures", go well beyond what formal semantics and filling in can tell us about the meaning of an utterance, are fully dependent on contextual and world knowledge, and seem to require the listener's full cognitive

apparatus to compute (Grice, 1975; Sperber & Wilson, 1995). It is quite common to separate the process of computing literal utterance meaning from the process of computing the full speaker meaning (e.g., Stanley, 2000). In the framework that Taylor adapts from Recanati (1993), the process of filling in the slots in the expression meaning to yield utterance meaning is referred to as a "primary pragmatic process". The process of computing speaker meaning from utterance meaning is then referred to as a "secondary pragmatic process".

Recanati's semantic framework, which Taylor also adopts, is based on direct reference, proposing that directly referring expressions (in this case names) carry a semantic feature **REF** which indicates that the expression in question "contributes just its referent and not any mode of presentation [roughly, the descriptive content of a definite description] thereof to the proposition expressed by any sentence in which it occurs" (Taylor, 2000, p. 18). Taylor's only extension to Recanati's semantics is to propose that empty (non-referring) names carry the semantic feature **REF** as well – that is, empty names do not have a special semantic structure. But Taylor maintains that there are important differences in the processing of empty names, due to the fact that sentences containing empty names (when contextualized) fail to express propositions. A proposition must be about something, and empty names fail to contribute something for the proposition to be about. So Taylor refers to the proposition-like entities expressed by sentences containing empty names as "propositions in waiting". For Taylor, Sentence (16), reproduced below, expresses a proposition in waiting with an empty slot where the referent of "Stuart Little" would have gone. This empty slot is marked with the name

from which it was derived ("Stuart Little") and the semantic feature **REF**, indicating that the name was used in a directly referential way.

(16) Stuart Little wears Converse sneakers.

Since Recanati's framework is based on direct reference, the process of filling in referents for proper names works very much like the process for filling in the referents of pronouns. That is to say, obtaining the full utterance meaning from a sentence containing a name, like (18) below, requires the same kind of primary pragmatic slot-filling as does a sentence containing pronouns, like (17), reproduced below. The referent of "Anne" has to be computed from context in much the same way as the referent of "she". When the sentence contains an empty name, like in (16), the primary pragmatic process fails to fill in a referent, yielding a proposition in waiting rather than a full proposition.

(17) She'll trip over that.

(18) Anne will trip over that bag.

This completes the account of Taylor's semantic framework. What he still needs to account for is how we come by our strong intuitions of the truth and falsity of utterances containing empty names. If these sentences don't express propositions, then they cannot literally have truth values. So how do our judgments of truth work in the case of empty names? Taylor's proposal is that there is a "one and a half" stage pragmatic process of "pseudo-saturation" that kicks in when the first stage process of filling in contextual referents fails. According to Taylor, the concepts that underlie proper names are access points to "conceptions", containing encyclopedic information about the appropriate referents.[12] "Pseudo-saturation 'fills' the unfilled object slots in a proposition-in-waiting not with an object, but with descriptive contents drawn from the conception, if there is one, which is labeled and accessed via the relevant name" (Taylor,

2000, p. 33). So someone who utters (16) might, in Taylor's terminology, pseudo-assert a proposition with the content that, for instance, the talking mouse who was adopted by the Little family wears Converse sneakers.

How well Taylor's solution actually works is debatable. Taylor (2000, p. 34) gives only one example of his own, shown in (19) below, and shows how his framework can account for our intuition that it expresses a true proposition.

(19) Santa Claus isn't coming tonight.

But note that this is a negative statement which is easy to make come out true in the pseudo-asserted proposition. The pseudo-assertion would be something like the quantificational proposition that no jolly old elf is bringing us presents tonight. To deal with positive statements like (16), in which the pseudo-assertions seem to come out false when they should be true, may require further pragmatic interpretation – that is, unless you subscribe to Russell's intuition that statements like these are indeed false.

But in any case, Taylor's proposal makes some testable psychological predictions. For a sentence that contains no empty names, the first stage pragmatic process will fill in the required referents and pass its output to the second stage process. For a sentence that contains an empty name, the first stage process will fail to provide a referent for that name, yielding only a proposition in waiting. At this point, the one and a half stage process will kick in to recruit a description from the listener's encyclopedic knowledge relating to that name. This description will be inserted in place of the referent in the proposition in waiting to yield a more complex proposition to be passed to the second stage process. On the entirely reasonable assumption that more processing takes more time, and that more complex propositions take more time to evaluate, Taylor's special

case solution predicts a potentially measurable processing difference for sentences containing empty names.

A Pseudo-Russellian Semantics[13]

This section discusses Fodor and Dretske's pseudo-Russellian suggestion that normal referring terms correspond to simple, atomic, word-like mental representations, whereas non-referring terms correspond to complex, non-atomic, phrase-like mental representations. This suggestion, which appears in both Fodor (1990, p. 124) and Dretske (1981, pp. 222, 230), is not one which is widely associated with Fodor and Dretske, and thus will require some explanation. Explaining Fodor and Dretske's suggestion itself does not take long (in fact I've already done it). Explaining what led Fodor and Dretske to suggest it and why they were compelled to suggest it will take a little longer, but is worth the effort, as it contains another example of the sorts of things that can go wrong when you don't stick to an internalist program. As with Russell and Meinong, I think Fodor and Dretske were driven to their special case solution via an inappropriate running together of philosophical projects. Russell and Meinong mixed semantics and metaphysics with dubious results, whereas Fodor and Dretske mix semantics and what is often called "metasemantics" with equally dubious results.

Up until now I have been discussing reference-based semantic theories and their troubles with non-referring terms. Reference-based semantic theories are theories about what words and sentences mean, rather than about how words and sentences come to have their meanings. According to reference-based semantics, what words mean is what they refer to – there is a connection between words and objects in the world, and it is in this connection that you will find meaning, implying, as Putnam famously put it, that

"meaning ain't in the head" (Putnam, 1975). But where do these word-world connections come from and how are they maintained? Where does meaning itself come from? These are the questions that drive the separate and distinct project of metasemantics.

The classic way to approach the metasemantics of reference is to say that meaning arises from some kind of causal chain connecting words to their referents in the world. As explored by the likes of Saul Kripke (1980) and Hilary Putnam (1975), the fact that, for example, the name "Karl Marx" means Karl Marx can be accounted for by a causal chain of naming that goes back to the original, "baptismal" event in which the name "Karl Marx" was first applied to that individual. Karl Marx's parents first caused the name to be associated with him by themselves associating it with him, and that caused other people to associate the name with him, which in turn caused others to make the association, and so on, leading up to my association of the name with the individual. So when somebody first introduced me to the name "Karl Marx" they hooked me into the causal chain of naming practices in virtue of which that name refers to the individual that it does. This account can also be extended to natural kind terms as well. The word "fish", for instance, comes to refer to fish because of its history of use as a label applied to objects of a particular natural kind.

Note that this metasemantic project is logically independent of semantics. Though it may not ultimately be philosophically satisfying to merely account for word meaning as reference, the question of *what* words mean is still logically separate from the question of *how* they get their meanings. This is important because, although the relationship between semantics and psychology may not always be direct, the relationship between metasemantics and psychology is, in principle, even more tenuous. It is difficult to

imagine what direct role the most distant elements of the causal-historical chains that fix meaning could possibly play in the cognition of an individual in the here and now. Although the parental baptizing of Karl Marx certainly played an important role in determining that everyone in the world now conventionally applies the name "Karl Marx" to him, the proximal causes of most individuals' coming to associate his name with him couldn't have had anything directly to do with such a far off event. So while semantics might reasonably be expected to have something to say about psychology, metasemantics is less likely to do so. It's a good guess that the facts about what a word means play a much bigger role in basic cognitive processes than the metasemantic facts about how it came by that particular meaning.[14] So why bring up metasemantics at all? Because, I maintain, it was an inappropriate mixing of semantics and metasemantics that lead both Fodor and Dretske to propose their special case solution to the problem of non-reference.

Fodor and Dretske are Conceptual Atomists. Conceptual Atomism is both a psychological and semantic theory. Conceptual Atomists think that words stand in a more or less one to one correspondence to concepts, and that these concepts are atomic – that is, they have no syntactic structure and are not dependent on any other concepts for their meaning. Atomic concepts get their meaning through reference, just like words are thought to in reference-based semantic theories. Conceptual Atomism has always been a minority view in cognitive science, developed primarily by Dretske (1981, 1986, 1988), Fodor (1975, 1987, 1990, 1994, 1998), and Millikan (1984, 1989, 1990, 1998), as well as winning a few other supporters along the way (e.g., Margolis, 1998; Usher, 2001). Within this psychologized version of reference-based semantics, telling a story about how the

concepts get their meaning comes to seem much more important. Reference-based semanticists of natural language, due in part to the fuzzy relationship between semantics and psychology, seem to be able to get away with claiming that words just refer, and passing the buck to psychology for an account of how this plays out (or not) in cognitive terms. But the Conceptual Atomists, who are concerned with the psychological mechanisms themselves, have no one to pass the buck to. They need their own story about how unstructured atoms can appear to be meaningful all on their own, hence their strong interest in developing an appropriate metasemantics. In what follows, I will briefly sketch out Fodor and Dretske's metasemantic account, and show how it leads to a special case solution for non-referring concepts.

Dretske was the first to formulate the metasemantics of Conceptual Atomism using information theory (Dretske, 1981). According to his informational view of conceptual content, a concept C represents some X in the world only if C carries information about X. In technical terms, if X and only X causes C then C represents X. The formulation is meant to be counterfactual supporting. So if X and only X *would cause* C, then C represents X. Left like this, Dretske's solution has two big problems. The first is the problem of misrepresentation. The condition that only X would cause C is too strong to apply to real cognitive agents in noisy environments. Even when things are working as they should, the sight of a horse on a dark night might still cause a tokening of one's COW concept, which, in this metasemantic framework, seems to imply that COW would mean the same as "cow or horse on a dark night". The second problem is with non-referring concepts. If the only way a concept gets to have content is by being caused by its referent, then non-referring concepts must either be semantically empty or, worse

yet, have other representations as their content, since only representations of nonexistent objects could cause the concepts in question. But as I've already noted, non-referring concepts can't have representations as their content.[15]

Fodor's solution to these two problems is known as Asymmetric Causal Dependence (Fodor, 1987, 1990, 1994, 1998).[16] The idea combines Dretskian informational semantics (a concept C means X if it's a law that X's cause C's) with an asymmetric dependence condition (Y's that cause C's only do so because X's cause C's and not vice versa). This takes care of misrepresentations such as horses on dark nights causing COW tokens. Mistakenly tokening the COW concept at the sight of a horse is dependent on cows causing COW tokens but correctly tokening a COW concept at the sight of a cow is not dependent on horses (on dark nights) causing COW tokens. The solution is also extendible to explain various kinds of "robust" tokenings (Fodor's term for non-X-caused C tokenings that are nevertheless not error cases – for instance, COW tokens that are caused by pictures of cows or thoughts about farms).

The problem of non-reference for Fodor's solution is immediately apparent. For example, how can non-smurf-caused tokenings of one's SMURF concept be asymmetrically dependent on smurf-caused tokenings when there are no existing smurfs? Fodor thinks that this objection can be answered by reminding us that, like Dretske, he is telling a "nomic" story:

> It can be true that the property of being a smurf is nomologically linked with the property of being a cause of SMURFs even if there aren't any smurfs… There wouldn't be non-smurf-caused SMURF tokens but that smurfs would cause SMURF tokens if there were any smurfs. (Fodor, 1990, p101, emphasis removed, single quotes changed to small caps for consistency, and "unicorn" changed to "smurf" because I'm sick of unicorns.)

I don't think that this response will do.[17] In this smurf-free world, all valid SMURF tokenings must be robust tokenings. The acquisition of the concept SMURF in the absence of smurfs comes from exposure to representations (visual or verbal) of smurfs. Having learned about smurfs from books and cartoons, if a smurf suddenly popped into existence in front of you, it would likely cause a SMURF token. So we have two valid causal routes to SMURF tokens: one from representations of smurfs, and one from possible real smurfs that you might encounter in the future (if smurfs begin to exist). To apply Asymmetric Causal Dependence, we have to know what would happen if we broke either of these two causal links. Would breaking the causal link between future smurfs and SMURF tokens break the link between representations of smurfs and SMURF tokens? My intuition is that this scenario doesn't even make sense, but suppose for the sake of argument that breaking the smurf/SMURF link would break the representation/SMURF link. Then SMURF tokens are causally dependent on (future) smurfs.

But what would happen if we broke the causal link between representations of smurfs and SMURF tokens? According to Asymmetric Causal Dependence, if SMURF is to mean smurf, then this should not affect the causal link between future smurfs and SMURF tokens. But it obviously does. In a world presently without smurfs, if you don't learn about them from representations of them then you don't learn about them at all. This means that if a smurf suddenly popped into existence in front of you, you wouldn't have the concept SMURF to apply to it. Maybe it would cause tokens of SMALL, BLUE, MAN or whatever, but it wouldn't cause a SMURF token because you wouldn't have one for it to cause. So in the best case, causal dependence runs both ways and Asymmetric Causal Dependence doesn't apply. In the worst case (where you don't buy the story about

breaking the link between future smurfs and SMURF tokens) you have Asymmetric Causal Dependence running in the wrong direction and implying that SMURF has representations of smurfs as its content, which, as I've already pointed out, is pretty clearly false – SMURF has smurfs as its content.

There is a way out of this trap for an extremely radical nativist. Fodor (e.g., 1998) entertains, though he does not currently endorse, the possibility that we are born with a stock of atomic concepts waiting to be triggered by the right sort of content-fixing experiences. Applying this idea to non-referring concepts, if we all have built-in SMURF concepts that just need to be triggered somehow, then maybe, even in a world presently without smurfs, our first encounter with a newly existing one would cause a SMURF token after all. Of course we wouldn't have a word for this concept, but that is irrelevant. So Asymmetric Causal Dependence would be satisfied by assuming that we are born with a lifetime supply of concept types for as yet unreal things, that already have their nomic triggering conditions fixed.

Radical nativism is not a popular option in cognitive science (Laurence & Margolis, 2002). Though Fodor correctly points out that whether (or to what extent) nativism is true is an empirical question, it seems very unlikely to most researchers that the empirical facts will bear the theory out. Furthermore, if the project is to naturalize content, then all radical nativism does is open up new questions. We are now owed a naturalistic account of how it can be the case that an individual is born with a large stock of mental states, potentially standing for entities that have never before existed, that already have the appropriate nomic connections. It seems unlikely that such a story is forthcoming (Scott, 2002a), and without such a story, all we have reduces to the

statement that SMURF means smurf because it has a set of properties that causes it to mean smurf.[18]

The solution of last resort, that is still consistent with at least a weakened version of Conceptual Atomism, is the pseudo-Russellian special case solution. This solution, proposed by Fodor (1990, p. 124) and Dretske (1981, pp. 222, 230), is to allow some concepts to be non-atomic, structured entities built out of atomic components. So the word "smurf" actually translates into a phrasal entity in the language of thought, assembled out of primitive atoms. That is, non-referring concepts on this view are definitions, or definite descriptions.[19] Fodor fails to provide any serious defense of the position, except to state that he thinks the situation in which a complex concept would be required is "*very, very rare*" (1990, p. 124, his emphasis). Dretske also balks at defending it: "I hope [the compositional solution] is sufficiently plausible not to *need* argument" (1981, p. 222, also his emphasis). But contrary to Fodor, non-referring terms are quite common. And contrary to Dretske, the definitional, or definite description solution is not sufficiently plausible not to need argument. (Again, there will be more on this in Chapter 3.) Furthermore, Fodor and Dretske are driven to this semantic solution by a problem internal to their separate problem of metasemantics.[20]

Like Taylor's proposal, Fodor and Dretske's proposals are also psychologically testable. While Taylor's solution predicts only a processing difference for non-referring concepts, Fodor and Dretske's solution predicts both processing and structural differences, both of which ought to be detectable in psychological experiments.

Predictions

So I now have, from formal semantics and philosophy of mind, a set of psychologically testable answers to the main questions posed by this dissertation. The Psychological Meinongian position predicts that there should be no structural or processing differences between referring and non-referring concepts. Taylor's pseudo-Russellian pragmatic solution predicts an identical psychological structure, but more involved processing in understanding utterances containing non-referring terms. Finally, Fodor and Dretske's pseudo-Russellian semantic solution predicts a radically different psychological structure between referring and non-referring concepts, plausibly leading to measurable processing effects as well. The next chapter looks at some of the possibilities from psychological research into conceptual structure, and lays out a basic strategy for the ensuing empirical investigation.

Chapter 3: Concepts and Categories

Gordie: All right, all right. Mickey's a mouse. Donald's a duck. Pluto's a dog. What's Goofy?

Teddy: Goofy's a dog, he's definitely a dog.

Chris: He can't be a dog. Wears a hat and drives a car.

Vern: God, that's weird. What the hell is Goofy?

– *Stand by Me* (Columbia Tri Star, 1985).

Categorization

The last chapter concerned one of the most important things we do with the concepts underlying natural language terms: We use them to refer to things. I looked at some of the lessons from the philosophical and linguistic literature on referring terms, ending with a number of opposing predictions about the psychological structure of non-referring concepts. This chapter concerns another of the most important things we do with concepts: We use them to categorize the world around us. I will look at some of the lessons from the psychological literature on concepts, ending, once again, with a number of predictions about the structure of non-referring concepts.

As far as the psychological literature on concepts goes, the questions posed in this dissertation about non-referring concepts are entirely new. As a result, this chapter will primarily be concerned with what is known about the structure of referring concepts. The game plan is as follows. First, I will review three different views of conceptual structure, as well as the evidence base for those views. The three views are the classical view that concepts have a definitional structure, and the two more recent views that concepts have either a prototype structure, or an ad hoc structure. Each of these views opens up

empirically verifiable possibilities for the structure of non-referring concepts. Based on that review, and the predictions from reference-based semantics reviewed in Chapter 2, I then develop two working hypotheses about non-referring concepts, one concerning structure, the other concerning processing. It is these two hypotheses that will be carried forward to the new empirical work on non-referring concepts described in Chapter 4.[21]

Concepts as Definitions

In Chapter 2, I mentioned that one of the reasons Russell's theory of disguised definite descriptions has been rejected by most philosophers and psychologists is the downfall of the definitional, or classical, view of conceptual structure. According to the classical view, a word can be exhaustively decomposed into a summary list of *necessary and sufficient* conditions for its application. The classic, and perhaps most successful, example of such a decomposition is the definition "a bachelor is an unmarried man". This definition implies two necessary and sufficient conditions for bachelorhood, namely being unmarried and being a man. The conditions are *necessary*, in the sense that an object must meet each of them in order to be correctly classified as a bachelor. Together, the conditions are also *sufficient*, in the sense that if an object meets both conditions then that is enough to truthfully apply the word "bachelor" to it. If this idea about words having definitions is interpreted in a psychological way, then the classical view amounts to the claim that the concepts underlying terms in natural language have a structure much like a short dictionary definition or definite description.

In philosophy, the classical view was the prevailing wisdom from antiquity until only recently, in part as an explanation for the apparent existence of analytic truths. Statements like (20) below feel like they are true by definition, or "analytic".

(20) All bachelors are unmarried.

More formally, a statement is an analytic truth if it is a logical truth, or if you can turn it into a logical truth by substituting synonyms. If "bachelor" is synonymous with "unmarried man", then the latter can be substituted into (20) to yield (21) without changing its truth value.

(21) All unmarried men are unmarried.

But (21) is a logical truth. It is a tautology to say that all objects with a certain property have that same property. Therefore, according to this story, (20) is an analytic truth as well – something that is true by definition.[22] In linguistics, the classical view served as the starting point for a number of classic works in semantics, for example by Katz (1966), Schank (1973), and Miller and Johnson-Laird (1976), to name but a few.[23] Today, the related practice of representing the meanings of words by decomposition into semantic primitives is still in use in many symbolic artificial intelligence and natural language processing systems (Allen, 1995; Wilks, Slator, & Guthrie, 1996), and still has some currency, though perhaps in a weakened form, among modern philosophers and linguists (e.g., Jackendoff, 1990; Sperber & Wilson, 1995).

Criticism of the Classical View

The classical view always had its critics, but it was not until the mid-20th century that it came under a sustained attack from philosophers, linguists, and psychologists alike, including Quine's attack on the notion of analyticity (Quine, 1953), Wittgenstein's attack on definitions (Wittgenstein, 1953), Rosch's ground-breaking empirical work on conceptual prototypes (to be reviewed in a moment), and Fodor's crusade against linguistic decomposition into semantic primitives (Fodor, 1970, 1981; Fodor, Fodor, &

Garrett, 1975; Fodor, Garrett, Walker, & Parkes, 1980). Many important criticisms of the classical theory were raised in the course of this attack, but three in particular involved a crucial contribution from experimental psychology (for more complete reviews see Laurence & Margolis, 1999; or Smith & Medin, 1981).

1. *Concepts don't seem to have definitions*. It seems that no one has ever managed to successfully define any interesting lexical concept. Despite a truly heroic effort by a number of philosophers and semanticists, very few concepts from ordinary language have been successfully decomposed into exhaustive and uncontroversial sets of necessary and sufficient conditions. Famously, Wittgenstein illustrated the futility of this kind of analysis for the object concept GAME (Wittgenstein, 1953), and Fodor did the same for the action concept PAINT (Fodor, 1981). Even the traditional example BACHELOR may not have as airtight a definition as was previously thought. The pope, Tarzan, and an 80-year-old widower are all unmarried men, but are they bachelors? Opinions on this vary. This problem with the lack of definitions was underscored by similarly unsuccessful attempts to find psychological evidence for a definitional structure for concepts. For example, Kintsch (1974) used a phoneme monitoring task to explore whether words that seemed to have more complex definitions would take longer to process. He found no such effects (see also the review by Fodor et al., 1975, as well as counter-arguments by Jackendoff, 1983).

2. *Concepts are fuzzy*. The classical view predicts that concepts will have a binary nature. An instance or exemplar either does or does not satisfy a list of necessary and sufficient conditions.[24] But the evidence shows that people tend to disagree on certain categorizations and will even contradict themselves on different occasions (Barsalou,

1989; McCloskey & Glucksberg, 1978; Medin, 1989). Is a rug furniture? How about a television? These and other cases of borderline category membership have lead many researchers to conclude that concepts have fuzzy boundaries rather than the sharp ones predicted by the classical view.

3. *Some exemplars are better than others*. The classical view further predicts that no instance or exemplar should be a better example of a given concept than any other. But this prediction can be falsified by the evidence as well. In a famous series of experiments, Eleanor Rosch and her colleagues showed that most people find it quite natural to rate some instances or exemplars as "better examples" or more "typical" of a given category than others. So whereas figs and apples both satisfy the concept FRUIT, an apple is judged by most North Americans to be a much more typical fruit, a much better exemplar for the concept FRUIT, than a fig (Rosch, 1973b, 1975). Note that this phenomenon is not the same as conceptual fuzziness. Almost everyone agrees that a penguin *is* a bird. It does not lie in some kind of fuzzy category border. But nevertheless, almost everybody also agrees that a penguin is not a particularly *good example* of a bird.

The Classical View Today

For the reasons cited above, among others, most philosophers and psychologists have abandoned the classical theory of conceptual structure as unworkable, at least on its own. But that's not necessarily the last word on the classical theory. A number of researchers still hold out the possibility that classical definitional information plays at least some role in conceptual structure, some for intuitive reasons (e.g., Armstrong, Gleitman, & Gleitman, 1983; Osherson & Smith, 1981), others because semantic decomposition, at least into necessary properties, can do a lot of work in explaining the generalizations that

drive theoretical approaches to semantics and pragmatics (e.g., Jackendoff, 1983, 1990, 1992). And of course, the default position in philosophy of language, as reviewed in Chapter 2, sometimes cashes out into a prediction of some kind of definitional structure, at least for non-referring words. So despite the evidence marshaled against concepts as definitions in general, the idea that non-referring concepts could have a definitional structure remains on the table.

Psychological Essentialism and the Classical View

One further way in which the definitional view of concepts has been kept alive is in the neo-classical interpretation of what has come to be known as the "psychological essentialist" view of conceptual structure (Medin & Ortony, 1989). The basic idea of Psychological Essentialism is that people assume that what binds together the members of a category is some kind of hidden essence. They may not know themselves what this hidden essence is, but they are sure there is one. For example, the common claim, "I may not know art, but I know what I like," seems to imply a belief that the concept ART has a definition or essence that, while it may not be known to the speaker, is presumably known to some expert in the community (cf. Putnam, 1975) or at the very least is *knowable* in principle (cf. Rey, 1983). This proposal is also quite compatible with the referential accounts of meaning reviewed in the last chapter (e.g., Laurence and Margolis, 2002),[25] but it has often been interpreted as lending support to some kind of classical, definitional theory based on critical properties (Malt, 1994; Pothos & Hahn, 2000).

In the original formulation of Psychological Essentialism, Medin and Ortony (1989) were careful to distance their proposal from two related claims. First of all, they were not endorsing *philosophical* essentialism. It is "not that *things* have essences, but

rather that people's *representations* of things might reflect such a belief" (Medin & Ortony, 1989, p. 183, their emphasis). Secondly, they were not claiming that people's concepts actually represent what the essence might be, but rather that people have, at minimum, an "essence placeholder" that may or may not be filled (Medin & Ortony, 1989, p. 184). The proposal is intuitively plausible (at the very least, it explains the tenacity of the classical view) and it also has some empirical support. For instance, to most children, a dog is still a dog after its fur is removed and it is painted with stripes, but it ceases to be a dog if its insides are removed and replaced with something else (see Keil, 1989, or Rips, 1989, for a review of similar results). From this type of evidence, it seems that the assumption of hidden essences plays a role in categorization behavior, even for young children.

The idea of Psychological Essentialism has received wide support, particularly among researchers studying lexical acquisition (Bloom, 2000), but some researchers have further interpreted the claim as implying that concepts contain explicit representations of critical properties or essences, much like the necessary and sufficient conditions of the classical view. This interpretation followed work by Rips (1989) that seemed to show evidence for the psychological reality of a particular critical property in categorization. Rips found that participants considered the ability to mate successfully with a member of a given species to be a critical property in determining whether an animal was a member of that species or not. Though Rips himself was cautious in his conclusions, many researchers took his finding as support for a new kind of definitional view of concepts in which the definitions are comprised of single critical properties rather than lists of necessary and sufficient conditions. But recent work by Pothos and Hahn (2000) has

shown that, in fact, mating ability is considered by most people to be a criterion for category membership that is neither necessary nor sufficient. Similar negative results concerning essences have been obtained for other biological and non-biological categories as well (Ahn, Kim, Lassaline, & Dennis, 2000), including the philosophers' favorite example, WATER, which is supposed to be essentially H_2O (Malt, 1994). These results do not contradict the original formulation of Psychological Essentialism, but they do serious damage to the neo-classical interpretation.

Psychological Essentialism, at least in its non-classical interpretation, constitutes another possibility for the structure of non-referring concepts, namely that these concepts have explicitly represented essence placeholders. In fact, the hypothesis is interesting, albeit quite counter-intuitive. It may be that people believe that things like dogs and hammers have some kind of hidden essence that puts them in their categories, and this belief may be reflected in the structure of the underlying concepts – but could it be that people believe that things like sea monsters and time machines have such hidden essences as well, and that even if they don't themselves know what that essence is, they could appeal to a sea monster or time machine expert for the answer? This prediction is an odd one, but odder things have been predicted and have turned out to be true. At the very least, essences, along with definitions, remain a theoretical possibility for the structure of non-referring concepts.

Concepts as Prototypes

Eleanor Rosch and her colleagues provided some of the first solid empirical evidence that the structure of at least one type of concept, namely those that are about concrete objects, might not be classical in nature. Rosch instead proposed a new kind of conceptual

structure based on concept *prototypes* to explain many of the experimental effects she was finding in categorization tasks. Since most non-referring concepts encode object categories, Rosch's work on object categories provides the first prediction for the structure of non-referring concepts that has strong empirical support – namely, that they will turn out to be structured as prototypes as well.[26] Her work also provides a model experimental approach, elements of which will be taken up and used in the empirical work described in Chapter 4.

Perceptual Prototypes

The insights that originally led Rosch to her ground-breaking research came from cross-cultural investigations into the nature of color and form categorization. Berlin and Kay (1969) demonstrated that there are cross-cultural regularities in how human beings carve up the perceptual color space for categorization. Though a given language might have as few as 2 or as many as 11 basic color terms, the extensions of these terms are not arbitrary. The basic color terms are centered around certain "focal" hues that are common to all languages, and cross-linguistic evidence also suggests that new color terms are added to a language in a constrained sequence. A language with 2 color terms will encode "dark" and "light", a language with 3 will add a term for red, and so on up to the observed maximum of 11 basic terms found in most European languages.[27]

Rosch showed that focal colors both attract attention and are remembered more easily than non-focal colors (Heider, 1971, 1972).[28] In later work with the Dani, a stone-age people with only two basic color terms in their native language and no terms for geometric shapes, Rosch (1973a, 1973b) showed that artificial categories based on the focal colors identified by Berlin and Kay and so-called "good forms" from Gestalt

psychology (circle, square, triangle, etc.) were more easily learned than categories that were not. Furthermore, when learning categories that were not built around focal colors and good forms, Dani participants tended to distort the categories they learned towards these more perceptually salient "natural prototypes" of color and form.

Typicality Effects

The discovery of what appeared to be natural prototypes in the domains of color and form raised questions as to whether other semantic categories, in particular, object concepts such as BIRD and DOG, might also be structured as prototypes rather than as definitions (as the classical view would have it). The first type of evidence that this might be the case was the discovery of typicality effects in categorization. In a paper that has now become a classic, Rosch (1973b) first asked a group of American college students to rate words representing exemplars of the categories FRUIT, SCIENCE, SPORT, BIRD, VEHICLE, CRIME, DISEASE and VEGETABLE[29] using a 7-point scale to indicate how well each exemplar matched their "image or idea" of the category.[30] The participants not only performed the task without hesitation, giving different ratings to different exemplars, but also showed a high level of agreement in their ratings of the exemplars. Rosch's basic findings with words as exemplars were also supported in more detailed follow-up work by Rosch (1975) and in a number of other studies since (e.g., Barsalou, 1985; Hampton & Gardiner, 1983; Malt, 1994; McCloskey & Glucksberg, 1978). The same pattern of results was also obtained using pictures of instances instead of words (Rosch & Mervis, 1975) and in artificially constructed categories of various types (Neumann, 1974; Reed, 1972; Rosch, Simpson, & Miller, 1976).

A second experiment in Rosch's original study (Rosch, 1973b) provided further evidence for the psychological reality of typicality. In this experiment, the results of which have been reproduced many times (Chang, 1986), Rosch tested participants' categorization speed for more and less typical exemplars of a given class. Participants were presented with sentences of the form "An [exemplar] is a [category]", and were to press buttons labeled "true" or "false" accordingly. High and low typicality stimuli were matched for word frequency, though not for word length. Both adults and 9- to 11-year-old boys were tested, and the results showed that for both groups, "true" responses were slower for sentences involving less typical exemplars ("A fig is a fruit") than for those containing more typical exemplars ("An apple is a fruit"). In contrast, there was no effect in the "false" reaction times for sentences involving less typical or more typical exemplars (e.g., "A fig is a crime" vs. "An apple is a crime" – in this case the typicality level is relative to the typicality of the items within their proper category, FRUIT, not the foil category, CRIME). Finally, children, but not adults, made significantly more errors with less typical items, suggesting that typical exemplars tend to be those that were acquired first in concept learning. Rosch (1975) also found similar reaction time effects using a task in which adult participants were primed with a category name, and then responded "same" or "different" to pairs of word or picture stimuli, indicating whether or not they belonged to the same or different categories. In these studies, participants were faster to respond when the stimuli were typical category members than when they were atypical. Finally, and in anticipation of the evidence for the notion of "family resemblance" to be reviewed in a moment, this basic pattern of reaction time data was also confirmed in a study by Rips, Shoben, and Smith (1973), who correlated reaction

times with participants' ratings of semantic relatedness between category and exemplar words. In summary, Rosch and others found that certain category exemplars are consistently rated more typical than others, and that these ratings can consistently be used to predict the trends in reaction time data on various tasks.

Converging evidence for the psychological reality of typicality effects has also come from other areas of research. Kahneman and Tversky (1973), followed by Rips (1975), showed that participants' inductive judgments about category exemplars show typicality effects as well. People are more ready to generalize from typical exemplars to atypical exemplars than the other way around. For example, if told that all the robins (typical birds) on a fictional island have developed a new disease, participants are much more likely to predict that the ducks (atypical birds) on the island will catch it from the robins than they are to predict that the robins will catch the disease from the ducks when the roles are reversed. Cherniak (1984) found that under time constraints, even participants' deductive reasoning abilities can be affected by prototypes. Given a deductive proof task (e.g., "All A's are robins, therefore all A's are birds: true or false?") and limited time, participants tended to work through an example chosen for high typicality rather than to produce a formally correct proof. Kelly, Bock, and Keil (1986) showed that typicality can affect sentence production as well. In sentence recall tasks, participants tended to change the order of items so as to place the more typical exemplars ahead of the less typical ones, and judged sentences to be more natural if the exemplars in the sentence were ordered according to their typicality.

Are Participants Really Rating Typicality?

A very common initial reaction to the phenomenon of typicality ratings is to suggest that experimental participants might be providing ratings based on criteria other than typicality. One early suggestion was that participants may have rated the exemplars based on how much they personally liked them. But Rosch (1973b) showed that when asked to rate how much they liked or disliked each exemplar, participants produced a different pattern of ratings and showed a high level of general disagreement about them.

Another suggestion is that participants may be rating their familiarity with the exemplars – that is, their frequency of exposure to words or instances from the categories. To rule out frequency of exposure to the words, Mervis, Caitlin, and Rosch (1976) computed Spearman rank-order correlations between participants' typicality ratings and the Kučera and Francis (1967) estimates of written word frequency, and found no evidence for a correlation. To rule out frequency of exposure to category instances, Rosch, Simpson, and Miller (1976) experimented with artificial categories built around visual prototypes. Instances of a category were generated by modifying a visual prototype in various minor ways, and participants learned a concept related to the category by controlled exposure to these instances. They found that after the learning phase, participants rated the prototype instance higher in terms of typicality than any other instance of the same category, even though they had not actually seen the original prototype during the learning phase of the experiment. A similar result, also involving artificial categories, was obtained by Neumann (1974). This type of finding was taken as evidence that, at least for artificial categories, frequency of exposure does not play a role in concept learning.

However, in contrast to the results for artificial categories, experiments with natural semantic categories have occasionally turned up a correlation between instance familiarity and typicality. Ashcraft (1978a, 1978b) found that the "property dominance" correlated with typicality ratings, where the property dominance of an exemplar is defined as a function of the number of properties participants could list for it. Under the assumption that participants will be able to list more properties for exemplars they are more familiar with, Ashcraft concluded that typicality effects were actually an effect of familiarity. Furthermore, McCloskey (1980) in reanalyzing and extending the experimental results of Smith, Shoben, and Rips (1974) found that stimulus familiarity, measured in terms of reaction times and participants' ratings, was confounded with semantic relatedness, a psychological variable quite similar to typicality. More recently, Janczura and Nelson (1999) have demonstrated that the frequency of very recent experience with an exemplar seems to correlate with its typicality rating in a subsequent test, suggesting that frequency effects can be selectively amplified.

But a number of other studies have also come to different conclusions. A careful study by Malt and Smith (1982) concluded that, whereas familiarity may play a role, it cannot explain the entire pattern of typicality ratings. They did find correlations between familiarity and typicality for some of the categories they studied (values ranged from -0.06 for FISH to 0.62 for FRUIT and VEHICLE), but only when the lists of exemplars presented to participants consisted of randomly selected exemplars from lists generated by a separate group of participants (Battig & Montague, 1969). In contrast, when the lists of exemplars were manipulated by the experimenters to consist only of those that were likely to be known to all the participants, familiarity did not seem to play a role in

typicality ratings. These results suggest that familiarity can have an effect on typicality judgments, but one that is only noticeable in the presence of large familiarity contrasts. Clearly, there is more to typicality ratings than simple familiarity or frequency of exposure. Malt and Smith (1982) also pointed out that an exemplar's typicality can change relative to its category. For instance, "chicken" was rated by participants in the Rips et al. (1973) study as more semantically related to (i.e., more typical of) the concept ANIMAL than BIRD, while the reverse was true for "robin". This phenomenon should not occur if familiarity with either the word "chicken" or actual chickens was the only determining factor of their typicality or semantic relatedness.[31]

A number of other independent studies have also found minimal or negligible correlations between familiarity and typicality. A study on British participants by Hampton and Gardiner (1983) found that neither familiarity ratings nor property dominance correlated with typicality ratings. A study by Barsalou (1985) found that familiarity played little role in accounting for typicality effects in object categories such as DOG, though a related measure, frequency of instantiation, did play a role in accounting for such effects in "goal-directed" or "ad hoc" categories such as THINGS TO DO FOR WEEKEND ENTERTAINMENT (a result to be discussed below). All this converging evidence seems to suggest that the roles played by various types of instance and word familiarity cannot explain the entire pattern of typicality results for the object categories studied by Rosch and her team. Typicality effects may sometimes be confounded with familiarity effects, but they cannot be entirely explained by them.

Family Resemblance

To sum up so far, Rosch and her colleagues found that participants willingly provide consistent and graded typicality ratings for lexical concepts, that these typicality ratings cannot be explained away simply as frequency or familiarity effects, and that typicality predicts a number of performance effects, such as reaction time in categorization. Other researchers subsequently demonstrated the existence of a number of other psychological phenomena that also appear to be influenced by typicality. But none of these findings suggest a positive account of what the causal determinants of the typicality effects might be. One of the first studies to shed light on this question was Rosch and Mervis' (1975) classic work on what, following Wittgenstein, they called "family resemblance". It was in this work that the first real evidence was found indicating that the concepts corresponding to object categories might be structured around similarity, and it has been claimed that these results suggest that concepts are structured as prototypes (e.g., Hampton, 2000; Rosch & Mervis, 1975; or see the review by Laurence & Margolis, 1999).[32]

An important precursor to Rosch and Mervis' work on family resemblance was work being performed at the same time, also by Rosch's team, providing evidence for different psychological levels of categorization (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). "Basic" level categories are those that tend to jump to mind for a given object, such as CHAIR and APPLE, while the corresponding "superordinate" (higher-level) categories, FURNITURE and FRUIT respectively, seem less salient at first. There are also "subordinate" (lower-level) categories like KITCHEN CHAIR and CRAB APPLE which are also less salient than the basic level. Like typicality effects, the existence of a basic category level is a robust and intuitive result, and further evidence has tended to refine

rather than contradict the general picture developed by Rosch's team.[33] The exact details of this work on category levels are not crucial to what follows, but the notions of basic and superordinate categories are important as they come up repeatedly in what follows and in future chapters as well.

In their classic family resemblance study, Rosch and Mervis (1975) asked participants to produce lists of properties for 20 basic level exemplars in each of six superordinate level categories taken from an earlier study by Rosch (1975): FURNITURE, VEHICLE, FRUIT, WEAPON, VEGETABLE, and CLOTHING. Each attribute that was listed in a superordinate category by at least one participant got a score from 1 to 20 representing how many exemplars in that superordinate category were assigned that attribute. Then each exemplar was assigned a "family resemblance" score based on the sum of its attribute scores. For example, suppose the only properties any of the participants could think of for "apple" were "sweet" and "round". Now suppose that the attribute "sweet" had been listed by at least one participant for 15 of the 20 fruits, while "round" had been listed for 10 of the 20 fruits. Then "apple" would receive a family resemblance score of 15 for the attribute "sweet" plus 10 for the attribute "round" for a total family resemblance score of 25. Rosch and Mervis correlated these family resemblance scores with the previously obtained typicality ratings (Rosch, 1975) and found high Spearman rank-order correlations for all categories (scores ranged from .84 to .93, $p < .001$). A later experiment by Malt and Smith (1984) replicated these basic findings for both superordinate and basic level categories.

Rosch and Mervis (1975) hypothesized that not only would a good member of a superordinate level category have a high family resemblance to other exemplars of the

category, but it should also have a low resemblance to members of contrasting superordinate level categories. This was verified in a task in which participants listed three ranked superordinate categories for a given basic level exemplar. A "category dominance" score was then computed for each exemplar which measured the difference between the number of times the main superordinate category was listed compared to the next two most frequent listings. The idea was that a more typical exemplar should be more salient in its one true superordinate category than in any other. Therefore, the salient category should be listed first most often, and other categories should be listed with a very low frequency. On the other hand, a less typical exemplar might be a little more salient for other categories as well, and would be listed in those categories with a slightly higher frequency. So the prediction was that more typical exemplars should have higher category dominance scores, and indeed, the category dominance scores correlated strongly with typicality as well (with Spearman rank-order correlations ranging from .71 to .83, $p < .001$).

Rosch and Mervis' study provided some of the first strong evidence that similarity, possibly to some kind of prototype, binds some categories together. But later work by Malt and Smith (1984) showed that the family resemblance picture may actually be a bit more complicated. Rosch et al. (1976) had suggested, but not pursued, the idea that properties in natural categories tend to occur in clusters, and might therefore be correlated to one another in the structure of the corresponding concepts. For example, animals with feathers tend to have beaks and wings. Hence, in categorization and the computation of family resemblance, participants may be able to make use of such property correlations. Malt and Smith (1984) later provided evidence for the existence of

property clustering. They found a number of correlated pairs in participants' lists of properties for various categories, and also found some limited evidence that these correlations might be important to typicality judgments. They went on to suggest the possibility that a single category could have multiple sub-clusters of properties that affect typicality judgments in non-linear ways.

Abstract Concepts as Prototypes

So far, most of the work reviewed has concerned concrete object concepts, that is, concepts which are satisfied by concrete objects, such as BIRD and FURNITURE. But the family resemblance studies only ever involved concrete object concepts. There are lots of abstract concepts that, while they do refer, do not refer to concrete objects. Examples of such categories include NUMBER, SONG, CRIME, JOB, and so on. (Though its use is a little vague, the term "abstract concepts" is usually meant as a catch-all, applying to all ordinary concepts that are not satisfied by concrete objects.) Soon after the work of Rosch and her team on family resemblance in concrete object concepts, James Hampton (1981) addressed the question of whether or not abstract concepts showed the same sort of effects.

Hampton (1981) first had a group of participants assign exemplars of each category to four different levels of typicality. Then a second group produced property lists for the exemplars. The most frequently mentioned properties were assembled into a subset, and a third group of participants provided ratings indicating how true they felt each property was of each exemplar. The resulting property ratings were combined to give a property score for each exemplar, which was analogous to the family resemblance score used by Rosch and Mervis. Like family resemblance, Hampton's property score is a

rough subjective estimate of how similar each exemplar is to other exemplars in the category. Then the property scores were used to place the exemplars into four ranked groups, with the top group containing the items with the highest property scores and the bottom group containing the items with the lowest property scores, and tested how well these groups matched the groups formed on the basis of typicality. Although this methodology was somewhat different from that used by Rosch and Mervis, Hampton (1979) had previously used it to find a significant match between property score levels and typicality levels for concrete object concepts, thus both supporting the conclusions of Rosch and Mervis and validating his own methodology.

The abstract concepts Hampton (1981) selected to investigate were A WORK OF ART, A BELIEF, A CRIME, A JUST DECISION, AN INSTINCT, A RULE, A SCIENCE, and A KIND OF WORK. Hampton found statistically significant differences in the property scores across the four typicality levels for five of these eight concepts. For the remaining three categories (A RULE, A BELIEF, and AN INSTINCT), the tests were not significant, though the trend was still in the right direction. Hampton argued that although he had failed to show a significant prototype effect for these three concepts, the possibility that they have a definitional structure was not supported by the data either. In the end, though, Hampton was forced to concede that prototype structure had not been "convincingly demonstrated" for abstract concepts in general. Hampton's work thus suggests that there are boundaries to the generality of Rosch's prototype model of conceptual structure, and that certain types of concepts may fall on either side of that boundary. The interesting question, from the point of view of this dissertation, is which side of the boundary non-referring concepts will fall on.

The Importance of Rosch's Work

Rosch and her team found evidence in support of the conclusion that ordinary object concepts are structured psychologically as prototypes rather than classical definitions. The graded typicality of exemplars in categorization is now one of the most robustly established results in the field of cognitive psychology. Participants easily take to the task of providing typicality ratings, and show a high level of agreement in their results, whether the stimuli are words for category exemplars or pictures of category instances. Typicality ratings for most lexical concepts are not reducible to familiarity or frequency effects, and they can be used to account for a wide range of other psychological phenomena, including reaction times on categorization tasks, grammatical intuitions, and some aspects of performance on inductive generalizations and deductive proofs. Typicality effects are found at all categorization levels and correlate with independent measures of the so-called family resemblances that bind a category together, thus supporting a new view of conceptual structure based on similarity to a prototype. Hampton extended these findings to show that some, but not all, "abstract" concepts show the same type of relationship between typicality and family resemblance.

Now much of the necessary information is in place to begin forming a working hypothesis regarding the structure of non-referring object concepts – namely that they will show the same sort of evidence for prototype structure as their referring counterparts. On the other hand, from the work on abstract concepts, I also found reason to believe that non-referring object concepts might fail to show that structure. The prototype hypothesis will be discussed and refined considerably in a moment, but before I get to that, there is a third important proposal about conceptual structure that needs discussion.

Ad hoc Concepts

Lawrence Barsalou raised a well known and often cited challenge to the importance of Rosch's work (Barsalou, 1987, 1989). According to Barsalou, typicality judgments are inherently unreliable both within and between participants, and inherently unstable across semantic contexts. He concludes that concepts are an "analytic fiction" – that is, the assumption underlying most accounts of conceptual structure both before and after Rosch's work, that concepts are mental representations with a content and structure that varies little over time, is false. According to Barsalou, what we think of as concepts are actually representations that are constructed as needed out of a partly context-dependent subset of our full knowledge. They are not representations with invariant structures that are retrieved intact from memory. This line of reasoning eventually culminated in Barsalou's empiricist proposal for a "perceptual symbol systems" model of human cognition, in contrast to the "amodal" theories favored by cognitive science (Barsalou, 1999; see also Prinz, 2002).

Commenting on Barsalou's most recent proposals would be outside the scope of the current work, which is situated well within the amodal representational tradition. But Barsalou's earlier work on categorization is important to this dissertation for at least two reasons. First of all, he demonstrated convincingly that typicality ratings for different kinds of categories do not always have the same structural determinants. This consideration will be important in developing the experimental plan for Chapter 4. Second, he demonstrated the existence of a previously unrecognized type of concept, namely ad hoc concepts. Ad hoc concepts are those that underlie ad hoc categories. Rather than being static and encoded in the lexicon, like MAFIA and SMURF, ad hoc

concepts are thrown together for a particular purpose and expressed phrasally, such as WAYS TO AVOID BEING KILLED BY THE MAFIA or THINGS TO BRING ON A SMURF HUNT. It turns out that the concepts underlying ad hoc categories are quite different in structure from the object concepts studied by Rosch and the abstract concepts studied by Hampton. Barsalou's work thus contributes another set of possibilities for the structure of non-referring concepts. Barsalou's eventual position was that concepts are an analytic fiction, or to put it another way, that *all* concepts have a changeable, ad hoc structure – not just the ones that underlie what he had previously labeled "ad hoc" categories. I think that although the original work on ad hoc categories was interesting and useful, the attempt to extend the results to all other concepts does not go through. The next sections will outline and respond to Barsalou's three main arguments for his radical conclusion. These arguments concern: 1) variable participant behavior; 2) contextual effects on concepts; and 3) multiple determinates of typicality ratings. Along the way, I will highlight both the methodological lessons of Barsalou's work, and two new possibilities for conceptual structure.

Variability

Barsalou's first claim is that the typicality ratings, on which Rosch based much of her work, are unstable both within and between participants. Barsalou (1987) reports that participants seem to agree with each other less than Rosch originally reported, and even seem to disagree with their own ratings on different occasions. Reviewing his own unpublished work, Barsalou reports that the mean between-participant correlations for 20 different groups ranged from .30 to .60. Rosch's (1975) original reports of high levels of between-participant agreement reflected a different, possibly misleading statistical

approach in which the stability of the mean was measured by split-group correlation. The problem with this type of statistic is that increasing group size tends to increase the stability of the mean, potentially producing an arbitrarily high appearance of inter-participant agreement. In contrast, Barsalou's statistics resulted from calculating between-participant correlations for every possible pair of participants and then computing the mean of these correlations to predict the mean level of correlation between any two participants' typicality ratings. Even more potentially damaging is Barsalou's report (again referring to his own unpublished work) that participants' typicality judgments can differ on different occasions. He had groups rate typicality on the same set of items twice, with varying delays of an hour to a month. With the exception of the 1-hour delay, the individual participants' ratings in the two sessions correlated to a mean level of .80. Barsalou concluded that typicality judgments are too unstable to be useful in determining the underlying structure of a concept.

Related studies by Bellezza (1984a, 1984b, 1984c) also lend general support to Barsalou's results. Bellezza showed that participants tend to be highly variable in their performance on two types of production tasks related to categorization: exemplar listing for categories (Bellezza, 1984a) and property listing for nouns and proper names (Bellezza, 1984b, 1984c). In this series of experiments, participants performed an identical task on two separate occasions, one week apart. The degree of overlap between the two sessions on the exemplar-listing task, as measured by a common-element correlation, averaged .69 within participants and .44 between participants. For the property listing tasks, the numbers were approximately .50 within participants and .20 between participants.

What should we conclude from all this? First of all, the fact that participants disagree with each others' typicality ratings is not particularly surprising, and does not necessarily invalidate the mean ratings over a large group of participants, particularly when these mean ratings prove to be predictive of other psychological effects. Experimental psychology is a statistical science designed to measure tendencies across large groups. Between-participant variance of the kind Barsalou reports is neither unusual nor troubling. Nor should it be particularly troublesome that participants show a high variability in production tasks, which are probably affected by all sorts of subtle priming effects as well as the participants' moods and preoccupations on the day tested. It is only the finding of within-participant disagreement on the typicality ratings in particular that might really pose a problem. If participants can't even agree with themselves on different occasions, then maybe typicality ratings do not result from the structure of relatively invariant concepts after all.

But there are two reasons to doubt the significance of the problem. First of all, within-participant reliability for typicality, as measured by Barsalou, seems significantly higher than within-participant reliability on Bellezza's production tasks. This difference shows that participants are more variable in free recall than they are in rating items, which is exactly the right sort of trend if the act of rating exemplars causes the participants to access some kind of invariant underlying structure. Secondly, Barsalou (1987, 1989) reports that high- and low-typicality items showed little variance within participants. Rather, the majority of the within-participant disagreement was on the intermediate-typicality examples, a result similar to that found by McCloskey and Glucksberg (1978) in their study on fuzzy category membership. This phenomenon may

simply reflect an artifact of the way the participants were using the rating scale. If participants tend to rate typical exemplars very high on the scale and atypical exemplars very low, then that leaves them with a lot of room to maneuver in the middle section of the 7- or 9-point scales that are usually used, and they can be expected to produce different values in the middle range on different occasions. Later work by Barsalou (1989) also supports this interpretation. When participants rank exemplars in order of typicality, they show less variance, both within- and between-participants, than when they use a rating scale. So the reliability of typicality ratings may not be a major cause for alarm after all. Barsalou raised a useful criticism of the statistics used in the earlier work on typicality effects, and one that should be kept in mind for the current work, but he may have over-emphasized its importance.

Context

Barsalou's second claim (Barsalou, 1987; also discussed in Barsalou, 1989, and Barsalou & Medin, 1986) is that sentential and situational context will affect typicality judgments. As evidence, he cites Roth and Shoben (1983), who found that when a concept word appears in a sentence, the sentential context affects participants' typicality ratings. For example, participants were shown the sentence, "During the midmorning break the two secretaries gossiped as they drank the beverage" and were then asked to use a 9-point scale to rate how well given category members fit their "idea or image of what the category term [i.e., beverage] refers to in the sentence" (Roth & Shoben, 1983, pp. 358-359). The results showed that normally atypical exemplars for BEVERAGE, such as "coffee" or "tea" were rated as much more typical in context than when no context was provided. In other experiments, Roth and Shoben also found context-dependent

differences in reaction times that were consistent with a shift in typicality across contexts. Barsalou (1987; Barsalou & Medin, 1986), again reviewing his own unpublished work, reports that typicality judgments for common semantic categories can change when participants are instructed to take another person's point of view. For example, American students judged robins and sparrows to be highly typical birds from an American point of view, whereas swans and peacocks were judged to be highly typical from a Chinese point of view.

The problem with these types of studies is that it is not at all clear that the concept itself is the same across contexts. It seems quite likely that participants are inferring what Barsalou might call ad hoc concepts such as BEVERAGE THAT SECRETARIES DRINK IN THE MORNING or BIRD FROM A CHINESE POINT OF VIEW and rating typicality against those concepts rather than the target concepts BEVERAGE and BIRD. If so, then these studies do not in any way prove that conceptual content is dependent on context. It would be inappropriate to conclude that context *can't* play any role,[34] but no evidence to date conclusively shows that the concepts themselves can actually change in context.

Ad Hoc Categories

The evidence in support of Barsalou's third claim is the most interesting for the current work. Barsalou (1983, 1985) had previously found typicality effects in ad hoc and goal-directed concepts, such as THINGS TO DO ON THE WEEKEND.[35] As discussed earlier, typicality effects, both in object categories and some abstract categories, have been shown to correlate consistently with family resemblance, rather than any simple measure of frequency or familiarity. But this relation does not seem to be a valid possibility for goal-directed categories, whose exemplars tend to bear little resemblance to one another.

For example, what do money, children, jewelry, photo albums, and pets have in common? On the face of it, they seem to have very little in common at all, but they are all intuitively good exemplars of the category THINGS TO SAVE FROM A BURNING HOUSE. On the other hand, chairs, televisions, and pots do not seem like such good exemplars of this category, even though they could all conceivably be rescued as easily as the good exemplars. What, then, are the sources of typicality effects in these domains?

To answer this question, Barsalou (1985) looked at four possible predictors of graded typicality judgments for a given category exemplar:

1. *Central tendency*, a measure that is somewhat analogous to family resemblance, but is derived from participants' direct ratings of the similarity between pairs of exemplars rather than on the degree to which properties are shared among exemplars,
2. *Closeness to an ideal*, such as *zero calories* for the category THINGS TO EAT ON A DIET,
3. *Frequency of instantiation*, measured by participants' own estimates of how often an exemplar occurs as a member of a given category, and
4. *Familiarity*, measured by participants' own estimates of how familiar they were with an exemplar.

In accordance with many earlier studies on family resemblance, Barsalou found that central tendency was most predictive of typicality for object concepts[36], but that both frequency of instantiation and closeness to an ideal were much more predictive of typicality for the goal-directed concepts. Of course, this result is easily explainable given that there is no reason to believe that ad hoc concepts are in any way psychologically basic. Whereas the object concepts he studied are in common use and tend to be lexicalized, goal-directed concepts are uncommon and generally have to be expressed

using long phrases. It is quite possible that these concepts function just as Barsalou suggests. That is, they have little or no invariant structure, but rather are constructed on the fly when participants' attention is called to their possibility. But the predictive value of family resemblance and central tendency seems to argue against the same conclusion for lexicalized object concepts.

However, Barsalou's main point was to prove that the determinants of graded structure can change. Having established that ideals can determine graded structure for certain categories, he constructed an experiment in which two groups of imaginary individuals were structured so that they could be sorted equally well by overall family resemblance or by considering a single property dimension. Each individual was represented by a list of values for five properties, encoding how often (daily, weekly, or monthly) they engaged in activities such as dancing, renovating houses, writing poetry, watching movies, and so on (Barsalou, 1987, p. 642). The target property dimensions for the two categories were how often the individual read the newspaper and how often they jogged – the two groups of individuals could be sorted equally well by overall family resemblance, or by separating them into groups based on how often they jogged and read the newspaper. In the "ideals" condition, participants were told that the categories were CURRENT EVENTS TEACHER and PHYSICAL EDUCATION TEACHER. In the "central tendency" condition, they were told that the categories were Z PROGRAMMER and Q PROGRAMMER. After studying the descriptions of individuals in the two categories, the participants were asked to rate them for typicality. Not surprisingly, in the ideals condition, participants' ratings correlated with the values of the relevant defining dimension (that is, how often they jogged or read the newspaper). In the central tendency condition, their ratings

correlated with overall family resemblance. Barsalou concluded that this was further evidence for the flexibility of conceptual structure.

The main problem with this study was that the participants in the ideals condition were not learning arbitrary concepts from scratch, like the participants in the central tendency condition. Suppose, as is quite likely, that the participants already had pre-existing concepts for PHYSICAL EDUCATION TEACHER and CURRENT EVENTS TEACHER that were organized around family resemblance principles. The participants were then placed before descriptions of individuals that gave values for a mere five property dimensions. It is quite likely that very few of these dimensions would seem relevant to categorization decisions based on these pre-existing categories. (Does a physical education teacher watch more or fewer movies than a current events teacher?) Under these conditions, it's quite plausible that only the target properties of the individuals seemed salient given the category's pre-existing central tendencies. If so, the fact that the participants appeared to sort using that property says nothing about the general organization of their concepts. In contrast, participants in the central tendency condition were learning two new (and quite closely related) categories, and thus approached the problem as a new learning situation, basing their typicality judgments on all the properties at once.

The Importance of Barsalou's Work

In the end, none of Barsalou's three conclusions are warranted, hence neither is his suggestion that invariant concepts are an "analytic fiction", or to put it another way, that *all* concepts have an ad hoc structure containing very little invariant information.[37] But his finding of graded structure in ad hoc categories, and their non-prototype causal determinants, is very important. First of all, it seems that participants have a strong

intuition that exemplars and instances in any category, regardless of the underlying conceptual structure, can be ordered by typicality.[38] But the information that participants access when grading exemplars for typicality may not be the same across different types of categories. In some cases, typicality judgments for ad hoc concepts were found to be predicted by closeness to an ideal, while in others they were predicted simply by a measure of the frequency of instantiation of the exemplar in the target category. On the other hand, typicality judgments for lexicalized object concepts have consistently been shown to be predicted by family resemblance. These different findings may reflect a different structure for the concepts underlying these different types of categories.

As an aside, there is a simple account of typicality that unites all the findings reviewed so far, namely that typicality ratings represent the ease with which participants can judge an item to be a member of the category. In prototype categories, this judgment is linked to how similar the exemplar is to some kind of conceptual prototype. In ad hoc or goal-directed categories, it is linked either to how close the exemplar is to an ideal value along a single dimension, or simply to how easily information about the exemplar can be accessed (i.e., how frequently it is encountered as a member of the category). Thus Barsalou's work, while failing to invalidate the very idea of concepts, or to establish that all concepts have an ad hoc structure, does raise the possibility that because non-referring categories are slightly unusual and subject to much greater variation between individuals, the concepts underlying them will be structured like the concepts underlying ad hoc categories.

Five Possibilities, Two Hypotheses

At this point, there are five distinct accounts of conceptual structure on the table. In this section, I will formulate and defend two working hypotheses about the structure of non-referring concepts to be carried into the next chapter, where they will frame my own experimental work on non-referring concepts.

Five Possibilities

The five accounts of conceptual structure currently on the table are:

1. *Definitions.* Not many people support the idea that classical, definitional structure is constitutive of more than maybe a few concepts, but some linguists and philosophers still regard the role of definitions in conceptual structure as an open empirical question, and the default position from reference-based semantics is that non-referring concepts do have this kind of structure.
2. *Essences*. The idea here is that people's conceptual structure reflects their belief that categories have essences that bind them together, and that this belief is reflected in the presence of an essence placeholder in the underlying conceptual structure. There is conflicting evidence for the psychological reality of essences, but the possibility remains that non-referring concepts might have essence placeholders.
3. *Prototypes*. Rosch's team (supported by the work of Hampton, Malt and Smith, Barsalou, and others) convincingly demonstrated a structural role for similarity in determining typicality judgments both for object concepts and for some abstract concepts. Most of the concepts they studied were encoded in the lexicon as single words. The default interpretation of this evidence is that these concepts are structured as prototypes (more on what this means in a moment). Because many non-referring

concepts are similar to object concepts (ZOMBIE, GRIFFIN, etc.), prototype structure is a natural prediction for these concepts as well.

4. *Ideals*. Barsalou found that many ad hoc concepts are rated for typicality according to their similarity to an ideal value along a single dimension. It could well be the case that some non-referring concepts are structured in terms of ideals as well. For example, the single, continuously graded property *strong* could determine how good an example something is of the category SUPERHERO.
5. *Ad hoc structure*. Hampton was unable to demonstrate any kind of structural determinant for some of his abstract concepts, and Barsalou found that frequency of instantiation, which is unlikely to be a *structural* variable, was the best predictor of typicality for some of his ad hoc concepts. Hence, these concepts don't seem to be structured according to definitions, essences, prototypes, or ideals. Perhaps they are unstructured, or have a variable structure, or perhaps they have some kind of structure that has yet to be discovered. Non-referring concepts could be something like these ad hoc and abstract concepts. For now, I'll just refer to this possibility as the possibility that non-referring concepts have an "ad hoc" structure.

Hypothesis I (First Try)

Recall Question 1 from Figure 1-2: What is the psychological structure of non-referring concepts, and does it differ in any way from ordinary, referring concepts? The last section outlined five possibilities for that structure, found in the psychological literature on concepts. Now, here's a first stab at a hypothesis about the structure of non-referring concepts, following from the third possibility:

> Hypothesis I (first attempt): *Non-referring object concepts are prototypes, just like referring object concepts.*

Evidence for a hypothesis like this would be evidence favoring one of the five structural possibilities, and disfavoring all the others. It would be (arguably non-conclusive) evidence against definitional structure and quite good evidence against concepts based on essences, ideals or ad hoc structure, at least insofar as all of the hypotheses under consideration are intended to be exhaustive descriptions of conceptual structure. It would also be evidence favoring Psychological Meinongianism and disfavoring special case solutions to the problems of non-referring terms reviewed in Chapter 2. As it stands, however, this first formulation is much too strong, for reasons that will emerge below. But before refining it, I will explain why I chose a hypothesis of this form, as opposed to one about, say, definitions or essences. Why focus on prototype concepts?

The initial working hypothesis focuses on prototypes for two main reasons. First of all, most non-referring concepts seem to be about nonexistent but otherwise concrete objects – things that, if they existed, would be concrete objects. These are things such as goblins, smurfs, superheroes, Pokémons, and the like. In contrast, although there are certainly non-referring abstract concepts, such as the concept of a whole number between four and five, there don't seem to be very many big and well-established categories for these sorts of entities. So the empirical work will most likely have to focus on non-referring object, or natural kind, concepts. Therefore, it makes sense to adopt the hypothesis that these concepts will turn out to be similar to their referring counterparts. The second advantage of choosing prototype structure as a working hypothesis is that there are well-established empirical methods for determining whether something has a prototype structure, and whether that structure plays a role in processing. My own

empirical work reviewed in the next chapter will more or less follow one of those methods.

First Refinement: Similarity-based Theories

The first try at the hypothesis was too strong. One reason is that the inclusion of the term "prototype" seems to imply that I have some particular structure in mind as the "correct" view of conceptual structure based on similarity. But in fact, at this point it is not necessary to buy into any one particular similarity-based view of conceptual structure. The basic idea behind prototype theory is that a concept is a structure that encodes information that is typical of, rather than necessary or sufficient for, membership in the corresponding category. Categorization proceeds by comparing this category prototype with a mental representation of an instance or exemplar (in the latter case, the representation would likely be a prototype as well, since exemplars are themselves concepts at a more subordinate level). If the instance or exemplar is sufficiently similar to the concept prototype, then it satisfies the concept, and can be categorized using the corresponding concept label. A number of proposals have been made regarding the exact structure of prototypes, most of which are easily categorized either as property list accounts or conceptual space accounts.

According to property list accounts of prototype structure (Hampton, 1993, 1995; see also Sperber & Wilson, 1995), a prototype is a list of weighted properties, with the weights corresponding, for example, to the frequency of occurrence of that property among the category exemplars. Typically, similarity is computed by comparing the properties associated with the incoming representation with the list of prototype properties, summing the weights for the properties that match. If the goal is to categorize,

then the next step is to compare that result to a threshold value. If the sum of property weights exceeds the threshold, the incoming representation is categorized using the concept label of that prototype. This account is directly related to the family resemblance idea suggested by Rosch and Mervis (1975) – the instances or exemplars with high family resemblance are likely to be highly similar to the prototype. According to conceptual space, or distance-based, accounts of prototype structure (Gärdenfors, 2000; Osherson & Smith, 1981), prototypes are points or regions in a multi-dimensional property space, instances and exemplars are points or regions in the same space, and similarity is computed by measuring the distance (usually the Euclidean distance) between them. If the goal is to categorize, there is a distance threshold within which the incoming representation is categorized using the concept label of the prototype.

So there are a number of different ways in which a prototype theory can be realized. But prototype theories do not exhaust all of the similarity-based structural options either. Some researchers have claimed that rather than corresponding to a single prototype, a single concept in fact corresponds to multiple prototypes or exemplars. This type of proposal is usually called an "exemplar-based" rather than "prototype" theory. Like prototype theories, exemplar-based accounts come in both property list varieties (Smith & Medin, 1981) and conceptual space varieties (Nosofsky, 1988, 1992), but tend to be more popular among researchers interested in artificial category learning. In any case, for now the point is just that in order to choose between the five possibilities outlined above, I need not frame the working hypothesis in terms of prototypes per se. What really distinguishes the third possibility is not that concepts have any particular

kind of prototype structure, but that they have some kind of similarity-based structure. So here's the second formulation of the working hypothesis:

> Second try: *Non-referring object concepts have a similarity-based structure, just like referring object concepts*.

Evidence favoring a hypothesis like this would still be good evidence against the four theories that compete with similarity-based theories such as prototype theory, but the new formulation is a little more cautious about what its confirmation would constitute evidence *for*.

Second Refinement: Hybrid Theories

Unfortunately, the second formulation of the working hypothesis is still too strong. Simply demonstrating the presence of both typicality effects and a link between typicality and family resemblance may count as evidence in favor of a similarity-based structure, but it is not necessarily evidence in favor of an exclusively similarity-based structure. The possibility remains open that the similarity-based structure is but one component in a hybrid structure. A number of researchers from all areas of cognitive science have recently endorsed hybrid models of conceptual structure, in which similarity plays a role, but not an exclusive one (Anderson & Betz, 2001; Michalski, 1993; Murphy, 2002; Smith, Patalano, & Jonides, 1998; Smith & Sloman, 1994). I shall argue in some detail in Chapter 5, based on my own work and on the work of others, that categorization behavior is often affected by knowledge that is not directly encoded into a similarity-based structure. This claim will turn out to be important in explaining my results for the processing of non-referring concepts. I don't necessarily endorse the position that this non-similarity-based knowledge is constitutive of the concept, but in principle, it could be. So the third, and final, formulation of the first hypothesis is as follows:

> Hypothesis I (final version): *Non-referring object concepts include similarity-based structures, just like referring object concepts*.

The above formulation is more appropriate, but at a price. It is a little weaker with respect to its ability to discriminate between the five possibilities outlined earlier. Evidence for this hypothesis would still be evidence in favor of prototypes or some other similarity-based structure, and would still be evidence against ad hoc structure, as well as against any of the other three options when they are interpreted as exhaustive descriptions of conceptual structure. However, showing that similarity-based structures are *included* in the structure of non-referring concepts would not completely rule out definitions, essences, or ideals as also being *included* in that structure. On the other hand, evidence for this formulation of the hypothesis would still be evidence in favor of Psychological Meinongianism and against at least the structural version, though perhaps not Taylor's processing version, of the pseudo-Russellian special case solutions to the problems of non-referring terms outlined in Chapter 2.

Note: Hybrid Theories are not Dual Theories

Before moving on to introduce a second hypothesis, I pause to note that "hybrid theories" of conceptual structure, as I am defining the term, are not the same as the popular "dual theory" accounts of conceptual structure. I endorse the possible correctness of a hybrid account, but I am much more skeptical about dual theory accounts. The basic idea of a dual theory is that the representations we use to identify instances for categorization are separate from the representations we use to perform more general tasks. This split was first proposed in semantics by Miller and Johnson-Laird (1976) and was first pursued in the categorization literature by Smith and Medin (1981), to be endorsed later in a number of other places (Armstrong et al., 1983; Medin, Wattenmaker, & Hampson, 1987; Medin

& Smith, 1984; Osherson & Smith, 1981) as a plausible way to explain typicality effects while leaving room for classical accounts alongside similarity-based accounts. The difference is that dual theory proposals would push similarity-based structures out of the core of the concept into a peripheral role only, whereas hybrid proposals would leave them in the core of the concept alongside other sorts of information. So in reality, dual theories were always intended to show that similarity *does not* play a core role in conceptual structure.

There is actually little evidence for the purported split between core and identification procedures. Participants will rate instances and exemplars for typicality whether the stimuli are pictures, in which they would probably make use of only the identification procedure (Rosch & Mervis, 1975), or words, in which they would most plausibly access the core (Rosch, 1973b). A study that looked specifically at the difference between pictures and words found that participants' typicality ratings are identical no matter how the stimuli are presented (Smith, Balzano, & Walker, 1978). Hampton (1981) also showed that some abstract concepts, in which the notion of an "identification procedure" makes less sense, show significant prototype effects as well. Furthermore, recall that Rosch (1973b) found reaction time differences that correlated with typicality judgments in categorization tasks involving sentences – a task which probably forces participants to access the cores of their concepts. Finally, Rosch (1975) also found explicit evidence that typicality effects in processing were a core rather than identification effect. In a task where participants were primed with a category name, and then asked to judge whether pairs of pictures were the same or different, she found that reaction time varied with typicality only if participants were instructed to interpret

"same" to mean that the pictures were members of the same category. The effect did not show up when participants were instructed to interpret "same" to mean the pictures were physically identical.

Dual theory's partial return to classical theory also brings back many of the problems of the original version – for instance, our inability to determine definitions is as much an issue for dual theories as it is for the full classical theory. It is also an open question whether positing definitional cores really solves the main problem that its proponents wanted it to solve, namely the compositionality problem (Fodor, 1981, 1998; Fodor & Lepore, 1996; Lawrence & Margolis, 1999; Osherson & Smith, 1981; Rey, 1983; Smith & Medin, 1981). According to proponents of the arguments from compositionality, prototype theory predicts that a concept like PET FISH ought to have a prototype something like a goldfish, whereas the PET prototype is probably something like a dog and the FISH prototype might be a trout. So how do you compose dogs and trout to come up with goldfish? The problem is even worse for exemplar accounts where you have to somehow compose two whole sets of exemplars to come up with a new set for the conjunction. This dissertation is not the place to review the entire debate over compositionality, though it is not totally clear that there isn't an empirical solution to this problem, at least within the framework of a prototype theory (see Hampton, 2000, or Prinz, 2002, for reviews of the evidence). But granting for the sake of argument that compositionality is a problem for similarity-based accounts, it is not at all clear how dual theories solve it. The trouble is that if compositional concepts show prototype effects, as the PET FISH argument requires, then these concepts must have similarity-based structures used for identification. And we would still need an account of how *these* structures

compose – it's just that they would be composing for identification rather than for core semantic processing.

Hypothesis II

The basic intuition behind the first working hypothesis is that non-referring object concepts are very similar in structure to their referring counterparts. This hunch about the psychology of concepts is also compatible with the Psychological Meinongian position outlined in Chapter 2. If, as is predicted, the structure of non-referring and referring concepts is quite similar, then a processing prediction follows, namely that there will be no measurable differences in how these two types of concepts are processed. Recall Question 2 from Figure 1-2: Are there any other (i.e., non-structural) significant psychological differences between referring and non-referring concepts? The Psychological Meinongian position leads naturally to a second hypothesis in answer to this second question:

> Hypothesis II: *The processing of non-referring object concepts is no simpler or more complicated than the processing of referring object concepts*.

Notice that this hypothesis makes a prediction that contradicts that of Taylor's special case solution reviewed in Chapter 2. Taylor proposed that non-referring concepts have a structure identical to their referring counterparts, but that they go through an additional step in pragmatic processing in which a proposition is formed using descriptive information about the concept drawn from long-term memory. This is a process that will take time, and this time should be measurable in a reaction-time task.

Predictions

This chapter reviewed the main arguments and evidence for five different accounts of conceptual structure: definitions, essences, prototypes, ideals, and ad hoc structure. Two working hypotheses were adopted:

> Hypothesis I: *Non-referring object concepts include similarity-based structures, just like referring object concepts do*.
>
> Hypothesis II: *The processing of non-referring object concepts is no simpler or more complicated than the processing of referring object concepts*.

The first hypothesis can be verified by performing some of Rosch's experiments on typicality effects and family resemblance using non-referring concepts. If verified, I will have an argument in favor of the proposition that non-referring object concepts have a structure similar to their referring counterparts, and against the claim that non-referring concepts are ad hoc or structured exclusively as definitions, essences, or ideals. I will also have an argument in favor of Psychological Meinongianism and against special case solutions to the problems of non-referring terms. The second hypothesis can be verified by comparing reaction times in a task that requires participants to verify propositions involving non-referring concepts to the reaction times for the same task performed with referring concepts. If verified (or more precisely, since it has the character of a null hypothesis, if it cannot be rejected on the basis of experimental data) the second hypothesis will provide a further argument that non-referring object concepts have a structure similar to their referring counterparts, in favor of Psychological Meinongianism, and against the claim made by Taylor that interpreting propositions containing non-referring concepts requires an extra processing step.

Chapter 4: Non-Referring Concepts

> Marge: Homer, there's someone here who thinks he can help you.
>
> Homer: Batman?!
>
> Marge: No, it's a scientist.
>
> Homer: Batman's a scientist
>
> Marge: It's not Batman.
>
> – *The Simpsons* ("Marge vs. the Monorail")

The Plan

In the last two chapters, I reviewed the philosophical, linguistic, and psychological literature looking for insights regarding non-referring concepts that might help formulate and empirically test some hypotheses relevant to the main questions posed in this dissertation. That discussion culminated in two hypotheses about non-referring concepts – namely, that they include similarity-based structures just like referring concepts (Hypothesis I), and that they are no simpler or more complex to process than referring concepts (Hypothesis II). Chapter 3 also pointed the way towards an experimental program that could put these two hypotheses to the test. This chapter describes the methodology, results, and implications of each of four main experiments as well as two follow-up experiments (Experiments 3a and 4a). Experiments 1 to 3 replicate Rosch and Mervis' (1975) classic work on typicality and family resemblance using non-referring concepts, and the results together provide strong evidence in favor of Hypothesis I. Experiment 3a addresses a concern about the methodology of those first three studies. Experiment 4 is a speeded categorization study, the results of which provide further

evidence confirming Hypothesis I, but, surprisingly, seem to disconfirm Hypothesis II. Finally, Experiment 4a is an attempt to rule out a possible alternative explanation for that latter result.

Experiment 1: Associative Frequency

Experiment 1 was a preliminary study to collect usable stimuli for Experiments 2 to 4. There were two sub-goals to the experiment. The first was to determine some usable non-referring category names (those that contain a large number of psychologically salient, non-referring exemplars), and the second was to determine which lexical items participants would list as members of those categories. (Given that there is no wholly participant-independent fact of the matter concerning what might or might not be a member of a non-referring category, exemplars could not be determined in any way other than by asking the participants themselves.) These two goals were achieved by way of an associative frequency study in which participants listed exemplars for various referring and non-referring categories.

Though there have been at least six major associative frequency studies in the past (Battig & Montague, 1969; Hampton, 1983; Hunt & Hodge, 1971; Loess, Brown, & Campbell, 1969; McEvoy & Nelson, 1982; Shapiro & Palermo, 1970), [39] only 3 of the 388 categories examined in these studies were non-referring, and participants almost never produced non-referring terms in otherwise referring categories. The three non-referring category names, were “mythological characters”, used by Hunt and Hodge (1971), and “mythical being” and “extinct animal”, used by McEvoy and Nelson (1982).[40] These three categories are not appropriate for further study in the current experiments since they have composite (multiple-word) category names. Sticking to

single-word non-referring category names is important for now, both because concepts with composite labels are often not structured by similarity, and because the nature of what is going on in conceptual combination is quite controversial (see the related discussions in Chapter 3). What was needed for this study was a set of associative frequencies for non-referring concepts with single-word English names.

Method

Participants

The participants were 51 undergraduate student volunteers, 16 male and 35 female, with a mean age of 23.0 years. One (female) participant was discarded for performing the task incorrectly, so the number of usable responses was 50. Participants were recruited from first- and second-year undergraduate classes and tested in small groups either immediately after class or during the class break. All participants either listed English as their mother tongue or claimed fluency in English, and most participants (44 of 50) were either born in Canada or had lived in Canada for more than 10 years prior to the study.

Stimuli

The stimuli were 10 category names. The target categories were non-referring categories with single-word English names. “Aliens”, “gods”, “monsters”, and “superheroes” were the only four English words identified during brainstorming that seemed to denote potentially large non-referring categories (i.e., there seemed to be a large number of lexical items that could potentially be listed as names of exemplars).[41] The category name “dinosaurs” was also included, since although it could have referred at one time, it now fails to do so. The category name “imaginary or mythical animals” was also included to

try to capture some of the more common non-referring terms in the philosophical and linguistic literature (i.e., “unicorn”, “centaur”, etc.). This category was not intended to be used on its own in future studies, since it does not have a single-word name, but some of the exemplars generated could be mixed in with real animals to form a hybrid set of stimuli under the category name “animal”.

In addition to these six non-referring category names, four referring category names were included, both to space out the non-referring categories and to provide data for some referring categories for future studies. These category names were “animal”, “fish”, “pet” and “pet fish”. Note, however, that there are plenty of existing associative frequency lists for referring categories, so the data collected under these category names was not absolutely essential to the current work.

Procedure

The procedure closely paralleled the classic study by Battig and Montague (1969). Participants were given a booklet with the name of one of the categories at the top of each page. The pages were ordered randomly, but each participant received them in the same order.[42] For each category, participants were given 45 s to write down as many exemplars or instances as they could think of. The 45 s time limit kept the study short enough that it would not seem too unattractive to potential volunteers.[43] The instructions, printed on the front of the booklet and read aloud from the front of the room, were as follows:

> The purpose of this experiment is to find out what items or objects people commonly give as belonging to various categories or classes. The procedure will be as follows: First you will read the name or description of the category printed at the top of each page of this booklet. Then you will be given 45 seconds to write down on that page as many items included in that category as you can, in whatever order they happen to occur to you. For example, if you were given the

> category “furniture”, you might respond with such items as chair, table, desk, bed, sofa and so on. For “fictional characters”, you might respond with Sherlock Holmes, Harry Potter, and so on. The words are to be written in this booklet, using a different page for each category. When you hear the word “stop”, you are to stop writing and turn over immediately to the next page of the booklet, where the next category name will be written. Again you are to write the names of as many members of that category as you can think of. This procedure will be continued through a total of 10 categories, and you are to use a different page of the booklet for each category. Now please wait for the experimenter’s signal to turn the page and start with the first category.

Results

Associative Frequency

The associative frequency score for an instance or exemplar in a given category is equal to the number of participants who listed that item in that category. A full list of all items produced by participants in this experiment, sorted by associative frequency, can be found in Appendix A. Words and phrases were coded as written, even if they seemed to be anomalous or incorrect. Variations in spelling were lumped together under the “correct” dictionary spelling, but only if it was very clear that the items were indeed spelling variations and not distinct items. As expected, no participant produced a non-referring term in an otherwise referring category, and referring terms in non-referring categories were exceedingly rare and could usually be attributed to a metaphorical reading of the category name.

Comparison to Past Studies

Most of the 10 category names in this study had not been tested for associative frequency before, making comparison difficult. However, the “fish” category has been used a number of times in the past. As a historical curiosity, Table 4-1 compares the top 10 fish listed by participants in this study as well as three other studies from about 30 years ago

and collected in various different regions of North America. Note that there is remarkably little variation in the 10 most salient fish across all these studies. ("Sunfish" was the only item generated in the top 10 of the current study that did not make the top 10 of at least one of the previous studies.)

Table 4-1
Top 10 Fish in Four Associative Frequency Studies

Ohio, 1969[a]	Maryland & Illinois, 1969[b]	Georgia, 1971[c]	Canada, 2002[d]
trout	trout	bass	goldfish
bass	bass	trout	salmon
shark	shark	catfish	shark
perch	herring	shark	trout
goldfish	catfish	perch	tuna
salmon	perch	salmon	bass
tuna	salmon	flounder	catfish
catfish	tuna	goldfish	pike
swordfish	goldfish	bream	swordfish
pike	swordfish	brim	sunfish

[a] (Loess et al., 1969). 470 participants wrote four exemplars per category.
[b] (Battig & Montague, 1969). 442 participants had 30 s per category.
[c] (Hunt & Hodge, 1971). 400 participants wrote four examplars per category.
[d] (Scott, current study). 50 participants had 45 s per category.

Category Selection

The first goal of this experiment was to descriptively analyze the results in order to choose promising non-referring categories for further investigation. The first part of this analysis involved determining the sizes of the categories (i.e., the total number of items listed for each category). Figures for category size are shown in Table 4-2 below. On the basis of relative category size alone, DINOSAUR was dropped immediately as a potential category. Of those remaining, the exemplars listed under "gods" were judged likely to be unfamiliar to most participants (or at least known by name only), and the list of actual category members (as opposed to properties and free associations) produced under

"aliens" was judged far too small for that category to be usable. So the categories GOD and ALIEN were also excluded. On the other hand, the participants produced a large list of seemingly salient and familiar items in response to the category names "monsters" and "superheroes", so the categories MONSTER and SUPERHERO were chosen to go forward into the future studies.

Table 4-2
Sizes of the 10 Categories

Category	Size
animals	121
fish	101
gods	85
monsters	76
imaginary or mythical animals	75
superheroes	65
aliens	65
pets	61
dinosaurs	40

For the referring categories, ANIMAL and FISH were chosen to match MONSTER and SUPERHERO, respectively. Intuitively, MONSTER seemed to be a more superordinate category, like ANIMAL, because participants tended to name kinds of monsters (e.g., "werewolf") rather than individual monsters (e.g., "Frankenstein"). On the other hand, SUPERHERO seemed to be a more basic level category because participants almost exclusively listed names of individuals, which are subordinate by definition. There was a further advantage to using ANIMAL in that some of the exemplars from the IMAGINARY OR MYTHICAL ANIMALS category could also be included to make the category an interesting hybrid of referring and non-referring exemplars.

The second goal of the experiment was to identify lexical items in each of the categories to be used in Experiment 2. For the chosen categories, the participants in this

study generated between 65 and 121 possible lexical items. The process used to select a subset of these items for further study is described in the "Basic Stimuli" section of the discussion of Experiment 2 below.

Experiment 2: Typicality

The second experiment had three main goals. The first was to collect typicality ratings for at least 20 exemplars and instances (hereafter referred to as "items") from each of the four categories identified in Experiment 1. I expected to meet this goal with no trouble – there is no example in the literature of participants failing to provide consistent typicality ratings for any kind of category. The second goal was to double-check category membership for the items in the non-referring categories. Since there is no wholly participant-independent way to determine whether something counts as a monster or a superhero, it was important to make sure that the items generated for those categories by the participants in Experiment 1 would indeed be accepted as category members by the independent group of participants in Experiment 2. Finally, the third goal was to use the typicality and categorization data to serve as an early check on the degree of similarity in the participants' behavior towards referring versus non-referring categories.

No previous study to date has collected typicality data for naturally occurring, non-referring categories, although such data has been collected for a number of artificial categories and for artificial exemplars and instances of natural referring categories.[44] However, the relevance of these studies to the current work varies. In most artificial category studies, the instances consist of random dot patterns, or combinations of form and color that don't appear to represent anything other than themselves, so it is not clear whether the concepts learned by participants in these experiments should properly be

considered referring or non-referring. But in some cases, the stimuli are pictures or descriptions that could be construed by the participants as representing nonexistent objects. For example, the cartoon-like animal drawings used by Medin, Wattenmaker, and Hampson (1987) or the stick figures used by Rosch, Simpson, and Miller (1976) could be seen to represent nonexistent individuals, although those particular studies didn't include any methods for influencing or determining whether the participants were rating the pictures themselves (that is, the forms on paper) or the nonexistent items they appeared to represent. There have also been a few artificial category studies in which nonexistent category exemplars and instances were introduced via short stories or property lists, and participants were specifically instructed to imagine the described item when providing typicality ratings (e.g., Ahn, Kim, Lassaline, & Dennis, 2000; Malt & Smith, 1984). But it is an open question whether artificial category studies, using stimuli introduced recently under experimental conditions, are causing participants to construct or access the same types of concepts as natural category studies, in which participants are assumed to be accessing concepts that were formed over a relatively long period of time and persist after the study has concluded.

In any case, participants in artificial category studies have never had trouble generating typicality ratings for exemplars and instances of artificial categories, presumably with the same level of agreement between participants as for ordinary (referring) natural categories. So if the artificial category studies are measuring anything at all similar to the current study on naturally occurring non-referring categories, then their results make it seem even less likely that the participants in the current study would have any trouble generating typicality ratings.

Method

Participants

The participants were 49 undergraduate students, 24 male and 25 female, with a mean age of 21.9 years. They were tested in small groups, and most performed the task for extra academic credit. All participants either listed English as their mother tongue or claimed fluency in English, and most participants (45 of 49) were either born in Canada or had lived in Canada for more than 10 years prior to the study.

Basic Stimuli

The four categories, selected based on the results of Experiment 1, were ANIMAL, MONSTER, FISH, and SUPERHERO, labeled on the experimental materials as "an animal", "a monster", "a fish" and "a superhero". The indefinite article was included in the category names because of the ambiguity of the word "fish". It was hoped that "a fish" would tend to focus the participants on fish as discrete objects rather than fish as a kind of meat. Twenty stimuli were selected from the word lists generated by participants in Experiment 1 using the following procedure:

1. For the categories ANIMAL and FISH, an initial list was generated, consisting of random selections from the lists generated by the participants in Experiment 1. One exemplar was chosen from each level of associative frequency (i.e., one exemplar with an associative frequency of 50, if there was one; one with an associative frequency of 49, if there was one; and so on – see Appendix A). For the smaller categories of MONSTER and SUPERHERO, the initial lists included all the words and phrases generated by the participants in Experiment 1.

2. For all categories, a number of items were dropped from the initial lists. Obvious non-members were excluded (e.g., “Joker” names a villain rather than a superhero, “superwoman” has no obvious narrative source, and “Hitler” is a monster only on a metaphorical reading of the term).[45] Pairs of overlapping items were identified, and the less specific item from each pair was excluded (e.g., “bird” names category that is superordinate to the category named “canary”, and the category named “ghoul” probably includes the category named “ghost”). Obscure items or items with unfamiliar names were excluded (e.g., “Mr. Freeze” names a little-known superhero, and “Sully” names a movie monster from *Monsters, Inc.* who many participants described, but did not know the name for). Ambiguous items were excluded (e.g., “bass” names both a fish and a musical instrument, and “Rage” names both a superhero and an emotion). Finally, phrasal descriptions were also excluded (e.g., “the monster under the bed”).
3. The ANIMAL category, which had a very large initial list, was also pared down by two further methods. Pairs of words for very similar animals, such as “mouse” and “rat”, were identified, and the one with the higher written word frequency according to Kučera and Francis (1967) was removed. Also, no more than two animals of the same type (rodent, bird, etc.) were allowed in the final list.
4. In the FISH category, the list was reduced by removing some of the fish that would probably be more familiar to the participants as a kind of meat than as a kind of living animal (e.g., “halibut”), and by removing any item that more than 50% of McCloskey and Glucksberg’s (1978) participants had rejected as a member of the category (e.g., “lobster”).

5. Finally, both the MONSTER and SUPERHERO categories contained a mix of terms picking out exemplars (e.g., "X-man", "ghost") and instances (e.g., "Superman", "Dracula"). An attempt was made to "clean up" these categories so that MONSTER would contain only exemplars and SUPERHERO would contain only instances. This was not totally successful for the MONSTER category – in order to get a list of twenty stimuli, some MONSTER instances (those with the highest associative frequency) were included along with the exemplars.

"Mixed" Category Stimuli

As mentioned in the discussion of Experiment 1, the ANIMAL category was turned into a "mixed" category by including five of the more frequently mentioned non-referring terms from the IMAGINARY OR MYTHICAL ANIMALS category of Experiment 1. To maintain the apparent symmetry between the ANIMAL and MONSTER category, five "monstrous" animals were selected from Experiment 1 to be added to the list of monsters. So in the final lists of stimuli, the more superordinate categories of ANIMAL and MONSTER contained 25 items each, while the more basic level categories of FISH and SUPERHERO contained only 20 items each. (See Appendix B for the full lists of stimuli.)

The point of the mixed categories was to see whether participants would accept the mixing of referring and non-referring terms in a single category, as previous work seemed to suggest that they might. For example, Malt and Smith (1984) collected typicality ratings for novel nonexistent items, introduced with property lists, in otherwise referring categories (i.e., novel descriptions of birds, fruit, furniture, etc.). The participants in that study found the task sensible enough to produce graded typicality data for these anomalous stimuli. Thus, it seems that participants, at least under experimental

conditions, will sometimes allow their existing natural categories to include both existent and nonexistent items. This result is mirrored for natural categories by the work of McCloskey and Glucksberg (1978), in which both categorization judgments and typicality ratings were collected for categories containing a small number of non-referring terms alongside a larger number of referring terms. For example, the term "unicorn" was included in the category ANIMAL. The participants in that study not only accepted "unicorn" as an animal (83% in the categorization task), but they also gave it a mean typicality of 6.14 out of 10 – higher than a number of referring animal terms, such as "lobster", "worm", and "sponge".

Procedure

The procedure was very similar to most previous typicality studies (e.g., Rosch, 1973b). Participants were given a booklet with the name of a category at the top of each page and the items listed down the left-hand side. To the right of each item was a 7-point scale for rating the typicality of the item in the given category (7 = very good example, 1 = very poor example) as well as an "×" symbol that participants could circle if they felt the item did not belong in the category and a question mark that they could circle if they did not know what the item was. The items in each category were presented to each participant in one of two different random orders, and each category was presented on a separate page. All 24 possible orderings of the four category pages were used at least once for each of the two random item orders. The instructions, printed on the front of the booklet and read aloud by the experimenter, were as follows:

> This study has to do with what we have in mind when we use words which refer to categories. Let's take the word *red* as an example. Imagine a true red. Now imagine an orangish red … imagine a purple red. Although you might still name the orange-red or the purple-red with the term *red*, they are not as good examples

> of red (not as clear cases of what *red* refers to) as the clear "true" red. In short, some reds are redder than others. The same is true for other kinds of categories. Think of birds. Everyone has some notion of what a "real bird", a "birdy bird" is. To me a sparrow or a robin is a very birdy bird while an ostrich is a less birdy bird. Notice that this kind of judgment has nothing to do with how well you like the thing; you can like a purple red better than a true red but still recognize that the color you like is not a true red. You might think that the ostrich is way cooler than other kinds of birds without thinking that it is the kind of bird that best represents what people mean by birdiness. This judgment also has nothing to do with how frequently you see or think about the thing; you could live on an ostrich farm and deal with ostriches every day, but still think that they are a pretty bad example of what people generally mean when they talk about birds.
>
> On this form you are asked to judge how good an example of a category various instances of that category are. At the top of the page is the name of a category. Under it are the names of some instances of the category. After each instance is a list of numbers from 1 to 7. You are to rate how good an example of the category each instance is by circling a number on this 7-point scale. A 7 means that you feel the item is a very good example of your idea of what the category is. A 1 means you feel the member fits very poorly with your idea or image of the category. A 4 means you feel the member fits moderately well. It is also possible that you may not know what the example is, in which case you should circle the question mark to the right of the scale. You also may feel that the instance is not actually a member of the category at all, in which case you would circle the ×. An example is shown below for the category "a bird".
>
> **Woody Woodpecker** 7 6 5 4 3 2 1 **?** **×**
>
> If *Woody Woodpecker* fits really well with your idea or image of *bird*, circle the 7; if he fits your idea really poorly, circle the 1, and so on. If you have never heard of him, you should circle the question mark. If you have heard of him, but you feel very strongly that he is not actually a bird at all, you should circle the ×.
>
> Don't worry about *why* you feel that something is or isn't a good example of the category (and don't worry about whether it's just you or people in general who feel that way) – just mark it the way you see it.

In addition to these written instructions, participants were informed of the names of the four categories in advance, warned that the task might seem odd (due to the nature of the non-referring terms), and urged to take the task seriously.

Results

The complete list of results, including mean typicality, numbers of "×" and question mark responses, and some means and percentages for all four categories are presented in Appendix B.

Categorization

One of the most immediately striking results is that participants rejected the mismatched items in the mixed categories far more often than they rejected other items. Rejection rates, as measured by the proportion of participants who circled the "×" option, were 40% for the non-referring terms in the "animal" category and 48% for the referring terms in the "monster" category (question mark responses were removed before calculating these percentages). In contrast, participants circled "×" in only 2% of the responses for the referring terms in the "animal" category and 7% of the responses for the non-referring terms in the "monster" category. This finding is in marked contrast with what was expected based on the behavior of participants in previous studies (e.g., the 83% of McCloskey and Glucksberg's participants who reported that a unicorn was an animal).

Even after excluding the mismatched items, participants rejected items slightly more frequently in the non-referring categories (6.0% vs. 2.8% "×" responses after the question mark responses and mismatched items were removed). Further discussion and interpretation of these and other results is left to the Discussion section below.

Typicality

In all four categories, participants produced graded mean typicality ratings, though they tended to restrict their responses to the upper range of the 7-point scale. In the referring categories, mean typicality with the mismatched items removed ranged from 2.55 to 6.78,

while in the non-referring categories it ranged from 3.34 to 6.78. The mean of the mean typicality ratings over the 20 items was 6.04 (*SD* = 0.74) for the animals, 4.96 (*SD* = 0.84) for the monsters, 5.35 (*SD* = 1.11) for the fish, and 5.30 (*SD* = 0.99) for the superheroes. The following procedure was followed to determine whether these means significantly differed from one another. The typicality ratings given to each item by each participant were analyzed within a three-factor ANOVA in which the category factor (animals, monsters, fish, and superheroes) was regarded as a fixed effect, and both item stimuli (i.e., 20 items nested within each of the four category factor levels) and participants were treated as random effect factors. The category effect was tested using the quasi-F ratio recommended by Clark (1973; see the discussion of this test later on with respect to the analysis of the Experiment 4 data). All cases in which no typicality rating was available from a participant for a certain item were replaced with the group mean for that item in order to run the analysis. In this analysis, the main effect of category was significant, $F'(3, 104) = 4.07$, $p < .01$. A set of post hoc tests were then carried out by deriving six separate quasi-F ratios for each of the different tests between all of the pair-wise category means and comparing each to a Scheffé criterion that constrained the family-wise error rate to be at most .05 (Hays, 1988). These tests indicated that the overall mean typicality rating for the animals was significantly different from that of the monsters, but no other pair of category means was significantly different.

Some previous studies have looked at correlations between typicality and both associative frequency and written word frequency. These correlations were computed to address the extent to which frequency of exposure might account for typicality ratings, and the extent to which associative frequency and typicality measure the same underlying

variable. In one study, Mervis, Caitlin, and Rosch (1976) took typicality data for eight categories (from Rosch, 1975) and computed Spearman rank-order correlations between mean typicality ratings and both associative frequency (from Battig & Montague, 1969) and written word frequency (from Kučera & Francis, 1967). They found clear evidence for a relationship between associative frequency and typicality (i.e., significant positive correlations: $.48 \leq r \leq .74$, $p < .05$ for all eight categories), but no evidence for a correlation between word frequency and typicality. They also noted a trend towards a higher correlation between word frequency and associative frequency than between word frequency and typicality. They concluded that typicality was not confounded with word frequency and that typicality "may be a more direct measure of the common factor underlying [typicality and associative frequency]" (Mervis et al., 1976, p. 284). Hampton and Gardiner (1983) also obtained significant correlations between typicality and associative frequency, using their own participants for both measures ($.57 \leq r \leq .90$, $p < .01$, for the 12 categories they examined).

Results for similar analyses using the data from Experiments 1 and 2 of the current study are presented in Table 4-3 below. Note that the pattern of results is slightly different from that reported by Mervis et al. (1976), both for the ANIMAL category, where word frequency was significantly correlated with both associative frequency and typicality, and for the MONSTER category, where no significant positive correlations were found[46] (note, however, that the number of items in each category in the Mervis et al. study ranged from 50 to 60).

Table 4-3:
Spearman Correlations: Mean Typicality, Associative Frequency, and Word Frequency

Category Label	AF[a] vs. Typ[b]	AF vs. WF[c]	WF vs. Typ
an animal	.89**	.72**	.69**
a monster	.16	-.17 (-.01)[d]	-.48 (-.45*)
a fish	.83**	.45 (.08)	.28 (.01)
a superhero	.79**	.00 (-.20)	.00 (-.40)

[a] Associative frequency.
[b] Mean typicality.
[c] Word frequency.
[d] Figures in parentheses are correlations using mean WF for multi-word terms.
* p < .05, ** p < .01.

Finally, a comparison was made between the participants' categorization behavior and typicality ratings. Rates of rejection (number of "×" responses) were found to be significantly correlated with the mean typicality rating in all four categories: $r = -.55$, $p < .02$, for the animals; $r = -.69$, $p < .01$, for the monsters; $r = -.75$, $p < .001$, for the fish; $r = -.63$, $p < .004$, for the superheroes. The rates of "don't know" responses did not correlate significantly with either rates of rejection or typicality.

Agreement Between Participants

A number of the original typicality studies looked at the agreement between participants with respect to their typicality ratings, usually reporting what appeared to be high levels of agreement. For example, Rosch (1973b) used split-half correlations to test between-participant agreement. For a split-half correlation, participants are divided into two groups, and correlations are computed between the mean typicality ratings for the two groups. In her studies (with 100 or more participants in each), Rosch always obtained high split-half correlations ($r > .90$ for all categories), and she presented this as evidence of high agreement. However, as noted in Chapter 3, Barsalou (1987) subsequently pointed out that split-half correlations measure the stability of the mean rather than the

agreement between participants, and that arbitrarily high correlations can be obtained by increasing the number of participants in a study. In the current study, with less than half the participants of Rosch's studies, split-half correlations were computed by separating participants two ways. One split was based on which of the two random stimulus orders the participants received, and the other split was random. The split-half correlations based on stimulus order were all above .80: $r = .94$, $p < .001$, for the animals; $r = .83$, $p < .001$, for the monsters; $r = .94$, $p < .001$, for the fish; $r = 0.86$, $p < .001$, for the superheroes. The random split-half correlations were all above .75: $r = .94$, $p < .001$, for the animals; $r = .77$, $p < .01$, for the monsters; $r = .92$, $p < .001$, for the fish; $r = 0.89$, $p < .001$, for the superheroes. Note that there appears to be a trend towards lower split-half correlations in the non-referring categories.

The other statistic used by Rosch (1973b) was a χ^2 test on the distribution of responses across the 7-point scale. The test was performed for each item, with the flat (or uniform) distribution as the null hypothesis, the reasoning being that if the typicality task made no sense to the participants, they would respond randomly, and the overall effect would be a flat distribution on all or most items. What Rosch found was that in all but a few cases, the null hypothesis of a flat distribution could be rejected. This same analysis was performed on the current data, with much the same result – the null hypothesis was rejected at the $p < .05$ level for 88% of the items in the study. The exceptions were "octopus" in the ANIMAL category; "ghost", "alien" and "witch" in the MONSTER category; "suckerfish" and "eel" in the FISH category; and "Tarzan", "Mighty Mouse", "Sailor Moon" and "Captain Kirk" in the SUPERHERO category. Note that the null

hypothesis of a flat distribution was retained for more items in the non-referring than in the referring categories.

Finally, following on his criticism of Rosch's use of split-half correlations, Barsalou (1987) suggested that a more appropriate statistic for measuring inter-participant agreement was the mean correlation between all pairs of participants. He reported that he had found mean correlations of between .30 and .70 on his own studies, using the averaging method of Guildford and Fruchter (1973). Positive mean pair-wise correlations (computed using the same method) were also computed for all four categories in the current study: r_{avg} = .49 for the animals; r_{avg} = .22 for the monsters; r_{avg} = .46 for the fish; r_{avg} = .37 for the superheroes.[47] Note the lower mean pair-wise correlations for the non-referring categories.

Discussion

The first two goals of the experiment were to collect graded typicality data for the four categories, and to check that participants would agree with the categorization judgments from Experiment 1. As expected, participants took to the task easily and generated typicality data for both referring and non-referring categories with levels of inter-participant agreement comparable to most previous studies. Furthermore, participants showed a strong tendency to accept the given items as members of the given categories, except in the case of the purposely mismatched items in the mixed categories of MONSTER and SUPERHERO. This finding was the one unexpected result in the experiment, and thus requires further discussion.

First of all, the rejection of the mismatched items does not necessarily represent any disagreement between the participants of Experiments 1 and 2. None of the

mismatched items were generated by Experiment 1 participants for the categories in which they appeared in Experiment 2, although the non-referring animal terms had been generated under a category name in which the head noun was "animal". But why *did* so many participants reject the mismatched items? One possibility, in principle, is that the participants were simply reluctant to allow a mixture of referring and non-referring terms in a single category. If this were the case, one would expect the same participants who rejected the non-referring animal terms to also reject the referring monster terms. But this was not the case.

Let's call those participants who failed to accept a single mismatched term in any given category "realists" for that category, as they didn't seem to want to mix real and imaginary things. And let's call the participants who failed to reject a single mismatched term in a given category "anti-realists" for that category, as they were quite happy to allow the real and the imaginary to coexist in the same category. An analysis of the ANIMAL category showed that 12 participants were animal realists, while 20 participants were animal anti-realists (with the remaining 17 participants falling in between, accepting some mismatched items while rejecting others). But only 3 of the 12 animal realists were also monster realists. In fact, two of them were monster anti-realists. And only 11 of the 20 animal anti-realists were also monster anti-realists, while 4 of them were monster realists. So the mixing of the real and the imaginary seems not to be the major issue for these participants. Whatever is behind the rejection of the mismatched terms, the non-referring animal terms are clearly being rejected for a different reason than the referring monster terms. This point is an issue that might be suitable for future investigation, but no more can be concluded from the current study.

The third goal of the study was to compare participants' performance between referring and non-referring categories. The most important finding here was that the ranges of mean typicality ratings over items were comparable between category types, indicating that participants were not treating non-referring categories radically differently from referring categories. But by all measures, there appeared to be a slightly lower level of agreement between participants for the non-referring categories. Participants were slightly more likely to reject items in the non-referring categories, reflecting a higher level of disagreement over the extension of the non-referring categories. They were also less harmonious in their rating behavior, as shown by the lower split-half and mean pair-wise correlations for the non-referring categories. Finally, they were slightly more likely to give the appearance of random behavior for some of the items, as measured by the χ^2 test on the distribution of responses. But this more inconsistent performance for non-referring categories is easy to account for because the terms in these categories are more open to interpretation. Everyone is free to have their own slightly different interpretation of what, say, a boogeyman might be like, but they are not so free to have their own interpretation of what a dog is like. In any case, despite the slightly more variable behavior of the participants for the non-referring categories, inter-participant agreement, by all measures, was within the normal boundaries for all four categories.

In addition to the findings related to the three main goals, there were a few other facts that stood out in the analyses. First of all, the MONSTER category was the only one for which there was no observed correlation between associative frequency and typicality. This result is unusual in the context of previous studies, but it may simply reflect the fact that participants had difficulty producing items in this category in the first

experiment. This difficulty may be indicative of the low saliency of the MONSTER category compared to the other three, but it gives no particular cause for alarm at this point. Secondly, the ANIMAL category differed from the other three categories in two respects: (a) the mean typicality across the items was significantly higher than at least one of the other categories; and (b) the animal typicality ratings correlated significantly with word frequency, whereas this was not the case for the other categories. But both of these effects are probably an artifact of the choice of stimuli for this particular category. When only 20 items are being rated, the results are more open to a stimulus selection bias than when there are 50 to 60 items involved, as in the classic studies by Rosch. In any case, the ANIMAL category is not a target category for this study. The purpose of including the two referring categories was to serve both as a control and as a check on the methodology in case the participants behaved in an anomalous way towards the non-referring categories. If they behave as expected, as they appear to have done, the results in the referring categories are of less concern.

Summary

Overall, the results of this experiment agree with the findings from previous typicality studies. Participants took to the task of providing typicality ratings without complaint and produced relatively consistent graded ratings across the items in each category. Furthermore, the mean typicality ratings appeared to be tracking categorization decisions somewhat, as shown by the significant negative correlations between the number of participants rejecting an item as a category member and the mean typicality rating assigned to that item by the remaining participants. The performance of the participants was comparable across both referring and non-referring categories, with the one minor

exception that both categorization and typicality rating behavior showed slightly more inter-participant variability in the non-referring categories. Thus this experiment, along with Experiment 1, succeeded in providing usable typicality data on non-referring categories which could be used to proceed with the family resemblance and speeded categorization analyses of Experiments 3 and 4.

Experiment 3: Family Resemblance

The goal of the third experiment was to look for evidence indicating that within category similarity is a causal determinant of the mean typicality ratings for the non-referring categories collected in Experiment 2. Given evidence about the causal determinants of typicality, I can draw tentative conclusions about the structure of the underlying non-referring concepts, and thus compare it with the structure of referring concepts. In particular, this experiment is an important test of Hypothesis I, the hypothesis that like referring concepts, non-referring concepts also include similarity structures.

A number of previous studies have found a significant correlation between the similarity of an exemplar to its category, measured by family resemblance or central tendency, and typicality ratings. Rosch and Mervis (1975) used their “family resemblance” measure to show that exemplars and instances that are better examples of a category tend to have more properties in common with other members of the category. This result was replicated and extended by Malt and Smith (1984) and also reproduced with some variations in methodology by Hampton (1979). Furthermore, Barsalou (1987) found that better examples tend to be judged by participants as more similar to other category members. These results have often been interpreted as supporting some kind of similarity-based view of the psychological structure of concepts (Goldstone, 1994;

Hampton, 2001; Rosch & Mervis, 1975). But to date, these results have not been extended to non-referring categories. The current study uses a methodology similar to both Rosch and Mervis (1975) and Malt and Smith (1984) to assess whether family resemblances, and hence similarities, play a role in the psychological structure of non-referring concepts as well. The experiment involves a property listing task followed by a family resemblance analysis on the results from that task.

Method

Participants

The participants were 45 undergraduate students, 22 male and 23 female, with a mean age of 21.8 years. They were tested in small groups, and most performed the task for extra academic credit. All of the participants in this study also participated in Experiment 2, which was conducted immediately after this experiment.[48] All participants either listed English as their mother tongue or claimed fluency in English, and most participants (42 of 45) were either born in Canada or had lived in Canada for more than 10 years prior to the study.

Stimuli

The stimuli used in this study were the same 90 stimuli used in Experiment 2, that is, the 25 animal, 25 monster, 20 fish, and 20 superhero terms listed in Appendix B. The "mixed" category stimuli, that is, the non-referring animal and referring monster terms, were included in the experimental task, but following the analysis of Experiment 2, which showed high rates of rejection of those items as category members, they were excluded from the family resemblance analysis.

Procedure

The procedure was very similar to previous family resemblance studies (Malt & Smith, 1984; Rosch & Mervis, 1975). Participants were given a booklet with the name of each instance or exemplar (hereafter referred to as "items") printed at the top of a separate page. Proper nouns were capitalized, and common nouns were printed with the indefinite article to focus participants on the object or count noun interpretation of the term (e.g., "a salmon" rather than "salmon"). Immediately underneath the name of the item was a box labeled "I've never heard of this", which participants were to check if they did not know what the item was. The rest of the page was blank, and participants were given 60 s per page to write down as many "attributes or characteristics" (that is, properties) of the item as they could think of. Each participant received 20 items, and each item was seen by 10 participants. No two participants saw the same 20 items. The instructions, printed on the front of the booklet and read aloud by the experimenter, were as follows:

> This is a very simple experiment to find out the characteristics and attributes that people feel are common to different kinds of things. The things you will be asked about will be either generic object types (like *bicycle*), individuals (like *Elvis Presley*), or particular objects (like *Eiffel Tower*). The things you will be asked about might be things that actually exist (like *Eiffel Tower*) or they might be things that actually do not exist (like *Bugs Bunny*).
>
> You will be asked to list all the characteristics and attributes you can think of for each thing. For example, for *bicycle* you might think of things bicycles have in common like two wheels, pedals, handlebars, you ride on them, they don't use fuel, etc. For *Bugs Bunny*, you might think of things that are characteristic or descriptive of him, like rabbit, rude, gray fur, talks, white belly, wears gloves, eats carrots, etc.
>
> There are 20 pages following this one. At the top of each page is the name of one thing. For each page, you'll have 60 seconds to write down all of the attributes or characteristics of that thing that you can think of. But try not to simply free associate – for example, if bicycles just happen to remind you of your father, *don't* write down *father*.
>
> It is possible that you may not have heard of one of the things you will be asked about. If this happens, please put an X in the box beside the words "I've never heard of this" at the top of the page, leave the rest of the page blank, and wait for

> the signal to turn to the next page. Only do this if you have *really* never heard of the item. If you have heard of it but you can't think of any properties, do not check the "I've never heard of this" box.
>
> Okay – you have 60 seconds for each page. You'll do 10 pages, then we'll take a short break, then you'll do another 10 pages. When I give the signal, turn to the first page, read the name of the first thing and write down the attributes and characteristics of that thing as fast as you can until you're told to turn the page again.

In addition to these written instructions, participants were warned that the task might seem odd or even silly, due to the character of the non-referring terms, but that they were urged to take the task seriously.

Raw Data

For this experiment, most of the interesting results come from the family resemblance analysis. However, there are a few things that can be noted about the raw data at this point. As shown in Table 4-4 below, participants generated between 2000 and 3000 properties in total for the 20 items in each of the four categories. As in Experiment 2, the participants' familiarity with the items seemed to be lower for the non-referring categories, but there was also evidence here for lower familiarity with the basic level categories of FISH and SUPERHERO. This difference can be discerned in Table 4-4 by looking at the mean number of participants out of 10 who recognized the items in each category (i.e., they did not check the "I've never heard of this" box). On the other hand, the mean number of properties listed per participant per item was more or less constant across the category types, though the figure was slightly higher for the ANIMAL category.

Table 4-4
Some Statistics from the Property Listing Task

Category	Properties[a]	Participants / Item[b]	Properties / Particiants / Item[c]
ANIMAL	2982	10.0	15.0
MONSTER	2405	9.3	13.0
FISH	2042	8.4	12.3
SUPERHERO	2229	8.3	13.4

Note. Mismatched items were discarded.
[a] Total number of properties generated by all participants over all items.
[b] Mean number of participants familiar with each item (out of 10).
[c] Mean number of properties generated per participant per item.

In a few cases, the items chosen for this study turned out to be unfamiliar to most participants. For example, 9 out of 10 participants either checked "I've never heard of this" for "neon tetra" or listed properties for a car rather than for a fish (they were presumably thinking of the Chrysler Neon). In the face of this, a criterion for the exclusion of items was developed. An item was excluded from the family resemblance analysis if half or more of the participants who received that item failed to produce usable properties. According to this criterion, four items (two fish and two superheroes) were excluded from the family resemblance analysis described below. The excluded items are shown in Table 4-5 below.

Table 4-5
Items Excluded from Family Resemblance Analysis

Category	Exp. 3[a]	Exp. 2[b]
FISH		
neon tetra	1	24
pickerel	5	40
SUPERHERO		
Green Lantern	3	31
Supergirl	4	44

[a] Number of participants (out of 10) who provided usable properties in Exp. 3.
[b] Number of participants (out of 49) who did not circle the question mark in Exp. 2.

Computing Family Resemblance

The goal of the family resemblance analysis is to assign to each item a numeric score that will be larger when the item has more properties in common with other members of the category, and larger still as the properties the item shares with other items are more common within the category. The computation of the family resemblance score has five steps: 1) list all the unique properties generated by the 10 participants for each item, counting the number of times each one is mentioned for that item, 2) assemble a "central list" of properties for each category, 3) assign a subset of the central properties to each of the category's items, 4) weight each of the central properties according to how common they are in the category, and 5) use the property lists from Step 3 and property weights from Step 4 to assign a family resemblance score to each item. The details of these steps were carried out as follows.

Step 1 – list the unique properties. This step was relatively straightforward, although many decisions had to be made about whether two responses were distinct or were expressing the same property. There seemed to be many cases of the same property being expressed in different ways by different participants, though in general, I tried to err on the side of caution. If I could think of a reasonable interpretation that made two responses distinct, I went with that interpretation. There were also a number of exclusions that had to be made. The majority of these exclusions were of entire property lists – cases where a participant had clearly misinterpreted the meaning of an item. For example, many participants interpreted "Green Lantern" as naming a type of bug light rather than a superhero, and a number of participants confused Godzilla with King Kong. In other cases, participants began to free associate or name subtypes of the item rather than listing

its properties. Again, I tried to err on the side of caution, only deleting a response if I could think of no way it might be used to name a property. I did not delete any properties just because I personally felt they were incorrect for that item. I only deleted responses that I felt did not actually name properties at all. In total, around two to three per cent of the approximately 10,000 properties were deleted in this way.

Step 2 – assemble a central list of properties for each category. The data in a task like this tends to be quite "noisy", so both of the studies on which this methodology is based included some method of cleaning it up. Rosch and Mervis (1975) went through the data from Step 1 and "corrected" it by hand, deleting properties they felt were wrong for a given item, and generalizing properties from one item to another. This process of correction seems like quite a dubious move, even for the referring categories Rosch and Mervis were investigating, but for non-referring categories, such a move seems even more dubious. As Experiment 2 showed, there is much less agreement between participants about non-referring categories, so intervening as an authority to "fix up" the data is out of the question. Malt and Smith (1984) used a slightly more principled methodology, assembling a "central list" of properties consisting of all properties mentioned by at least one third of their participants for a given item. I replicated this procedure, but using a lower cutoff value. A property made it into the central list as long as it was mentioned by at least two out of ten participants for a given item. For example, if two participants mentioned the property "furry" for "a bear", then "furry" was in the central list. But if only one participant mentioned the property "violent", then it would not be in the central list, even if one participant had also mentioned "violent" for another item. To make the central list, "violent" would need to be mentioned twice for a single

item. This cutoff seemed to screen out most of the more idiosyncratic or questionable responses, while still leaving most of the intuitively useful properties in the list. It was after this step that the excluded items from Table 4-5 were removed from further analysis.

Step 3 – assign subsets of the central properties to the items. Again following Malt and Smith (1984), once the central list was determined, each item was given a list of properties consisting of every property from the central list that was mentioned by at least one participant. So to continue the "bear" example from Step 2, if "violent" was mentioned twice for "a dog", then it would be in the central list. Then in the current step, both "a dog" and " a bear" would get the property "violent", even though it was only mentioned once for the latter item. However, if "violent" had only been mentioned once for each of "a dog" and "a bear", it would not have made the central list in Step 2, and thus neither item would get that property in this step. Step 2 can be thought of as increasing the precision or accuracy of the central property list, weeding out many of the properties that most participants would not agree with or think of, whereas Step 3 can be thought of as increasing recall or coverage of the individual items' property lists. Some examples of the resulting property lists can be seen in Table 4-6. In most cases, the lists seemed reasonable and accurate.

Step 4 – weight the properties. Each property in each category was then assigned an integer weight equal to the number of items in the category that had that property in their lists (maximum = 20). Thus the properties that were more common to items in the category received higher property weights. Table 4-7 shows the top 10 properties for each category.

Table 4-6
Example Property Lists

Item	Properties
a moose	animal, antlers, big, big nose, brown, Canadian symbol, four legs, fur, heavy, in Canada, in the forest, in water, long neck, smelly, strong, swims, tail
a troll	big eyes, big head, colorful hair, deformed, eats people, in a cave, in the forest, mean, scary, short, smelly, ugly, under a bridge, warts
a shrimp	edible, good with sauce, gray, in salt water, in schools, in the ocean, in water, pink, scales, small, smelly, swims, tail, tastes good, white
Spiderman	a man, blue, brave, climbs, fast, fictional, fights crime, in cartoons, in comic books, movie, on T.V., red, saves people, secret identity, shoots webs, strong, super powers, superhero

Table 4-7
Top 10 Properties by Weight for Each Category

ANIMAL		MONSTER		FISH		SUPERHERO	
property	weight	property	weight	property	weight	property	weight
fur	13	scary	18	in water	18	strong	14
big	12	fictional	14	swims	15	superhero	12
brown	12	myth	12	fish	14	fictional	11
four legs	11	big	11	ocean	13	hero	11
animal	9	evil	9	scales	13	powerful	11
white	9	movies	8	fins	12	TV	11
tail	8	green	7	gills	12	cartoon	10
fast	7	mean	7	small	12	super powers	9
gray	7	ugly	7	tail	11	attractive	8
mammal	7	monster	6	edible	9	muscular	8

Step 5 – compute family resemblance. Each item was given an integer family resemblance score equal to the sum of the weights of its properties.

In addition to the five steps outlined above, there was one piece of fine-tuning that needed to be performed. In every category, the name of the category itself appeared as one of the top 10 “properties” (see Table 4-7 above). This type of response is not logically inconsistent (“is an animal” is a perfectly good characteristic or attribute for “a dog”), but in the context of the current study, the presence of these properties with high

weights could have skewed the data towards a false positive result. For example, if the category name is more likely to be produced as a property for those items which are good examples, then including the category names could have the effect of inappropriately increasing the correlation between typicality and family resemblance. For this reason, the category name properties were removed before Step 5 above. The final family resemblance scores are shown in Appendix C, where they can be compared to the mean typicality ratings from Experiment 2. Appendix C also contains scatter plots of mean typicality against family resemblance for each of the categories separately.

Results

The main point of this experiment was to determine the relationship between typicality (from Experiment 2) and family resemblance (from the above analysis) for each category. Past studies have used both Spearman rank-order correlations (Rosch & Mervis, 1975) and Pearson correlations (Malt & Smith, 1984) between family resemblance and mean typicality. The argument for using a rank-order correlation is presumably that the typicality scale appears to be ordinal, which would make a rank-order correlation more appropriate. However, it is not clear that typicality really should be interpreted as ordinal, especially once the means for each item have been computed, and it is even less clear that the family resemblance score should be interpreted as ordinal. Thus, the Pearson correlation coefficient is probably the more appropriate measure, but for the sake of completeness and comparison, both correlations are shown in Table 4-8 below. Pearson correlations were positive and significant for three of the four categories: for ANIMAL ($r = .58$, $p < .01$), MONSTER ($r = .48$, $p < .04$), and SUPERHERO ($r = .67$, $p < .005$), but not for FISH ($r = .29$, $p > .25$). Spearman rank-order correlations were positive and significant

for MONSTER (r_s = .47, p < .04), and SUPERHERO (r_s = .64, p < .005), and were very close to significance for ANIMAL (r_s = .44, p = 0.053) and FISH (r_s = .47, p = 0.052).

Table 4-8
Correlations Between Mean Typicality and Family Resemblance

Category	N	Pearson	Spearman
ANIMAL	20	.58*	.44^{+}
MONSTER	20	.48*	.47*
FISH	18	.29	.47^{+}
SUPERHERO	18	.67*	.64*

$^{+}$ p < .055

The above analysis may be useful for comparison to past studies, but Lorch and Meyers (1990) have recently argued that this particular method of running a regression on repeated measures data tends to inflate the estimate of the percentage of variance accounted for by the predictor variables, and also to inflate the probability of a Type I error (in which the null hypothesis is rejected when it should not be). Therefore, a regression analysis was run using the method recommended by Lorch and Meyers (1990), which controls for excessive Type I errors by removing variance due to participants. In this analysis, the raw typicality ratings for each Experiment 2 participant were regressed against family resemblance as the predictor variable. For every category, partial correlations (controlling for participant variance) between raw typicality and family resemblance were lower than the previously computed correlations with mean typicality, though it should be noted that the total *r* (with all variables included in the model) tended to be a little higher (see Table 4-9). In the current statistical analysis (which also capitalizes on the increased number of observations used in the analysis to increase power when the null hypothesis is false), there was a highly significant main effect of family resemblance in every case: $F(1,48) = 50.69$, $p < .0001$, $b = .021$, for

ANIMAL; $F(1, 48) = 54.35$, $p < .0001$, $b = .021$, for MONSTER; $F(1, 48) = 55.38$, $p < .0001$, $b = .016$, for FISH; and $F(1, 48) = 86.41$, $p < .0001$, $b = .022$, for SUPERHERO. Note that the partial correlations are highest for SUPERHERO and lowest for FISH, reflecting the trends in the previous analysis. Table 4-9 summarizes the results of this and the previous analysis.

Table 4-9
Correlations Between (Mean and Raw) Typicality and Family Resemblance

	Analysis 1	*Analysis 2*		
Category	r_{MTFR} [a]	$r_{RTFR.PS}$ [b]	r_{ALL} [c]	*b*
ANIMAL	.58	.34	.71	.021
MONSTER	.48	.25	.54	.021
FISH	.29	.18	.54	.016
SUPERHERO	.67	.36	.60	.022

Note. The regression coefficient *b* is identical for both analyses.
[a] MT = Mean Typicality, FR = Family Resemblance.
[b] RT = Raw Typicality, PS = Participants.
[c] ALL = All variables entered.

Discussion

The high levels of significance obtained for every category in the second regression analysis, in which raw typicality was regressed against family resemblance, shows that family resemblance is a statistically significant predictor of participants' typicality ratings. The values of the partial correlations between family resemblance and typicality were comparable across referring and non-referring categories, and the size of the regression coefficient was approximately the same for three of the four categories, the exception being FISH, a result that will be discussed in a moment. The results for the first regression analysis of mean typicality against family resemblance, though it must be cautioned that they were the product of a much less powerful and less valid method, are nevertheless useful in enabling comparisons to previous studies. The sizes of the correlations for the non-referring concepts obtained from that analysis are well within the

range of those reported in the most comparable previous study on referring concepts, that of Malt and Smith (1984). Malt and Smith obtained significant Pearson correlations ranging from .42 to .69 for four of their six categories. Rosch and Mervis (1975) had previously obtained much higher Spearman correlations, ranging from .84 to .94, but, as mentioned above, their study involved a "fix-up" step in which the experimenters altered the data provided by the participants. So Rosch and Mervis' results may not be directly comparable either to the current study or to Malt and Smith's study. In any case, the pattern of results from both regression analyses is strong support for Hypothesis I, that similarity is a significant structural determinant for both the non-referring and referring lexical concepts in the study. This experiment has shown that one of the most robust results pertaining to referring lexical concepts can be replicated for non-referring lexical concepts as well. This evidence in favor of Hypothesis I significantly weakens any view that would postulate a special alternative semantic structure for non-referring concepts.

What remains to be discussed is the smaller effect found in the FISH category – the correlations were not significant in the first analysis, and significant but much smaller than the other three categories in the second analysis. Malt and Smith (1984) looked at six referring lexical categories and also found that for two of them, TREE and FLOWER, the Pearson correlations between mean typicality and family resemblance failed to reach significance, with $r = .23$ and $r = .33$, respectively. However, in a follow-up experiment, they used a property *rating* rather than a property listing task, recomputed the family resemblance scores, and did obtain significant Pearson correlations for the same two categories ($r = .56$ and $r = .57$, respectively), as well as higher correlations on two other categories. In a property rating task, participants are given a list of potential properties for

each item and asked whether that property is exhibited by that item. This task produces less noisy data because the participants are provided with the properties and do not have to retrieve them from memory themselves. There is no reason to believe that computing family resemblance using property ratings would not have the same effect here as it did for Malt and Smith, raising the correlation for FISH above significance in the first analysis, and possibly raising the value of the correlations and regression coefficients for all four categories in the second analysis as well.

But there is another reason why this correlation might have been particularly low for FISH, in particular. As the scatter plot for this category in Appendix C shows, there were three notable outlying points, corresponding to the items “shrimp”, “eel”, and “stingray”. These items received high family resemblance but low mean typicality ratings. Hampton (1979), using a different methodology, had previously found significant correlations for the FISH category. One of the differences in that study was that Hampton generated a central property list by interviewing participants about the category itself, rather than items within the category. He obtained a list of central FISH properties that closely resembled the list for this study, with the notable exception that his list contained shape properties (“streamlined”, and “oval, fish-like shape”) which, had the participants in the current study thought of them, could have differentiated the three outlying points from the other fish. In fact, when these two properties were added to the family resemblance analysis by hand (based on my own intuitions), significant Pearson and Spearman correlations between mean typicality ratings and family resemblance were obtained for the FISH category: $r = .52$, $p < .03$; $r_s = .61$, $p < .01$. In the regression using

raw typicality ratings, a partial correlation coefficient was obtained that was comparable to the other three categories in the same analysis, $r_{RtFR.PS} = .32$, $F(1,48) = 172.41$, $p < .0001$. The size of the regression coefficient was now also comparable to the other categories, $b = 0.23$. This raises the interesting possibility that in some cases, important properties are more salient for recall at the category level rather than at the item level. This makes a certain amount of intuitive sense – once a participant has responded with "fish" as a property of, say, a carp, why bother to mention "fish-like shape"? Thus, if this experiment were to be repeated using a property rating task, it might be appropriate to have some participants produce properties for the category names themselves, and to include these properties for rating.

Summary

There are two main findings from this experiment. First of all, family resemblance was found to be predictive of typicality for the non-referring categories, suggesting that non-referring concepts contain similarity structures. Secondly, the size of the effect was similar across the non-referring categories from this study, and the referring categories from this and the previous study of Malt and Smith (1984). These results suggest that non-referring concepts do not constitute a special case as far as their psychological structure is concerned. Thus the predictions of Hypothesis I are supported by this experiment – non-referring concepts contain similarity structures, just like referring concepts.

Experiment 3a: List Context

Experiment 3a was a follow-up to Experiments 2 and 3, designed to address a concern about the methodology of those studies. The categorization decisions made by participants in Experiment 2 sometimes appeared to be anomalous. For example, 16% of participants rejected Cookie Monster as an instance of MONSTER even though, from his name, Cookie Monster appears to be a monster by definition, and 35% of participants rejected a unicorn as an exemplar of ANIMAL even though 42 out of 50 participants in Experiment 1 listed it as some kind of animal, albeit an imaginary or mythical one. One possible explanation for this behavior is that rather than accessing a relatively stable concept for the target category in the typicality rating task, participants examine the list of items first and form an ad hoc concept based on what most of the items in the list appear to have in common. Thus, unicorns were rejected as a kind of animal because most of the other members of the list of "animals" were real animals, and Cookie Monster was rejected as a monster because he was the only cute and cuddly member of the list of "monsters".

This possibility raises two concerns. The first is a relatively minor concern about the current work. If participants are affected by the context of the list in Experiment 2, then the presence of the mismatched items in the MONSTER and ANIMAL category may have affected the typicality ratings in those categories. The concern is that if the data had been collected without the mismatched items, the mean typicality ratings would have been significantly different. This issue matters both to the findings of Experiment 3 and to the interpretation of the reaction time data in Experiment 4 below. The second concern is about the entire methodology of family resemblance studies. If participants form ad

hoc concepts for each category by scanning the list and forming a prototype based on what the exemplars and instances in that list have in common, then they can rate for typicality based on that contextually determined prototype. If this is the case, a positive correlation of those ratings with family resemblance (a direct measure of what the exemplars and instances have in common) becomes a self-fulfilling prophecy. To put it another way, perhaps Barsalou is right that all concepts are essentially ad hoc after all. As I argued in Chapter 3, this conclusion is certainly not warranted by Barsalou's own studies, but the new worry is that family resemblance studies as they have traditionally been conducted don't actually rule the possibility out either.

So the goal of the follow-up study described here was to check whether manipulating the context provided by the task would have a significant effect on the rating behavior of the participants. New typicality data was collected from two separate participant groups. Group A completed the task in the usual way, but the mismatched items were removed from the ANIMAL and MONSTER categories. Group B completed the task in a way that was designed to minimize any effect of list context. Instead of seeing a list of items on a page, participants saw the items one at a time displayed on a computer screen and responded using the computer keyboard. In this condition, items were presented randomly from 10 different categories to ensure that, on average, after responding to an item in any given category, participants would have to process nine more items from different categories before they saw an item from that first category again. To my knowledge, this is the first time that typicality data has been collected in a way that is specifically designed to upset the context of the exemplar list. If typicality ratings do cause participants to access invariant, stable concepts, then agreement within

participant groups should be comparable, and agreement between participant groups should be high (i.e., it does not matter how the typicality data is collected).

Method

Participants

The Group A participants were 32 student volunteers, 20 male and 12 female, with a mean age of 25.3 years. They were tested individually, and performed the task either for extra academic credit or for $10. All participants were native speakers or claimed fluency in English, and most (18) were born in the USA or Canada. All but five had lived in Canada or the USA for at least five years prior to the study. They performed the current rating task after having completed the speeded categorization task described in Experiment 4 below.

The Group B participants were 31 student volunteers, 13 male and 18 female, with a mean age of 20.5 years. They were tested individually, and performed the task for extra academic credit. All participants were native speakers or claimed fluency in English, and most (26) were born in Canada. All but one had lived in Canada or the USA for at least five years prior to the study.

Stimuli

For Group A, the stimuli were identical to those for Experiment 2, but excluding the mismatched items from the MONSTER and ANIMAL categories. Thus, the final lists consisted of 20 items in each of the four categories ANIMAL, MONSTER, FISH, and SUPERHERO. For Group B, the target stimuli were the same as for Group A, but also included six extra categories drawn from the data in the Appendix of Rosch (1975). 20

items were selected at random from the categories FURNITURE, WEAPON, VEGETABLE, SPORT, TOY, and CLOTHING.

Group A Procedure

For Group A, the instructions were identical to those given in Experiment 2, which were in turn based on the instructions that Rosch (1975) used. This time however, the instructions were read by the participants rather than being spoken out loud by the experimenter. Ratings were provided on paper in a booklet laid out exactly as in Experiment 2. Participants received the items within the categories in one of two random orders, and received the categories in a random order. No more than two participants saw the same random ordering of categories.

Group B Procedure

Group B provided ratings on a computer keyboard. Participants were seated in front of a 15-inch color monitor (AcerView 54e by Acer, Inc.) connected to a 133 MHz Pentium computer running SuperLab Pro, Version 2.0, by Cedrus Corporation. Stimuli were displayed in black on a white background. On each trial, participants first saw a category name (all lowercase) displayed using a 24-point, bold-faced Arial font, centered horizontally and 50 pixels above vertical center. Categories expressed as count nouns were displayed using the indefinite article (e.g., “a superhero”) while those expressed as mass nouns were not (e.g., “furniture”). The category name was followed, after 500 ms, by the name of the item to rate (all lowercase), centered both horizontally and vertically, and displayed using a 30-point, bold-faced Arial font with no indefinite article. The category name remained on the screen during the entire trial, and participants had an

unlimited time to respond. The response keys were chosen from the top row of a standard computer keyboard. These keys were relabeled such that seven keys represented the 7-point response scale numbered in increasing order from left to right. Two additional keys to the right were specially labeled as the "×" and "?" keys. After responding, there was a 1000 ms inter-stimulus interval before the next trial began.

The participants were given 10 practice trials consisting of categories and items not included in the main study. The main block of trials involved all 200 category/item pairs, presented in a different random order for each participant. The instructions to participants in Group B were provided on paper for participants to read individually. They were almost identical to Group A's (and Rosch's) instructions, but modified to reflect the fact that the participants were responding on a computer keyboard. The instructions were as follows:

> This study has to do with what we have in mind when we use words which refer to categories. Let's take the word *red* as an example. Imagine a true red. Now imagine an orangish red … imagine a purple red. Although you might still name the orange-red or the purple-red with the term *red*, they are not as good examples of red (not as clear cases of what *red* refers to) as the clear "true" red. In short, some reds are redder than others. The same is true for other kinds of categories. Think of birds. Everyone has some notion of what a "real bird", a "birdy bird", is. To me a sparrow or a robin is a very birdy bird while an ostrich is a less birdy bird. Notice that this kind of judgment has nothing to do with how well you like the thing; you can like a purple red better than a true red but still recognize that the color you like is not a true red. You might think that the ostrich is much more interesting than other kinds of birds without thinking that it is the kind of bird that best represents what people mean by birdiness. This judgment also has nothing to do with how frequently you see or think about the thing; you could live on an ostrich farm and deal with ostriches every day, but still think that they are a pretty bad example of what people generally mean when they talk about birds.
>
> In this study you are asked to judge how good an example of a category various instances of that category are. The computer will present a category followed by an item. You are to use the buttons labeled 1 to 7 to rate how good an example of the category each instance is. A 7 means that you feel the item is a very good example of your idea of what the category is. A 1 means you feel the member fits very poorly with your idea or image of the category. A 4 means you feel the member fits moderately well. It is also possible that you may not know what the

> example is, in which case you should press the button labeled "?". You also may feel that the instance is not actually a member of the category at all, in which case you would press the button marked "X". You can go at your own speed, and you will be given a chance to practice before the experiment starts.
>
> Don't worry about *why* you feel that something is or isn't a good example of the category (and don't worry about whether it's just you or people in general who feel that way) – just mark it the way you see it.

Results

The full set of mean typicality ratings for both groups can be found in Appendix D. Table 4-10 below shows the four target categories from Experiment 2 and the set of Pearson correlations between group mean typicality ratings. The responses were substantially and significantly correlated across all three groups of participants for every category (the mean correlations between the three pairs of participant groups were: r_{avg} = .94, for ANIMAL; r_{avg} = .92, for FISH; r_{avg} = .85, for MONSTER; and r_{avg} = .92, for SUPERHERO). Table 4-11 shows the correlations between the Group B participants and the 209 participants from Rosch's (1975) study for the six non-target categories. Once again, all Pearson correlations are significant and substantial, ranging from r = -.81, p < .01, for TOY to r = -.97, p < .01, for WEAPON (these correlations are negative because Rosch used an inverted 7-point scale). Spearman rank-order correlations (preferred by Rosch for these types of studies) are also shown in Table 4-11, along with the same correlations for the four target categories.

Table 4-10
Correlations Between Groups for Mean Typicality Ratings

	Category			
Groups	ANIMAL	MONSTER	FISH	SUPERHERO
Exp. 2 & Grp. A	.96	.80	.96	.89
Exp. 2 & Grp. B	.95	.91	.90	.92
Grp. A & Grp. B	.92	.85	.90	.95

Note. All correlations are significant (p <.01).

Table 4-11
Group B Mean Typicality Correlations

Category	Pearson	Spearman
Correlations with Rosch (1975b)		
CLOTHING	-.94	-.94
FURNITURE	-.93	-.92
SPORT	-.89	-.79
TOY	-.81	-.82
VEGETABLE	-.84	-.71
WEAPON	-.97	-.97
Correlations with Group A		
ANIMAL	.95	.93
FISH	.90	.87
MONSTER	.91	.85
SUPERHERO	.92	.91

Note. All correlations are significant ($p < .01$).

As for within-group agreement, Table 4-12 shows the available data for split-half correlations over all 10 categories. No exact values are available for Rosch's six categories, except for her report that all split-half correlations were above $r = .90$. The values from the other three groups reflect split-half correlations based on random splits. Table 4-13 shows mean pair-wise correlations calculated using the method of Guildford and Fruchter (1973) following Barsalou's (1987) suggestion. This time no data is available from Rosch, who never calculated this statistic. However, Barsalou (1987) reported that he had found mean correlations of between .30 and .70 in his own studies involving numerous categories and participant groups. The final test for agreement within groups is the χ^2 test for every item against a flat distribution. No figures are available for the current data from Rosch, though she did report in a separate study (with 113 participants) that 96% of the tests for her items were significantly different from the flat distribution (Rosch, 1973b). In Experiment 2, with 49 participants, the corresponding

figure was 88%. In this study, 83% of the tests reached significance in Group A, but only 65% reached significance in Group B, although that figure rises to 75% if only the four categories from Group A are included. (Note that χ^2 tests are highly dependent on sample size. The implications of this fact will be discussed further below.)

Table 4-12
Random Split-Half Correlations

Category	Rosch	Exp. 2	Grp. A	Grp. B
ANIMAL		.94	.94	.94
CLOTHING	>.90			.97
FISH		.92	.78	.84
FURNITURE	>.90			.87
MONSTER		.77	.81	.78
SPORT	>.90			.98
SUPERHERO		.89	.83	.93
TOY	>.90			.94
VEGETABLE	>.90			.83
WEAPON	>.90			.96

Note. All correlations are significant ($p < .01$)

Table 4-13
Mean Pair-Wise Correlations

Category	Exp. 2	Grp. A	Grp. B
ANIMAL	.49	.64	.52
CLOTHING			.70
FISH	.46	.52	.39
FURNITURE			.64
MONSTER	.22	.24	.22
SPORT			.58
SUPERHERO	.37	.45	.42
TOY			.43
VEGETABLE			.61
WEAPON			.56

Table 4-14 shows the Pearson correlations between family resemblance (from Experiment 3) and mean typicality across the three participant groups, using the simple correlation method of Rosch and Mervis (1975). Correlations with family resemblance

for the ANIMAL and SUPERHERO categories were again significant for both Group A and B. Correlations for the FISH category were again not significant for Groups A and B, likely for the same reasons as noted in the discussion of Experiment 3. Only the MONSTER category showed a change in the size and significance of the correlations between family resemblance and typicality. Whereas in Experiment 2 this correlation had been significant, these correlations were not significant for either Group A or B of the current study. Scatter plots of typicality against family resemblance for Groups A and B for the MONSTER category can be found in Appendix D.

Table 4-14
Correlations Between Mean Typicality and Family Resemblance

	Exp. 2		Grp. A		Grp. B	
Category	r_{MTFR}	*b*	r_{MTFR}	*b*	r_{MTFR}	*b*
ANIMAL	.58**	.021	.67**	.033	.48*	.018
MONSTER	.48*	.021	.26	.010	.37	.016
FISH	.29	.016	.26	.016	.23	.010
SUPERHERO	.67**	.022	.58*	.014	.55*	.021

Finally, Table 4-15 shows the correlations between raw typicality and family resemblance using the more sensitive method recommended by Lorch and Meyers (1990), as discussed in the Results section of Experiment 3. All correlations in this table are significant ($p < .001$) and the comparisons across the three groups are more consistent than for the correlations involving mean typicality. In particular, the MONSTER correlation for Group B is now closer to the result for the Experiment 2 participants under this analysis.

Table 4-15
Correlations Between Raw Typicality and Family Resemblance

Category	Exp. 2		Grp. A		Grp. B	
	$r_{RTFR.PS}$	*b*	$r_{RTFR.PS}$	*b*	$r_{RTFR.PS}$	*b*
ANIMAL	.34	.021	.45	.033	.29	.018
MONSTER	.25	.021	.11	.010	.24	.016
FISH	.18	.016	.15	.016	.14	.010
SUPERHERO	.36	.022	.29	.014	.30	.021

Note. All results are significant ($p < .001$).

Discussion

The purpose of this follow-up experiment was to evaluate the extent to which the context of the list of items in a standard typicality rating task affects the performance of participants. For the target categories, the within- and between-groups correlations were comparable across all three groups (most striking are the strong significant correlations between Group B and Rosch's participants, respectively located in Ontario in 2002 and California in 1975). The only exception was that the number of items with significant χ^2 tests was smaller for the participants in Group B. But this result was most likely due to the much smaller number of participants (if all the groups had been as large as Rosch's, 1973b, comparison group, the percentage of items reaching significance, assuming the same distribution of responses, would have been 97% for Experiment 2, 95% for Group A, and 96% for Group B).

In most cases, the ratings of all three groups of participants were also correlated to about the same extent with the family resemblance scores from Experiment 3. The only significant change was in the MONSTER category where the mean responses of the Experiment 2 participants were correlated significantly with family resemblance, but the mean responses of Groups A and B from this experiment were not. This result did not extend to the analysis recommended by Lorch and Myers (1990), in which all

correlations were strongly significant, but note that the value of this correlation for the Group A participants is quite low. However, these lower correlations may just reflect the fact that the MONSTER category is the least stable of the categories across participants, as shown by most measures of inter-participant agreement in Experiment 2, and by the lower mean correlations between and within groups found in this experiment. Given this lower level of agreement, the MONSTER category was the most likely to fail to show a correlation between the family resemblance scores computed from the property lists gathered from one group of participants and the typicality ratings gathered from a different group. Recall that there was considerable overlap between the participants used in Experiments 2 and 3 (i.e. 45 of the 49 Experiment 2 participants also took part in Experiment 3), but no overlap between those used in Experiment 3 and Groups A and B of Experiment 3a. Thus the lower level of correlation is probably a further reflection of lack of inter-participant agreement regarding the MONSTER category.[49]

There is little evidence from this study that either small or large modifications to the context provided by the task has a significant effect on the overall typicality ratings of the participants, and it seems reasonable to conclude that participants are in fact able to access a stable, relatively invariant concept when presented with a category name.

Experiment 4: Speeded Categorization

There were two goals to Experiment 4. The first was to replicate, for the non-referring categories, Rosch's (1973b) results involving speeded categorization. Rosch found that both adult and child participants were significantly faster at categorizing typical (or "central") than atypical (or "peripheral") category members, an effect that has since proven to be extremely robust and reproducible (see Chang, 1986). Rosch's result implies

that there are differential processing effects (for more and less typical category members) that are tied to the underlying similarity, or family resemblance, structure of any category, and a number of computational models have subsequently been developed to account for these results. Extending this result to non-referring categories would provide further support for Hypothesis I. The second goal was to investigate separately Hypothesis II, that non-referring concepts are no more complicated to process than referring concepts. If they are more complicated to process, a sensible place to look for evidence of the difference would be in reaction times to tasks such as speeded categorization.

Method

Participants

The participants were 20 student volunteers, 5 male and 15 female, with a mean age of 26.1 years. They were tested individually, and performed the task either for extra academic credit or for $10. All participants were native speakers of English, and had lived in Canada or the USA for at least 20 years. This set of participants represented all the native English speakers from a larger original pool of 32 participants. For this study, in which quick access to English lexical items was of paramount importance, it was imperative to exclude non-native speakers of English. (The original group of 32 also served as Group A of Experiment 3a. The Experiment 3a task was performed after the current one.)

Stimuli

The stimuli used in this study were drawn from the same set of stimuli used in Experiment 2. For each of the four categories, five high-typicality and five low-typicality stimuli were selected. The procedure for selection was as follows. Starting from the item with the highest mean typicality from Experiment 2 and moving down the sorted list towards the lowest mean typicality, the first five single-word items meeting the following criteria were selected: low written frequency, defined as less than 10 in the Kučera and Francis (1967) norms; and relatively high familiarity to the participants, defined as five or fewer question mark responses in Experiment 2. The low-typicality items were selected in the same way, but starting with the item with the lowest mean typicality and working upwards. In the FISH category, all items whose name contained the word "fish" were removed, and an exception to the familiarity rule had to be made in order to obtain 10 items. Thus "pickerel", which received nine question marks in Experiment 2, was allowed as one of the stimuli. For each stimulus, a "foil" category was also selected to be used to elicit negative categorization judgments from the participants. The list of all the stimuli can be found in Appendix E.

Tables 4-16 and 4-17 below show the mean word frequencies and number of question marks from Experiment 2 for each of the four groups of high or low typicality and referring or non-referring items (averaged across items and categories). Table 4-18 shows the mean typicality for the stimuli in each group of items. Note that no attempt was made to control for the word length of the stimuli, as it has been shown that categorization reaction times are not affected by word length (Terry, Samuels, &

Laberge, 1976). However, for completeness, Table 4-19 shows the mean word lengths for the four groups of items.

Table 4-16
Mean Word Frequencies for the Items in the Study

	Typicality Level		
Category Type	High	Low	All
Non-Referring	1.20	0.50	0.85
Referring	3.80	0.60	2.20
All	2.50	0.55	1.53

Table 4-17
Mean Number of Question Marks from Experiment 2
for the Items in the Study

	Typicality Level		
Category Type	High	Low	All
Non-Referring	0.80	1.30	1.05
Referring	1.40	1.10	1.25
All	1.10	1.20	1.15

Table 4-18
Mean Typicality for the Items in the Study

	Typicality Level		
Category Type	High	Low	All
Non-Referring	6.15	4.76	5.45
Referring	6.44	4.45	5.45
All	6.29	4.61	5.45

Table 4-19
Mean Word Length for the Items in the Study

	Typicality Level		
Category Type	High	Low	All
Non-Referring	7.70	6.50	7.10
Referring	5.70	6.30	6.00
All	6.70	6.40	6.55

Procedure

This experiment was a standard categorization task. Participants were seated in front of a 15-inch color monitor (AcerView 54e by Acer, Inc.) connected to a 133 MHz Pentium computer running SuperLab Pro, Version 2.0, by Cedrus Corporation. Stimuli were displayed in black on a white background. On each trial, participants would see a category name (all lowercase) with a question mark, displayed using a 24-point, Arial font, centered horizontally and 50 pixels above vertical center. This was followed, after 1000 ms, by a fixation cross, centered both horizontally and vertically, which was replaced after 1000 ms with the name of the item to categorize (all lowercase), displayed using a 32-point, bold Arial font. The category name remained on the screen during the entire trial. Participants had 5 s to respond once the item was displayed. They responded by pressing one of two colored buttons labeled “Y” (green) and “N” (red) on a Cedrus RB-620, six-button response pad. Participants used the index finger of their dominant hand, which rested between trials on a yellow button between the “Y” and “N” buttons. For half the participants, the “Y” button was positioned on the thumb side of their index finger (left for right-handed participants, right for left-handed participants), and for the other half, the “Y” button was positioned on the opposite side (right for right-handed

participants, left for left-handed participants). After responding, participants waited through a 1500 ms inter-trial interval before the next trial began.

Participants received four blocks of 80 trials, each containing all 40 positive and all 40 negative categorizations (using the category foils shown in Appendix E). Stimulus order was randomized differently for each block and for each participant. Between blocks, participants were allowed to take a break for as long as they wanted. Most participants completed the task in about 25 minutes. The instructions, given to the participants in print, were as follows:

> This experiment is designed to find out how quickly people can categorize various things. You will be seated in front of a computer with your dominant hand on a response pad. The pad has buttons labeled “Y” for “yes” and “N” for “no”. Rest the index finger of your dominant hand on the yellow button between the “yes” and “no” buttons.
>
> When the experiment is running, you will first see a category name with a question mark appear near the middle of the screen (for instance “fruit?”). Then you will see a cross appear below the category name. Focus on the cross. After a short time, the cross will be replaced by the name of an item (for instance “apple” or “martian”). Your job is to press the “yes” button if the item matches the category, or press the “no” button if it does not. So for “apple”, you should answer “yes” because it is a fruit, but for “martian” you should answer “no”. (Don’t worry, you’ll get to practice before the experiment begins.)
>
> Try to answer as quickly as you can, while making as few mistakes as possible.
>
> If you take too long (more than 5 seconds), the computer will go on to the next example on its own.
>
> It is possible you might not know what one of the items is. If this happens, don’t press either button. Just wait for the next example (this will take about 5 seconds).
>
> There will be a short practice session to make sure you are pressing the right buttons, then there will be a longer practice session to help you get used to the experimental task. The experiment itself consists of 4 sets of examples. Each set should take under 10 minutes, and you can take a

> break between sets – the computer will wait until you are ready to start again.
>
> That's it – please let the experimenter know you have finished reading this.

Results

Each participant responded four times to each item under objectively positive and negative categorization conditions, resulting in the collection of a total of 6400 response times. Responses that did not match the target response (e.g., a response of "N" to the category-exemplar pair "monster, frankenstein" or "Y" to "animal, witch") were removed, as were null responses (in which the participant did not hit a key). There were 213 mismatched responses and 32 null responses in the data. Two steps were taken to control for outlying data points. First, all responses slower than 3000 ms were removed from the analysis, though this criterion excluded only six responses. Second, the results reported below all involve the median response time over the maximum of four possible responses to each item rather than the mean (there were 1581 of these median values after the removal of the mismatched, null, and slow responses). Identical analyses were also performed on the mean scores with few inconsistencies in the results. The analyses involving means are only reported when they differ from the analyses involving the medians.

The median response times to each item for each participant were analyzed within two separate five-factor ANOVAs, one each for the positive and negative categorizations. Within each ANOVA, there were three factors regarded as fixed effects: typicality (high, low), reference type (referring, non-referring), and category level (superordinate, basic). Both the item stimuli (i.e., five items nested within each of the three fixed factors) and

participants were treated as random effect factors. The fixed effect factor of category level is of lesser interest to the current study than typicality and reference type, but was included in part to test the intuition, mentioned in the description of Experiment 2, that SUPERHERO and FISH are plausibly paired as basic level categories while MONSTER and ANIMAL can be paired as superordinate categories. One well-established result for referring concepts is an effect of category level – categorization at the basic level is faster than at the superordinate level. This is usually known as the "category size effect", since basic level categories are often nested within superordinate categories and therefore have fewer exemplars (Chang, 1986). So the inclusion of category level in the ANOVA checks for a source of variance that would be expected if the intuition about the levels of the non-referring categories is correct, and can also be used to check more thoroughly for possible interaction effects of this factor with typicality and reference type.

Table 4-20 shows these response times for each of the four conditions obtained by crossing the typicality and reference factors (i.e., averaged over categories, items, and participants within each condition). Table 4-21 shows the response times obtained by crossing the typicality and category level factors, and Table 4-22 show those obtained by crossing the reference and category level factors.

Table 4-20
Mean Median Response Times by Typicality and Reference

Positive Categorizations			
	Typicality Level		
Reference	High	Low	All
Non-Referring	856	922	889
Referring	782	859	820
All	819	891	854

Negative Categorizations			
	Typicality Level		
Reference	High	Low	All
Non-Referring	859	907	883
Referring	816	817	817
All	838	862	850

Table 4-21
Mean Median Response Times by Typicality and Category Level

Positive Categorizations			
	Typicality Level		
Category Level	High	Low	All
Basic	792	838	815
Superordinate	846	944	894
All	819	891	854

Negative Categorizations			
	Typicality Level		
Category Level	High	Low	All
Basic	839	838	839
Superordinate	836	886	861
All	838	862	850

Table 4-22
Mean Median Response Times by Category Level and Reference

Positive Categorizations			
	Category Level		
Reference	Basic	Super	All
Non-Referring	836	943	889
Referring	793	846	820
All	815	894	854

Negative Categorizations			
	Category Level		
Reference	Basic	Super	All
Non-Referring	861	905	883
Referring	816	817	817
All	839	861	850

To derive the appropriate F ratios and test the effects of interest within such a design (i.e., those involving the three fixed factors), the expected mean squares for the variance components associated with each of the effects were determined according to the method prescribed by Kirk (1995), and then appropriate quasi-F ratios were formed for each of the statistical tests (Clark, 1973; Kirk, 1995; Richter & Seay, 1987). Although such tests could be regarded as being somewhat conservative (i.e., the actual Type I error can sometimes be smaller than the nominal level; Richter & Seay, 1987),[50] their use ensures that the statistical results obtained should generalize to other samples of stimulus items (or, conversely, ensures that they do not simply represent some idiosyncratic selection of the stimulus items; Clark, 1973). For cases in which no response times were available from a participant for a certain item (14 out of a possible 800 for the positive categorizations and 5 out of a possible 800 for the negative categorizations), the group mean was substituted for that item in order to run the analysis.

In the analysis of the positive categorization response times, there were only significant main effects of typicality, $F'(1, 50) = 11.48$, $p < .01$ (participants were faster to categorize highly typical items), reference type, $F'(1,47) = 7.51$, $p < .01$ (participants were faster to categorize referring items), and category level, $F'(1,49) = 10.80$, $p < .01$ (participants were faster to categorize basic level items).[51] In the analysis of the negative categorization response times, there was only a significant main effect of reference type, $F'(1, 50) = 7.80$, $p < .01$, but not of typicality, $F'(1,47) = 1.36$, $p > .20$ or category level, $F'(1,45) = 1.21$, $p > .25$. There were no interaction effects in either positive or negative reaction times.

Discussion

The analysis described above treated both the participants and items as random variables, and thus can be expected to generalize both to new participant samples and new item samples. In the statistical analysis, a main effect of both typicality and category level was found for positive categorizations (highly typical items and basic level categories lead to shorter reaction times than less typical items and superordinate categories). These effects did not show up for negative categorizations. These analyses also showed a main effect of reference type for both positive and negative categorizations (response times were shorter for referring terms). There were no significant interactions between any of the three factors for positive or negative categorizations.

The main effect of typicality for the positive categorizations matches what has been found in past studies. The result indicates that items that are more typical can be categorized more quickly, and moreover, the lack of an interaction with either reference type or category level shows that both non-referring and referring categories exhibit this

effect to the same degree. As has also been found in many previous studies (Chang, 1986; Rosch, 1973b), there is no typicality effect on response time for the negative categorizations. This pattern of results can be explained by a number of computational models of similarity or prototype structure, such as the property weight summation process explored by Smith and Medin (1981) or Nosofsky's (1988) exemplar model. Thus, this finding lends further support to Hypothesis I, in that it provides evidence compatible with the inclusion of a similarity-based structure in both referring and non-referring concepts.

The main effect of category level indicates that the original intuition about the levels of the categories was correct. The items in the more basic level categories (i.e., FISH and SUPERHERO) were categorized significantly faster than those in the more superordinate categories (i.e., ANIMAL and MONSTER) for positive judgments, but not for negative judgments. Again, the lack of interaction with reference type shows that the non-referring SUPERHERO and MONSTER categories exhibited this effect to the same degree as the referring categories, in which FISH is known to be basic and ANIMAL is known to be superordinate (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). These findings are all consistent with the well-established effect of category level in the literature on semantic memory (Chang, 1986), and thus can be taken as evidence that the original intuition was correct: SUPERHERO is indeed basic and MONSTER is more superordinate. As such, this finding demonstrates another structural parallel between referring and non-referring categories.

The main effect of referential category type, for both positive and negative categorizations, is the big surprise of this experiment, as it contradicts Hypothesis II, that

non-referring terms are no more difficult to process than referring terms. Apparently non-referring terms do take longer to process, at least for the purposes of categorization. There are three possible ways to account for this effect. The first, obviously, is to accept that Hypothesis II is wrong. Despite having an identical structure, there must be something about non-referring concepts that makes them more difficult to process in categorization. This conclusion is very tempting because, in addition to the above analyses, there is another trend in the data that also supports it.

In the positive categorizations, items were necessarily paired with categories of the same reference type. But such pairings could not always occur for the negative categorizations. Some items had foil categories of the same referential type (e.g., “fish” with “bear”) whereas others had foil categories of the opposite referential type (e.g., “fish” with “witch”). When the negative categorization data is rearranged along the dimensions of foil and item type (referring vs. non-referring), an even more compelling trend emerges. Table 4-23 shows that the referential type of both the item and the foil category appear to have an effect on negative categorization times, with the greater effect coming from the referential type of the item and a slightly lesser effect coming from the referential type of the category foil. Of course the original experiment was not designed to evaluate this trend among negative categorizations, so the finding should be taken tentatively. Nevertheless, it is consistent with the hypothesis that participants simply take longer to make categorization decisions about non-referring terms.

Table 4-23
Mean Median Response Times for Negative Categorizations by Type of Item and Foil

	Item Type				
	Non-Referring		Referring		
Foil Type	Mean	N[a]	Mean	N	All
Non-Referring	955	7	821	13	862
Referring	844	16	798	4	833
All	883	23	817	17	850

[a] Number of items in each cell.

So there appears to be evidence that non-referring terms take longer to process than referring terms. But there are at least two other possibilities that need to be ruled out before this conclusion can be accepted. The first is that there is some known variable, that was not controlled for carefully enough, and the second is that there is some unanticipated variable that was not controlled for at all, that could account for the effect. Regarding the first possibility, the items were carefully selected to be matched as well as possible according to a number of variables. The most successful matching was by typicality. As Table 4-18 above shows, the mean typicality rating across the referring and non-referring items was exactly the same. Thus, a difference in overall typicality between referring and non-referring terms can be ruled out. But two other variables were not quite so perfectly well-matched. All the items were chosen to have both a low word frequency according to the Kučera and Francis (1967) norms, and a high familiarity to participants as measured by the number of question mark responses in Experiment 2. But as Table 4-16 above shows, the referring terms did have a slightly higher mean word frequency (2.2 as opposed to 0.9 for the non-referring terms), and Table 4-19 shows that the non-referring terms were slightly longer (7.1 vs. 6.0 characters in length). But these differences are quite small compared to the relatively large effect of the approximately 69

ms overall difference in median response time shown in Table 4-20, and are thus unlikely to fully account for the effect of reference type.

The possibility remains that an unanticipated, and therefore uncontrolled, variable slowed down the participants' responses to the non-referring terms. The most obvious candidate for such a variable is familiarity. It is possible that, despite the very low word frequencies for all the items in the study and despite the low number of question marks circled for the items in Experiment 2, participants were nevertheless much more familiar with the referring terms than the non-referring terms. As reviewed in Chapter 3, it is well-established that participant familiarity with test items can have an impact on their speed of recall and reasoning about the items (see also Chang, 1986). Note, as well, that it is quite plausible that word frequency tracks familiarity, and it is this fact that explains its power in accounting for reaction times in tasks such as lexical decision (e.g., Whaley, 1978). The overall levels of word frequency for the items in this study did not differ much, but most of them had a frequency of zero, simply because they were not common enough to appear in the corpora studied by Kučera and Francis (1967). Thus, there could well be a much larger difference in participant familiarity with the items than there appears to be from the word frequency counts and numbers of question mark responses in Experiment 2. This possibility was tested in a follow-up to this experiment, described as Experiment 4a below.

Experiment 4a: Familiarity

The purpose of this follow-up experiment was to collect familiarity ratings from participants for the items used in the previous studies, and then use those familiarity ratings to try to account for the effect of referential category on reaction times identified

in Experiment 4. Familiarity is impossible to measure objectively with any method other than simply directly asking participants how familiar they are with the items in question. This technique has been used in a number of previous studies, including studies by Barsalou (1985) and Hampton and Gardiner (1983), and it is the technique that was used in this experiment as well.

Method

Participants

The participants were the same individuals that formed Group B of Experiment 3a. They all performed the familiarity rating task immediately after the Experiment 3a typicality task.

Stimuli

The stimuli were the same 20 items in each of the four categories that were used in Experiment 3a, plus an additional four items representing the category names themselves: “monster”, “animal”, “fish”, and “superhero”.

Procedure

Participants were given a booklet with the 84 items arranged over three pages in one of two random orders. Items were not grouped by category. The items were listed down the left-hand side of the page. To the right of each item was a 7-point scale for rating the familiarity of the item (7 = highly familiar, 1 = highly unfamiliar) as well as a question mark that participants could circle if they had never heard of the item at all. The instructions, printed on the front of the booklet and read individually by the participants, were as follows:

> In this task, you will be presented with a list of individuals and categories, and you are to use a 7-point scale to rate them for familiarity. If you feel that you are very familiar with the individual or type of thing named, you should circle the 7. If you are not at all familiar, circle the 1. If you are moderately familiar, circle the 4, and so on. There is also a question mark that you should circle only if you have no idea at all what something is.
>
> How familiar you are with something has nothing to do with how much you like it or how well you know the word that names it. Your familiarity with it has more to do with how much you know about the thing itself, or how often you encounter it, see pictures of it, read about it, and so on.
>
> Here's an example. Suppose the item is "Bugs Bunny" and you have no idea who Bugs Bunny is. In that case you should circle the question mark. If you know who he is, but feel that you are very unfamiliar with him, perhaps because you know almost nothing about him, or perhaps because you have hardly ever seen him on TV, you should circle the 1. If you are very familiar, you circle the 7. If you are moderately familiar, circle the 4, and so on.
>
> An example of the layout of the items is shown below:

	very familiar						not at all familiar	don't know what this is
Bugs Bunny	7	6	5	4	3	2	1	?

> One more important note: In all cases, you should interpret the words as naming an object, category, or individual. For instance, the above example is talking about Bugs Bunny the animated rabbit character, not Bugs Bunny the TV show.
>
> Following this are 3 pages with about 80 items on them as shown above. Feel free to complete the form at your own speed, but keep in mind that even if you go quite slowly, the form won't take more than 5 minutes to fill out.

Results

The full set of familiarity ratings are shown in Appendix F alongside the corresponding mean reaction times from the positive categorizations in Experiment 4. A quick glance at these results shows that familiarity does not appear to track reaction time. The mean familiarity ratings are presented in Table 4-24 below, split up according to reference type and category level such that each cell represents a single category. Once again, a quick glance at these results shows that familiarity does not appear to track reference particularly well either. This latter intuition can be quantified. If participants are less

familiar on average with non-referring items, then most of the non-referring items should receive familiarity scores below the median. In fact, only about half (24 out of 42) of the non-referring terms were below the median familiarity score of 5.21. A Pearson correlation computed between familiarity and reference (coded as 1 and –1) was not significant: $r = .13$, $p > .20$. By the same measures, however, there does appear to be a relationship between category level and familiarity: 31 out of 42 basic level terms were below the median familiarity score, $r = .51$, $p < .001$. There is also some evidence of a relationship between familiarity and typicality. Table 4-25 below shows the correlations between the mean familiarity and the mean typicality ratings corresponding to the group of participants from Experiment 2, and Groups A and B from Experiment 3a.[52] Of the four categories, only the ANIMAL category showed any sign of a significant correlation.

Table 4-24
Mean Familiarity Ratings

	Category Level		
Category Type	Basic	Super	All
Non-Referring	4.59	5.24	4.92
Referring	4.38	5.98	5.18
All	4.49	5.61	5.05

Table 4-25
Correlations Between Mean Familiarity and Mean Typicality

	Exp. 2	Grp. A	Grp. B
ANIMAL	.62**	.53*	.80***
FISH	.03	.10	.20
MONSTER	-.28	-.14	-.26
SUPERHERO	.39	.30	.37

*** indicates p < .001.

To more rigorously check the possibility that what appeared to be an effect of reference type in Experiment 4 was actually a familiarity effect, a repeated measures

regression analysis was performed with familiarity and typicality as the main predictors. For each participant, median response times to each of the positive categorizations were regressed against typicality, familiarity, and their interaction (all continuous variables) using the method recommended by Lorch and Myers (1990). In the overall analysis, the interaction variable did not account for a significant amount of variance in the median response times, $F(1,19) = 3.81$, $p > .05$. The interaction term was thus removed from further analysis as prescribed by Pedhazur (1997). In the subsequent analysis, involving only main effects, there was a significant relation between median response times and typicality, $F(1,19) = 29.23$, $p < .001$, $b = -39.0$, but not familiarity, $F(1,19) = 0.95$, $p > .30$.

For comparison, a similar analysis was conducted with typicality (as a continuous variable) and reference type (referring, non-referring) as the main predictors. In the overall analysis, the interaction did not account for a significant amount of variance in the median response times, $F(1,19) = 1.26$, $p > .25$. The interaction term was thus removed from further analysis. In the subsequent analysis, involving only main effects, there was a significant relation between median response times and both typicality, $F(1,19) = 27.18$, $p < .001$, $b = -38.8$, and referential category, $F(1,19) = 11.68$, $p < .01$, difference = -69 ms.

For each participant, median response times to each of the negative categorizations were also regressed against typicality, familiarity, and their interaction using the same method. This time, the interaction variable was significant, $F(1,19) = 5.05$, $p < .05$, $b = -10.4$ ms, and could not be removed from further analysis. However, there was no main effect of either typicality, $F(1,19) = 4.14$, $p > .05$, or

familiarity, $F(1,19) = .24$, $p > .60$. The mean response times analysis did not agree for negative categorizations. The interaction variable was not significant, $F(1,19) = 3.67$, $p > .05$, and was dropped. In the subsequent analysis, typicality was not significant, $F(1,19) = .14$, $p > .70$, but familiarity was significant $F(1,19) = 9.06$, $p < .01$, $b = -16.2$.

Again for comparison, a similar analysis was conducted using typicality and reference type as the main predictors. The interaction variable did not account for a significant amount of variance in the median response times, $F(1,19) = 3.47$, $p > .10$, and was removed from further analysis. In the subsequent analysis, involving only main effects, there was a significant relation between median response time and referential category, $F(1,19) = 23.00$, $p < .01$, difference = -64 ms, but not typicality, $F(1,19) = .24$, $p > .25$.[53]

Discussion

It seems clear that the effect of referential type cannot be explained as a familiarity effect. In the regression analyses involving typicality and reference type, both factors were significant for positive categorizations, but when familiarity replaced reference in the analysis, the only significant factor was typicality. For negative categorizations, only reference type was significant in the analysis involving typicality and reference. In the analysis involving familiarity and typicality, the analysis of the mean response times did show a significant effect of familiarity in negative categorizations, but in terms of accepting familiarity as a predictor to replace reference type, this does not make up for the lack of familiarity effect for positive categorizations. In fact, the possible presence of the effect and the disagreement between means and median analyses is probably a further

symptom of the foil category and relatedness confounds discussed above for Experiment 4.

Summary and Implications

The results of Experiments 1 to 4 together provide strong confirmation for Hypothesis I, which stated that non-referring concepts include similarity-based structures, just as referring concepts do. Experiments 2 and 3 showed, for the two large non-referring categories MONSTER and SUPERHERO, that, just as for the referring categories, participants willingly produce consistent typicality judgments that correlate well with independent measures of family resemblance. Although in the typicality judgment study, participants showed more signs of disagreement and conceptual instability for non-referring than referring categories, in the family resemblance study, the non-referring categories were shown to be at least as tightly bound together by similarity as the referring categories. Further support for Hypothesis I comes from Experiment 4, in which the well-established effects of typicality and category level on positive categorization response times were replicated for both referring and non-referring categories.

The confirmation of Hypothesis I provides an argument against at least some of the special case solutions to the problem of non-reference (that is, those that make claims about a special structure) and in favor of the Psychological Meinongian position outlined in Chapter 2, that proposes little or no difference between referring and non-referring concepts. The confirmation of this hypothesis also counts against the competing theories of conceptual structure reviewed in Chapter 3 – ad hoc structure for non-referring concepts is refuted, and so is the idea that non-referring concepts are structured exclusively around ideals, essences, or definitions.

Experiment 4 was also designed to test Hypothesis II, which states that the processing of non-referring concepts is no more complicated than that of referring concepts. The finding of a main effect of reference for both positive and negative categorizations, and the resistance of this effect to explanation by other factors, does not support this hypothesis and leaves open the possibility that, although the structure of non-referring terms may not differ, they may still be more difficult to process than referring terms. Possible reasons for this and implications for future work will be explored in Chapter 5.

Chapter 5: Back to Reality

> The difference between fiction and reality? Fiction has to make sense.
>
> – Tom Clancy

A Quick Recap

The Main Questions:

1. What is the psychological structure of non-referring concepts, and does it differ in any way from that of ordinary referring concepts?

2. Are there any other significant psychological differences between referring and non-referring concepts?

Figure 5-1: *The two main questions addressed in this dissertation (reproduced from Figure 1-2).*

I started this project with the two questions reproduced in Figure 5-1. The motivation for these questions came from a problem with the highly influential reference-based approach to semantics: If most words get their meanings by referring to things, then how do seemingly meaningful words with no existing referents get their meanings? The default position in a reference-based framework is that non-referring words, and the concepts that they express, are semantic special cases. In Chapter 2, I reviewed some of the better-known proposals regarding the semantics of non-referring terms, drawing out a number of competing predictions about the non-referring concepts that they express, including Fodor and Dretske's prediction that non-referring concepts have a radically different structure from referring concepts, Taylor's prediction that non-referring concepts have an identical structure, but require an extra processing step, and my own tentative suggestion that there is no difference in either structure or processing – I called

this position "Psychological Meinongianism". Then in Chapter 3, I reviewed some of the most important psychological evidence to date regarding the structure of various kinds of concepts – object concepts, abstract concepts, and ad hoc concepts, in particular. That chapter ended with the predictions of Psychological Meinongianism encoded into two testable hypotheses:

> Hypothesis I: *Non-referring object concepts include similarity-based structures, just like referring object concepts do.*
>
> Hypothesis II: *The processing of non-referring object concepts is no simpler or more complicated than the processing of referring object concepts.*

Finally in Chapter 4, I reported the results of a set of experiments that provided support for Hypothesis I, but contradicted Hypothesis II. The results of that experimental work are summarized in Table 5-1.

In what follows, I will argue that these results are compatible with a weakened version of Psychological Meinongianism, according to which, although there is no difference in the structure of referring and non-referring concepts, knowledge about the real world, namely the knowledge that non-referring concepts have no real world referents, interferes with the processing of non-referring terms and causes a measurable processing difference.[54] This conclusion is incompatible with most special case solutions to the problems of non-reference from the philosophy of language. It is clearly incompatible with Fodor and Dretske's prediction of a substantial structural difference between referring and non-referring concepts, and it is also, I will argue, incompatible with Taylor's suggestion that non-referring concepts simply require an extra processing step.

Table 5-1
Summary of the Results of the Experiments Reported in Chapter 4.

Result	Experiment	Description
1	2	Participants willingly produced graded mean typicality judgments for non-referring categories.
2	3a	Mean typicality judgments were relatively stable across three groups of participants, even when collected using quite different methodologies.
3	1, 2	As with referring categories, typicality judgments for non-referring categories were significantly correlated with associative frequency but not with word frequency.
4	2	Participants were reluctant to accept nonexistent things as members of a normally referring category, and vice versa.
5	2, 3a	By all measures, agreement within and between groups of participants was lower on the typicality rating task for non-referring categories than for referring categories.
6	3	Typicality judgments for non-referring categories correlated significantly with an independently obtained measure of family resemblance within categories. These correlations were comparable in magnitude to the referring categories from this study and previous studies.
7	4	Typicality was predictive of reaction time in speeded categorization to the same extent for both non-referring and referring categories. Participants were slower to categorize less typical items. This effect showed up for positive judgments only.
8	4	Category level was predictive of reaction time in speeded categorization to the same extent for both non-referring and referring categories. Participants were slower to categorize items in superordinate categories. This effect was significant for positive judgments only, though there was a trend towards slower reaction times for superordinate foil categories in the negative judgments.
9	4	Reference type was also predictive of reaction time in speeded categorization. Participants were slower to categorize non-referring items both in positive and negative categorization judgments.
10	4	In negative judgments, there was also a trend towards slower reaction times if the foil category was a non-referring category.
11	4a	The effect of reference type on reaction time was not explainable as an effect of participants' self-reported familiarity with the stimuli.

Structure and Non-Reference

Results 1, 2, 3, 6, 7 and 8 in Table 5-1 strongly suggest that non-referring concepts have a structure very much like their referring counterparts. Participants willingly produced graded mean typicality ratings for non-referring concepts that were relatively stable within and between groups. These typicality ratings correlated significantly with the family resemblance scores derived from a separate property listing task, and both typicality and category level were predictive of reaction time in a speeded categorization task. As I highlighted in Chapter 3, empirical inquiries into the role of similarity in conceptual structure suggest that there are those concepts for which we can find evidence of similarity-based conceptual structures like prototypes (namely object concepts and some abstract concepts) and those for which we cannot find such evidence (namely ad hoc concepts and some abstract concepts). In this context, finding evidence of similarity-based conceptual structures for non-referring object concepts provides a strong argument that, as far as their structure is concerned, they should be placed in the former group, alongside referring object concepts, rather than in the latter, alongside ad hoc concepts.

The typicality and family resemblance effects displayed by non-referring concepts show that they have the same kind of structure as referring concepts. That's the basic lesson about structure. But there are a number of models of conceptual structure that are compatible with this basic lesson. Most obviously, the exemplar and prototype accounts I reviewed in Chapter 3 are explicitly similarity-based, but there are also a number of other models, primarily developed in response to other empirical and conceptual phenomena, that can also account for similarity effects. For example, Prinz's (2002) empiricist project led him to a model of conceptual structure quite similar to prototypes, that he calls

"proxytypes", Sperber and Wilson's (1995) concern with the role of inference in linguistic interpretation led them to a model of conceptual structure in which concepts are "mental addresses" pointing to lists of prototype-like information, and Jackendoff's (1983) concern with the phenomenon of preferred or default semantic interpretations led him to propose a system of "preference rules" that can also account for typicality and family resemblance effects. All of these proposals, while not designed explicitly or primarily to account for typicality and family resemblance effects, can also account for the main results regarding non-referring concepts. But whichever account of similarity in conceptual structure ultimately turns out to be preferable, the basic lesson, that non-referring concepts have a structure very much the same as the structure of referring concepts, still stands.

Processing and Knowledge

The structure story is relatively simple, but the story about processing is less so. Results 9, 10, and 11 in Table 5-1 strongly suggest that non-referring concepts are more difficult to process than referring concepts. Participants take significantly longer to respond in categorization tasks, for both positive and negative judgments, also showing a trend towards even longer reaction times for negative judgments when the foil category is also non-referring. This effect is not explainable as an effect of familiarity with the experimental stimuli. Though I believe this result can and should be investigated further (see the Future Directions section), in this section, I will assume for the sake of argument that it is firmly established.

Given the basic lesson about structure, the source of the processing difference between referring and non-referring concepts could not be any sort of difference in

conceptual structure. My suggestion is that the source of the difference is to be found in the knowledge of the real world that participants bring to the categorization task – knowledge that is external to the concepts themselves. I think that the real world knowledge that there are no smurfs, dragons, and so on causes participants to be a lot more cautious in their judgments about the terms in question. Generally speaking, things that we know to be nonexistent need to be ignored most of the time – at least they need to be prevented from playing a role in the explanations and inferences we use to navigate the real world. Reading a story in which one learns, for example, that a werewolf killed the Prime Minister of Canada, should have no effect on one's beliefs about the real Prime Minister of Canada. So it makes sense that we should be accustomed to thinking carefully about the interpretation of sentences or propositions containing non-referring concepts, and performing extra checks on the validity of those propositions. We know that, as a matter of fact, there are no werewolves, and that therefore the Prime Minister was not killed by one. My suggestion is that this kind of knowledge interferes with the processing of non-referring concepts, even in a task as simple as categorization.

Knowledge and Categorization

In recent years, the study of concepts has begun to take a serious look at the effects of world knowledge on categorization decisions and the role of knowledge-based inference in those judgments. The results of these studies have often been taken to be incompatible with similarity-based views of conceptual structure, and have given rise to "Theory Theory" views of conceptual structure (Gopnik & Meltzoff, 1997; Murphy & Medin, 1985). The basic idea of the Theory Theory is that concepts are not just constituted by lists of properties, as similarity-based theorists would tend to have it, but also by other

kinds of knowledge, such as knowledge about causal mechanisms or explanatory frameworks. This knowledge is encapsulated in mini-domain theories that develop from a naïve state in infants into a more sophisticated form in adults (Carey, 1985; Keil, 1989). In its strongest form, the Theory Theory claims that concepts just *are* the theories themselves. In weaker forms, the Theory Theory claims that concepts are partly constituted by theories, or are individuated by the roles they play in theories.[55] I think that the Theory Theory is misguided and probably wrong, but the results that have driven its development are relevant here.

In Chapter 3, I made a deliberate choice not to explicitly discuss the Theory Theory as a competing theory of concepts, except insofar as I discussed Psychological Essentialism, which is often taken to be a variant or element of the Theory Theory. There were two reasons for this omission. The first was that the discussion didn't fit with the thrust of the story I was telling, which was meant to focus on the most robust results in the literature on concepts for the purpose of identifying experiments that could be used to compare referring and non-referring concepts. The second reason was that the study of mental theories, domain knowledge, and inference is not necessarily the same as the study of the concepts that those theories, knowledge domains, and inferences concern. I think that the Theory Theory is based on a fundamental confusion between what is *constitutive* of a concept, and what is *known about the referent* of a concept. If the standard psychological notion of a concept as a sub-propositional mental representation (as I introduced in Chapter 1) is to remain coherent, then a distinction has to be made between the concept itself and the larger role that it plays in the mental life of an individual. This is not to say that the study of knowledge, folk theories, and inference is

not interesting and useful. Far from it. It's just that the topic in these studies is not necessarily the concepts themselves. I think this reflects a fundamental difference of perspective between at least the advocates of the strongest version of the Theory Theory and the rest of the research community studying concepts – a difference that is not likely to be resolved any time soon.

In any case, the proponents of the Theory Theory have done a good job of isolating and documenting the role of real world knowledge and theoretical preconceptions in categorization tasks. For example, Rips (1989) asked experimental participants whether a circular object, three inches in diameter, was more similar to a pizza or a quarter, and then also asked whether it was more likely to *be* a pizza or a quarter. Participants responded that it was more similar to a quarter, but also more likely to be a pizza, presumably reflecting the knowledge that quarters are much more fixed in size than pizzas. The obvious conclusion is that similarity is not always predictive of categorization decisions. Thus, one could argue, the QUARTER and PIZZA concepts are not just constituted by sets of properties, but also by knowledge about the production of pizzas and quarters and the social rules governing them – that is, by a "theory" of pizzas and quarters.[56]

There are a number of ways in which a prototype or exemplar theorist can respond to a challenge like this. First, like a number of the results that drive the Theory Theory, this one could be explainable by a suitably reconstructed version of prototype or exemplar theory – one in which properties are weighted for categorization. With appropriate property weighting, the *diameter* property of the concept QUARTER could be made very important for categorization, but less so in judging absolute similarity

(Gärdenfors, 2000; Hampton, 2001). Another possible response, which for the current purposes is more interesting, would be to deny that this kind of categorization decision is of the same type as speeded categorization. In speeded categorization tasks, participants tend to go with their first impressions. But in a task with no time pressure, they are free to think through the problem and make inferences, consulting stored knowledge about the world that is external to the concepts most immediately involved in the categorization. Perhaps people can either make quick judgments based on a computation of similarity over the concepts involved, or think through the issue more carefully, consult their real world knowledge about the referents of those concepts, and come up with a more considered decision, which may be different from the one based purely on similarity. Support for this possibility comes from a study by Smith and Sloman (1994) in which Rips' result was replicated when participants talked out loud while performing the task, but not when they performed it silently. Perhaps what we have here are two ways to categorize, reflecting two different ways of thinking.

There is other evidence to support this picture of human categorization behavior. For example, in McCloskey and Glucksberg's (1978) study of fuzzy concepts, they had one group of participants rate items for typicality, while another group judged category membership for those same items. Though McCloskey and Glucksberg didn't mention it, the results, reported in an appendix to their paper, appear to show evidence for a dissociation between categorization and typicality ratings in some cases. For example, 83% of the participants said that whales are not fish, while 80% said that lampreys are. But the evaluations of typicality seem to tell the opposite story: Whales were judged by a separate group of participants to be better examples of fish than lampreys, receiving a

mean typicality rating of 5.71 out of 10, as compared to 5.47 for lampreys. Similarly, while 65% of participants said that a fever is not a disease and 70% said that tooth decay is, fever was again judged to be a much better example of a disease than tooth decay, receiving a mean typicality rating of 6.25 compared to 5.50 for the latter.

Hampton (1998) performed a post hoc statistical analysis on this data and found what he referred to as "plenty of evidence for non-monotonicity" (Hampton, 2001, p. 23) in the relationship between mean typicality and probability of categorization. In a follow-up study, he identified and put to the test a number of other factors that could influence typicality and categorization judgments over and above similarity. Most important for the current discussion was his finding that the degree to which participants rated an item to be either "technically a member" or "technically not a member" of a given category were both significant predictors of residual categorization probability after the effect of typicality had been removed, particularly for biological categories. The interesting thing about the word "technically" is that it seems to imply the existence of some separate, perhaps highly specialized, body of knowledge that would yield a different judgment than more pedestrian ways of thinking about the concept. When typicality and categorization judgments come apart, it may be partly due to the interference of these different kinds of knowledge about the referents of the concepts.[57]

In another study, Armstrong, Gleitman, and Gleitman (1983) had two groups of participants produce typicality ratings for what the authors considered to be "well-defined" categories (ODD NUMBER, EVEN NUMBER, PLANE GEOMETRY FIGURE, and FEMALE). One group performed the task normally, while the other group was first asked, for each category, whether they thought it made sense to rate items for degree of

membership. Almost all the participants in the latter group answered "no" to the question about degree of membership and then went on to produce much more categorical (less graded) mean typicality judgments than the first group. They were much more reluctant to use any part of the typicality scale other than the value indicating "most typical". There are a number of ways to interpret this result,[58] but one possibility is that Armstrong et al. modified the behavior of their participants by calling their attention to their own beliefs about the essences of categories, causing them to pay more attention to, and think more carefully about, category membership, rather than just relying on similarity. By reminding participants of their own theories or domain knowledge, they changed their behavior. This conclusion is again compatible with the view that, by default, participants in these sorts of tasks tend to make use of the similarity-based structures of their concepts, but their judgments can be interfered with by real world knowledge involving those concepts. Again, the behavior of participants in this study is evidence for the existence of two different modes of thinking, both of which are available at all times, but one of which may be preferable in certain tasks or contexts.

Returning to the current work, consider Result 4 in Table 5-1, in which participants rejected non-referring items in the ANIMAL category and referring items in the MONSTER category. This result is also explainable by appealing to the interaction of knowledge and conceptual structure. In Experiment 2, nonexistent animals got the lowest typicality ratings and were rejected as category members by between 30 and 55 per cent of the participants. It might be tempting to suggest that this is simply an effect of similarity. According to prototype theory, the lower the typicality (which tracks similarity), the more likely it is that the exemplar will be rejected as a category member –

a trend documented in McCloskey and Glucksberg's (1978) data. Perhaps the mismatched items are just much less similar to the prototype than the others, due to a heavily weighted *exists* property or dimension. But what is striking about the mixed categories is the apparent discontinuity between typicality level and probability of negative categorization.

The discontinuity becomes clear when we look at the boundary between existent and nonexistent animals in Appendix B. Consider the exemplars that lie on either side of the boundary – "octopus" and "phoenix". The mean rated typicality for these two items was 4.15 and 3.91 respectively, and a paired samples t-test (involving all the participants who gave a numerical rating for both items) did not reach significance: $t(21) = .712$, $p > .40$. On the other hand, there was a marked discontinuity in the rate of categorization between the two items. Only 4% of participants rejected "octopus", while 36% rejected "phoenix" (after question mark responses were removed).[59] The same story can be told for the pairing of "octopus" with "unicorn", the next most typical of the non-referring items in the category. The mean rated typicality for these two items was 4.15 and 3.60 respectively, and a paired samples t-test (involving all the participants who gave a numerical rating for both items) did not reach significance: $t(27) = 1.270$, $p > .20$. But again, there was a large jump in the number of category rejections – 4% of participants rejected "octopus", while 35% rejected "unicorn". The discontinuity was not as sharp in the case of the existent and nonexistent monsters. Again, looking at the exemplars that lie on either side of the boundary ("Cookie Monster" and "alligator"), the mean typicality was 3.34 and 2.41 respectively, but a paired samples t-test over all the participants who gave a numerical rating for both did reach significance in this case: $t(25) = 2.184$, $p < .05$,

though there was still a marked discontinuity in rate of categorization between the two items – 16% of participants rejected "Cookie Monster", while 45% rejected "alligator".

At least in the case of the ANIMAL category, the data seem to suggest that there is more to the rejection of the mismatched items than a simple similarity calculation. In fact, if typicality is a reasonable measure of similarity, then it seems that the explanation for the negative categorizations must be found elsewhere. Once again, I suggest that the effect is explainable as an interference effect involving real world knowledge about the referents of the concepts. The impulse to keep non-referring terms out of everyday inferences causes participants to be reluctant to allow nonexistent objects into otherwise referring categories.[60]

What About Taylor's Solution?

Recall that the survey of the philosophical literature on reference in Chapter 2 did contain one suggestion that predicted the reaction time difference for non-referring terms in the speeded categorization task. According to Taylor's pseudo-Russellian account (Taylor, 2000), sentences containing non-referring concepts require an extra stage of pragmatic processing in which the offending concepts are replaced with descriptive information (a definition or definite description) drawn from the individual's conception of the nonexistent thing in question. In contrast, sentences containing referring concepts do not require this step, as they have referents that can be directly inserted into the proposition constructed from the original utterance. This proposal does correctly predict the longer processing times for non-referring concepts, under the assumption that categorization decisions require the participant to construct and evaluate a proposition (e.g., the proposition that a dragon is a monster).

Nevertheless, there are two reasons why Taylor's solution is not viable. The first is that it is incompatible with the results concerning the structure of non-referring (or referring) concepts. On its own, Taylor's framework does not provide a natural account of the fact that typicality ratings correlate with family resemblance and predict reaction times for positive categorizations in both non-referring and referring categories. The second reason is that Taylor does not have an obvious account of the fact that so many participants rejected the mismatched items from the MONSTER and ANIMAL categories of Experiment 2. Why did so many participants reject imaginary animals as members of the ANIMAL category? If "phoenix" unpacks into something like MYTHICAL BIRD REBORN FROM FIRE, then the question remains of why the concept MYTHICAL BIRD is not a restriction of the concept BIRD, thereby falling under the concept ANIMAL. We would need an account of why it is that this conceptual combination blocks such a natural inference.

A Partly Meinongian Conclusion

My first main question was about the psychological structure of non-referring concepts. The answer to this question, provided by the experimental work described here, is that non-referring concepts have the same structure as referring concepts. Human conceptual structure is Meinongian in the sense that it makes no distinction in kind between referring and non-referring concepts. If formal semantics is in some sense about human cognition, as many researchers feel it ought to be, then special case solutions to the problem of non-referring terms in natural language must be rejected. On the other hand, if formal semantics is not really about human cognition, then special case semantic solutions in linguistic semantics may be admissible. Opinions may vary on the most appropriate

interpretation of this work, but semanticists who also consider themselves to be cognitive scientists should clearly take note of this result.

My second main question was about non-structural differences between referring and non-referring concepts. I found evidence that such differences do exist. Although the human conceptual system has a Meinongian character, people as whole cognitive agents do not. Despite the fact that non-referring concepts have an identical structure to referring concepts, there is a measurable processing difference with respect to our deployment of them, and adults tend to make a strong mental separation between referring and non-referring concepts. I argued above that the best explanation for this phenomenon may well be that knowledge about the world, and about the referents of non-referring concepts, interferes with the normal course of reasoning involving these concepts, and has a measurable impact on processing. This possibility is intriguing, and could form the basis for a number of future studies concerning the role that real world knowledge plays in inference, and how this knowledge develops from childhood to adulthood.

Future Directions

Most of the future work I am proposing concerns the mechanisms by which people can resist being Meinongian even though they have a Meinongian conceptual system. The first two items below are concerned with strengthening and exploring the claim that there is a processing difference between referring and non-referring concepts, and that this difference is due to the interference of world knowledge. The third and fourth items are intended to more deeply investigate the nature and effects of this knowledge in adults and how it develops in children.

1. Strengthen the case for a processing difference. The effect of referential type on speed of categorization seems to be significant and substantial, but there are a couple of reasons the result is not as strong as it could be. The first problem is that the experiment only involved items in two categories – these are currently the only two identified categories containing a sufficiently large number of non-referring exemplars and instances that are well-known to potential participants. But perhaps this problem can be overcome by using a higher level categorization task, in which the task would be to categorize items as "real" or "imaginary". In this situation, monster and superhero instances could be mixed with other imaginary animals (e.g., "unicorn"), fictional characters (e.g., "James Bond"), and even imaginary artifacts (e.g., "time machine", "magic wand"). If participants were still slower to categorize the imaginary objects when the types of objects being categorized were more varied, this would lend weight to the claim that there is an effect of reference type on categorization.

The second weakness of Experiment 4 was that the choice of foil category in the negative judgments was uncontrolled and may have caused some of the problems and disagreements between the various analyses of the negative judgments. A new experiment could be designed to explore the trend noted in negative categorizations, by manipulating the type of the foil category and the type of the item independently so as to study both the effect of item type and the effect of foil type. If the trend towards slower categorization for non-referring items and categories was replicated, this would lend further support to the claim that the effect is due to the referential type of the items involved. Once again, the stimulus set could contain items other than monsters and

superheroes, and since it is well documented that typicality has no effect on negative categorizations, this factor could be ignored in choosing stimuli.

2. Strengthen the case for the interference of real world knowledge. If the slowdown in processing for non-referring items is due to the interference of real world knowledge in inferences concerning non-referring concepts, then perhaps the effect would disappear in a reaction time task which does not force participants to do much processing or inference involving the meanings of the terms involved. A straight lexical decision or naming task could perhaps be used to see if the effect of reference type disappears when participants are only required to make a decision involving the lexical item itself, rather than the nonexistent thing it denotes.

3. Investigate the precise nature of how participants process information about existent and nonexistent things. Moving beyond merely establishing the reliability of the effect of reference type, a new group of experiments could be performed to investigate more precisely what kinds of inferences participants think are and are not licensed concerning non-referring concepts. One set of experiments could involve presenting descriptions of novel animals and artifacts to participants. In Condition A, the items would be presented as new inventions or discoveries, while in Condition B, they would be presented as imaginary or fictional. Participants would be asked a series of questions designed to evaluate how they incorporated the new information into their existing knowledge structures (e.g., how similar are the novel items to familiar items or how likely is it that the novel items will have certain ordinary or unusual properties?). Another group of experiments could investigate how people manage the real/unreal distinction by

looking at the generalization of novel properties from real animals and artifacts to imaginary ones and vice versa (similar to experiments in Carey, 1985).

4. Apply and extend the results to the area of conceptual and representational development. Children as young as three seem to have a very clear sense of what is real and what is not (Woolley, 1997), but they are much less likely than older children to use that distinction in cognitive tasks such as picture sorting (Morison & Gardner, 1978). This suggests that children might lack a mature, fully-developed domain knowledge, or "theory" about the real/unreal distinction, much like they have been shown to lack fully developed theories about biological kinds (Carey, 1985), and to shift from reliance on characteristic properties of objects to more defining properties in making inferences (Keil, 1989). While there is an existing body of work on pretense and imagination in children, the focus has tended to be on children's understanding of the mental states involved in pretense and imagination, rather than on how they store and process information that they know is not about anything existent (Rosengren, Johnson, & Harris, 2000). The goal of this fourth set of studies would be to map out and analyze the developmental trajectory of children's processing of non-referring concepts, using the same types of studies as in the previous sets, adapted for younger participants where necessary.

Final Words

I noted early on that very little work has been done to date concerning the structure of non-referring concepts. To my knowledge, this dissertation marks the first time the suggestion has been made, outside of formal semantics and philosophy of mind, that non-reference, so conceived, might give rise to an interesting kind of concept or knowledge

structure at all.[61] I think I have shown that it does and that my findings are of direct relevance to psychology, philosophy, and linguistics. I have shown, contrary to what might be expected in the philosophical and linguistic camp, that human psychology is partly Meinongian in nature. There is no distinction in kind between referring and non-referring concepts *within* the human conceptual system. This finding should at the very least constrain future theories of reference, and could also be taken as evidence in favor of the internalist camp in the debate over the nature of cognitive science that I gestured towards in Chapter 1.

Despite this lack of distinction in their conceptual systems, both adults and children do manage to keep fantasy and reality separate most of the time, and my experimental work has begun to unearth some evidence about how exactly we do that. But there are many open questions concerning the extent to which referring and non-referring concepts are distinguished in the human mind, and the mechanisms by which this is accomplished, some of which I hope to address in future work. Moreover, beyond the future work items I laid out in the previous section, there lurk some even bigger cognitive, developmental, and even social questions. Adults and children often seem to learn things about the real world from educational and artistic portrayals of characters and events that they know do not depict anything real (children are taught some of what they know through educational cartoons). Under what circumstances do adults and children judge that it is okay to draw an inference about the real world from what is patently a work of fiction? Are there differences between children and adults in this regard? The answers to these questions could, for example, eventually help to shed light on the recurring question of the effect of exposure to television violence on children and

adults. So in addition to addressing a standard problem in formal semantics and opening up a new area of inquiry in cognitive and developmental psychology, the study of non-referring concepts that I have begun in this dissertation may also open up new possibilities for addressing some important social and political questions from a cognitive perspective.

Notes

Chapter 1 Notes

1. Concepts will always be denoted with small caps. Lexical items will be in quotation marks.

2. It could be objected that the optimal psychological form should tend also to be the most ecologically valid one. Obviously, concepts have to have a certain ecological validity or they would not be useful, but I see no reason to make the strong assumption that the psychologically optimal conceptual structure is always a perfect fit to the corresponding category, particularly when scientific investigation often reveals that our folk concepts are mistaken – that whales are not fish is a fact which a surprising number of people seem not to have absorbed (see the discussion in Chapter 5).

Chapter 2 Notes

3. Two recent introductory texts that discuss these issues from slightly different perspectives are Lycan (2000) and Stainton (1996).

4. Some logicians may deny that (5) is true, arguing that a sentence of that form is only true for genuine terms that denote existing objects.

5. Everett tries to explain away our intuitions of truth values by appealing to the Kripkean causal chains that connect our current uses of names to the original naming ceremonies. More on this in a moment.

6. Historically, Meinong's view came first, and was very similar to a view held by Russell for a time. But for my purposes it works better to introduce the two views in the reverse chronological order.

7. The material in this section is drawn in part from Scott (2000).

8. Depending on whose interpretation of Meinong you believe, it may or may not be a consequence of Meinong's view that nonexistent objects have some kind of "nonexistent being". Russell (1905) thought that it was, while Caplan (2002) and Parsons (1980) have argued that it was not, and that Meinong felt the same way. Meinong (1904) actually said that he was proposing a Theory of Objects that would be bigger in scope than metaphysics. He said that metaphysics is only concerned with what exists or is real, but that "the totality of what exists, including what has existed and will exist, is infinitely small in comparison with the totality of the Objects of knowledge" (Meinong, 1904, p. 79). The question is, does the Theory of Objects which Meinong envisaged concern only things (existent or not) that have some kind of being, or does it include Objects that don't exist and don't have any other kind of being? Parsons (1980, p. 10) thinks that this is only a terminological issue, and he may well be right. But to my mind, if nonexistent objects are to be the subject of a "Science of Objects", then they must have some kind of being. As Chisholm puts it, if Meinong is right, then it must be that a nonexistent object "is not a bare nothing… for when we refer to it, we refer to *something*" (Chisholm, 1960, p. 9, author's emphasis).

9. Caplan argues that some nonexistent objects can be referred to, though they don't have any sort of being at all; Deutsch says that some nonexistent objects live in a "fictional

plenitude"; Parsons takes on and develops most of Meinong's commitments, remaining neutral on the question of nonexistent being.

10. Also of note are Crittenden (1991) who thinks nonexistent objects are "grammatical objects", van Inwagen (2000) who argues from quantificational evidence that fictional objects really do exist, and Zalta (2000) who thinks that fictional objects are abstract objects. Lewis (1983) acknowledges that Meinongian solutions can be made to work, though he does not endorse such a solution, and Fine (1982) conducted a detailed investigation into the properties of nonexistent objects, while at the same time denying that there are any.

11. The scare quotes around the words "true" and "refer" are meant to indicate that if we take this internalist step, we are necessarily defining these terms in a different way than most semanticists would. "Reference" no longer requires a referent, and propositions can be judged to be "true" even if we can't find any principled way to account for their literal (metaphysical) truth.

12. This is a common assumption among many researchers looking into pragmatics. For example, Sperber and Wilson (1995, Chapter 2) make just such an assumption.

13. Much of the material in this section is adapted from Scott (2002a, 2002b).

14. For this reason, it is quite unlikely that appealing to these metasemantic, causal-historical chains (like Everett, 2000 does) will help to solve any of the semantic or psychological problems of non-reference. The fact that the names "Santa Claus" and "Father Christmas" share part of the same causal chain of naming may explain how they

can be co-referential cross-culturally, but it does not help us understand how the two names come to be co-referential in one person's head.

15. To be fair, Dretske did develop his theory further (Dretske, 1986, 1988) but none of the modifications contain a satisfactory solution to this problem (see Scott, 2002a for more detailed discussion).

16. Lately, Fodor (1998) prefers to talk about concept acquisition as a process of "locking on" to a relevant property. The new formulation addresses some concerns about nativism and ontology, but Fodor is clear that however locking on works, the meaning of the resulting concept is still grounded in an informational relationship, and Asymmetric Causal Dependence remains his most mature attempt to characterize that relationship.

17. Fodor has been attacked on related grounds before. Baker (1991) constructed a detailed argument based on possible worlds and judgments of which of various possible worlds is "closest" to our own.

18. Fodor also toys with a "mixed" version of Asymmetric Causal Dependence that would make it a condition that at least one actual smurf had caused a SMURF token, but that's pretty clearly a non-starter too.

19. "[The] idea that many terms express concepts that have internal structure is tantamount to the idea that many terms have definitions" (Fodor, 1981, p. 289).

20. According to personal correspondence reported by Prinz (2002), Georges Rey is driven by a defense of a theory of concepts as classical definitions, towards a position that may be quite similar. According to Prinz, Rey thinks that most concept have

external, public definitions, but non-referring concepts correspond to internal, private definitions.

Chapter 3 Notes

21. The one major kind of theory of concepts that I have chosen not to review in this chapter is the one that is usually known as the "Theory Theory". In terms of forming hypotheses about non-referring concepts and motivating the experimental approach I take in Chapter 4, I found little in Theory Theory accounts that was directly relevant. However, I do discuss the Theory Theory and the evidence for it in Chapter 5, where that material plays a role in evaluating the implications of my experimental results.

22. This is more often made as a point about $concepts_A$ than about $concepts_M$ (see Chapter 1), but most of the arguments against the classical view apply to both types of concept.

23. This point has been made in any number of places in the literature on concepts. For example, see summaries by Jackendoff (1983, pp. 112-115), Komatsu (1992), Laurence and Margolis (1999), Medin (1989) or Sperber and Wilson (1995, p. 90).

24. Following the common practice in the literature, I use the word "exemplar" for a subcategory of a category. So DOG is an exemplar of ANIMAL. The word "instance", in contrast, is used for a particular individual. So Fido is an instance of both DOG and ANIMAL.

25. Of interest to the discussion from Chapter 2, Laurence and Margolis (2002) suggest that combining Psychological Essentialism with prototype theory (which will be

reviewed in a moment) can provide a solution to the problem of how new primitive concepts (with appropriate Asymmetric Causal Dependence relations) could be learned. Thus, in their view, Psychological Essentialism helps save a referential, information-based account of concepts, while avoiding radical nativism. But it is worth noting that whether or not the account goes through for the learning of primitive concepts, the problem of non-referring concepts remains in full force (Laurence & Margolis, 2002, fn. 26).

26. For now, the notion of a prototype will be left intuitive, receiving a more detailed treatment at the end of this chapter.

27. Although the assumptions and conclusions of their original studies have come under attack recently (Saunders and van Brakel, 1997), Berlin and Kay assure us that their much more ambitious and thorough World Color Survey will, when complete, support their original conclusions (Kay and Berlin, 1997).

28. Before 1973, Eleanor Rosch wrote under the name Eleanor Rosch Heider.

29. The observant reader might have noticed that the concepts SCIENCE, SPORT, DISEASE, and CRIME are abstract concepts, rather than object concepts. My claim that Rosch only established convincing evidence for prototypes for object concepts rests on the fact that these abstract concepts were not carried forward from the experiments discussed in this section to the critically important family resemblance studies discussed later. These family resemblance studies were the first to look for the *causal determinants* of typicality judgments, rather than to simply establish typicality as a robust and important psychological variable.

30. Experiments on typicality don't actually ask the participants which examples are more "typical", as there is a sense in which "most typical" can mean "most frequent", which is not the variable we are after. What participants are asked to do is more like an evaluation of the "goodness of example" of each item as a member of a given category. I have chosen to use the word "typicality" almost exclusively, as it's a bit more convenient than "goodness of example" and also more common in the literature.

31. Note that the comparison set for rating the semantic relatedness of bird exemplars to the word "bird" was different than the set for relating bird exemplars to the word "animal". The "bird" list included only bird names, while the "animal" list included both bird and mammal names. But I think the point still stands because all the birds were included in both sets, so it was the relative order of the various birds that changed.

32. It should be noted that Rosch (1978) eventually took a much more cautious position regarding this interpretation of her empirical results.

33. Numerous studies have found a "category size" effect in speeded categorization, which became explainable after Rosch et al.'s work as a basic level categorization advantage (Chang, 1986). Some studies have shown an interaction between typicality and level of categorization (Jolicoeur, Gluck, & Kosslyn, 1984; Murphy & Brownell, 1985). For an atypical bird, such as a penguin, participants will show a subordinate level (PENGUIN) rather than basic level (BIRD) naming advantage. A related study by Glass and Meany (1978) showed an interaction between typicality and imagery such that low-typicality, high-imagery exemplars (e.g., penguins and blimps) were categorized at the basic level more slowly than low-typicality, low-imagery exemplars (e.g., grackles and

hovercrafts). One possible reason for these results is that some atypical exemplars, such as penguins and blimps, are initially placed in their own basic level category, to be hooked into a mature conceptual hierarchy after further learning, whereas this does not happen for other, perhaps less familiar, atypical exemplars such as grackles and hovercrafts. A more recent study showed that the basic level categorization advantage over the superordinate level disappears when objects are placed in the context of a scene containing many examples of the superordinate category (Murphy & Wisniewski, 1989). Finally, Tanaka and Taylor (1991) found that domain experts may have a psychologically basic level that corresponds to what would be a subordinate level for non-experts.

34. For instance, Barsalou (1982) showed that context can prime the verification of some properties of concepts but not others, suggesting a split between context-independent and context-dependent properties in the internal representation of concepts.

35. Barsalou made a distinction between ad hoc and goal-directed concepts. According to him, both kinds are directed towards a goal, but ad hoc ones are "those goal derived categories that have been constructed to achieve a novel goal and that therefore are not well-established in memory" (Barsalou, 1985, p. 632). I am going to talk about these two very similar concept types together, only distinguishing them where necessary.

36. Barsalou called them "common taxonomic categories", but of the 9 categories, 8 were object categories and one was abstract (SPORTS).

37 Barsalou and (more recently) Prinz have continued to argue that much of the structure of concepts in working memory is assembled on the fly rather than included in an invariant structure retrieved from long term memory. Both Barsalou's (1999) "simulator"

and Prinz's (2002) "proxytype" concepts are skeletal structures consisting of a small amount of context-invariant information packed together in long term memory. They are default concepts that are retrieved and integrated with other context-dependent information to form a newly modified concept in long term memory. On these accounts, concepts are highly flexible structures, and no two retrievals of the same default concept from long term memory are likely to result in exactly the same concept in working memory. Though weakened a little from Barsalou's original position, these theories still predict much more flexibility and variability in concepts than most other theories.

38. Work by Armstrong et al. (1983) also demonstrate this point for numerical categories. Also, see Lakoff (1987) for related examples.

Chapter 4 Notes

39. Previous studies of this type often go under the alternative title "Category Norms". I use the term "Associative Frequency", following Hampton (1983) because it is not clear that, with 51 participants, the experiment about to be described can properly be considered normative. But 50+ participants was more than enough for the current purposes.

40. Whether extinct animal terms like "dodo" really refer or not is a thorny metaphysical question (see Caplan, 2002, for a discussion), but from a psychological point of view, there seems to be little difference in the referential status of, for example, "dodo" and "dragon".

41. Of course, it's possible that "gods" might turn out to name a category with one or more existing exemplars, but no matter what religion turns out to be right (or not), there will still be plenty of names for gods that fail to refer.

42. The order was: "animals", "superheroes", "dinosaurs", "pets", "monsters", "gods", "fish", "aliens", "imaginary or mythical animals", "pet fish".

43. Initially, the time limit had been set at 30 s, mirroring Battig and Montague (1969), but participants in a pilot study reported that they could have listed many more items if they had had more time, so it was increased to 45 s.

44. By "artificial concepts", I mean those made up by experimenters for use in the lab. By "naturally occurring concepts", I mean those that the participants bring with them to the lab.

45. All superheroes were checked against an on-line dictionary of superheroes to see if the named items had indeed been introduced at some point as superheroes in some well-known narrative, and all unrecognized fish and monster terms were likewise checked with on-line dictionaries and web searches.

46. The ANIMAL category was the only one which contained more than a few items with non-zero frequencies according to Kučera & Francis (1967).

47. Correlations could not always be computed for all possible pairs, as some participants circled the same response (usually "7") for every item. This was most pronounced in the ANIMAL category.

48. There were 49 participants in Experiment 2, but only 45 were required for this experiment.

49. It should also be kept in mind that Groups A and B had around 30 participants each whereas Experiment 2 had 49 participants rating typicality, and also that Group A contained a slightly higher number of non-native English speakers than the other groups (38% as opposed to 25% for Group B and 22% for the Experiment 2 group). These factors could have made the overall pattern of ratings more variable than in the original group from Experiment 2, and thus led to lower correlations between typicality and family resemblance.

50. Controversy surrounding the use of this method of statistical analysis centers around two basic issues discussed in detail by Richter and Seay (1987). One issue involves the degree to which quasi-F's accurately approximate actual F sampling distributions, and the other issue involves the ramifications of assuming that the items were selected at random from the population of potential items within any condition when typically some systematic basis of item selection is used.

51. Note that the appropriate degrees of freedom associated with each quasi-F statistic must be estimated using the method prescribed in Kirk (1995).

52. Note that the Group B participants are the same as those who produced the familiarity ratings for this experiment.

53. This method of analysis provided another opportunity to check for an influence of factors such as word length and frequency on reaction time (see the discussion section of

Experiment 4). Repeating the analysis with typicality and reference type as the main predictors, but including word frequency and word length in the regression as covariates, did not change the overall pattern of statistical results for either positive or negative categorizations. This is further evidence that the effect of reference cannot be explained by other factors.

Chapter 5 Notes

54. The phrasing here is crude, of course, since ordinary folk don't "know" what non-referring concepts are, or what reference is, or even what concepts are in the technical sense intended here.

55. In semantics, the Theory Theory finds a close cousin in conceptual or functional role theories (e.g., Block, 1986, 1999; Goldstone & Rogosky, 2002; Harman, 1982).

56. It should be noted that Rips (1989) is more cautious in his conclusions than this. He argues that "because resemblance theory is precisely the claim that categorization can be reduced to similarity alone, resemblance theory must be false" (Rips, 1989, p. 36). That is, similarity does not explain all categorization behavior, which is exactly the line I am pushing for here. However, this experiment is often cited as evidence that motivates the Theory Theory (e.g., Prinz, 2002), and Rips did go on to eventually endorse a sort of hybrid version of the Theory Theory, in which a concept is individuated both by its mental symbol and by its role in a mental theory (Rips, 1995).

57. Hampton (1998, 2001) seems a little reluctant to draw this sort of conclusion from his data, though he does concede in a number of places that similarity-based categories

cannot account for every type of human categorization behavior: “It would probably be foolish to argue that all our categories are constructed on the basis of putting similar things together” (Hampton, 2001, p. 24).

58. Armstrong et al. (1983) famously argued that this result and others reported in the same paper were evidence against prototype, or “family resemblance” theory. They were right that typicality rating should not be taken as proof of prototype or similarity-based structure on its own (as I argued in Chapter 3), but they were wrong to think that this finding necessarily invalidates prototype theory. (See Laurence and Margolis, 1999, for further discussion.)

59. These results should be taken as rough illustrations only. The experiment was not quite the right design for this sort of comparison, since the typicality ratings involved in the t-test only came from the group that accepted both items as a member of the target category.

60. I pointed out in the discussion of the Experiment 2 results that there did not appear to be a group of “realist” participants who were consistent across categories (i.e., always rejecting the mismatched items in both categories). This observation is also compatible with the analysis I am developing here. Two modes of thinking, call them “similarity-based” and “inferential” for lack of better terminology, are available to all participants, and the impulse to resist the mixing of the existent and the nonexistent reflects the interference of the inferential mode, based on knowledge of the real world, with the similarity-based mode, based on the structure of the concepts involved.

61. Medin, Lynch, and Solomon (2000) don’t mention the possibility at all in their list of “kinds of concepts”.

References

> The first time I read the dictionary, I thought it was a poem about everything.
>
> – Steven Wright

Ahn, Woo-kyoung, Kim, Nancy S., Lassaline, Mary E., & Dennis, Martin J. (2000). Causal status as a determinant of feature centrality. *Cognitive Psychology*, 41, 361-416.

Allen, James. (1995). *Natural Language Understanding, Second Edition.* California: Benjamin/Cummings Publishing Company.

Anderson, John R. & Jonathan Betz. (2001). A hybrid model of categorization. *Psychonomic Bulletin and Review*, 8(4), 629-647.

Armstrong, Sharon Lee, Gleitman, Lila R., & Gleitman Henry. (1983). What some concepts might not be. *Cognition*, 13, 263-308.

Ashcraft, Mark H. (1978a). Property dominance and typicality effects in property statement verification. *Journal of Verbal Learning and Verbal Behavior*, 17, 155-164.

Ashcraft, Mark H. (1978b). Property norms for typical and atypical items from 17 categories: A description and discussion. *Memory and Cognition*, 6, 227-232.

Baker, Lynne Rudder. (1991). Has content been naturalized? In Barry Loewer & Georges Rey (Eds.) *Meaning in Mind: Fodor and his Critics* (pp. 17-32). Oxford, England: Blackwell.

Barsalou, Lawrence W. (1982). Context-independent and context-dependent information in concepts. *Memory and Cognition*, 10(1), 82-93.

Barsalou, Lawrence W. (1983). Ad hoc categories. *Memory and Cognition*, 11(3), 211-227.

Barsalou, Lawrence W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11(4), 629-654.

Barsalou, Lawrence W. (1987). The instability of graded structure: Implications for the nature of concepts. In Ulric Neisser (Ed.) *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization*. Cambridge, England: Cambridge University Press.

Barsalou, Lawrence W. (1989). Intra-concept similarity and its implications for inter-concept similarity. In Stella Vosniadou & Andrew Ortony (Eds.) *Similarity and Analogical Reasoning*. Cambridge, England: Cambridge University Press.

Barsalou, Lawrence W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences,* 22, 577-660.

Barsalou, Lawrence W. & Medin, Douglas L. (1986). Concepts: Static definitions or context-dependent representations? *Cahiers de Psychologie Cognitive,* 6(2), 187-202.

Battig, William F., & Montague, William E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monograph Supplements*, 80(3), Part 2.

Bellezza, Francis S. (1984a). Reliability of retrieval from semantic memory: Common categories. *Bulletin of the Psychonomic Society*, 22(4), 324-326.

Bellezza, Francis S. (1984b). Reliability of retrieval from semantic memory: Information about people. *Bulletin of the Psychonomic Society*, 22(6), 511-513.

Bellezza, Francis S. (1984c). Reliability of retrieval from semantic memory: Noun meanings. *Bulletin of the Psychonomic Society*, 22(5), 377-380.

Berlin, Brent, & Kay, Paul. (1969). *Basic Color Terms: their Universality and Evolution*. Berkeley: University of California Press.

Block, Ned. (1986). Advertisement for a semantics for psychology. *Midwest Studies in Philosophy*, X. Reprinted in Stephen P. Stich and Ted A. Warfield. (1994). *Mental Representation: A Reader* (pp. 81-141). Oxford, UK: Blackwell.

Block, Ned. (1999). Functional role semantics. In Robert A. Wilson & Frank C. Keil (Eds.). *MIT Encyclopedia of the Cognitive Sciences* (pp. 331-332). Cambridge, MA: MIT Press.

Bloom, Paul. (2000). *How Children Learn the Meanings of Words*. Cambridge, MA: MIT Press.

Caplan, Ben. (2002). *Empty Names: An Essay on the Semantics, Pragmatics, Metaphysics, and Epistemology of Empty Names and Other Directly Referential Expressions*. Ph.D. Dissertation. University of California, Los Angeles.

Carey, Susan. (1985). *Conceptual Change in Childhood*. Cambridge, MA: MIT Press.

Chang, Tien Ming. (1986). Semantic memory: Facts and models. *Psychological Bulletin*, 99(2), 199-220.

Cherniak, Christopher. (1984). Prototypicality and deductive reasoning. *Journal of Verbal Learning and Verbal Behavior*, 23, 625-642.

Chisholm, Roderick M. (1960). Editor's introduction. In *Realism and the Background of Phenomenology* (pp. 3-38). New York: The Free Press.

Chomsky, Noam. (2000). *New Horizons in the Study of Mind and Language*. Cambridge, England: Cambridge University Press.

Clark, Herbert H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 12, 335-359.

Crittenden, Charles. (1991). *Unreality: The Metaphysics of Physical Objects*. Ithaca, NY: Cornell University Press.

Deutsch, Harry. (2000). Making up stories. In Anthony Everett & Thomas Hofweber (Eds.) *Empty Names, Fiction and the Puzzles of Non-existence* (pp. 149-182). Stanford, CA: CSLI Publications.

Dretske, Fred. (1981). *Knowledge and the Flow of Information*. Cambridge, MA.: MIT Press.

Dretske, Fred. (1986). Misrepresentation. In R. Bogden (Ed.) *Belief, Form, Content and Function* (pp. 17-36). Oxford, England: Oxford University Press.

Dretske, Fred. (1988). *Explaining Behavior*. Cambridge, MA: MIT Press.

Everett, Anthony. (2000). Referentialism and empty names. In Anthony Everett & Thomas Hofweber (Eds.) *Empty Names, Fiction and the Puzzles of Non-existence* (pp. 37-60). Stanford, CA: CSLI Publications.

Everett, Anthony & Hofweber, Thomas (Eds.) (2000). *Empty Names, Fiction and the Puzzles of Non-existence*. Stanford, CA: CSLI Publications.

Faderman, Avrom. (2000). On myth. In Anthony Everett & Thomas Hofweber (Eds.) *Empty Names, Fiction and the Puzzles of Non-existence* (pp. 61-66). Stanford, CA: CSLI Publications.

Fine, Kit. (1982). The problem of non-existents I: Internalism. *Topoi*, 1, 97-140.

Fodor, Janet D., Fodor, Jerry A. & Garrett, Merrill F. (1975). The psychological unreality of semantic representations. *Linguistic Inquiry*, 4, 515-531.

Fodor, Jerry A. (1970). Three reasons for not deriving "kill" from "cause to die". *Linguistic Inquiry*, 1(4), 429-438.

Fodor, Jerry A. (1975). *The Language of Thought*. New York: Crowell.

Fodor, Jerry A. (1981). The present status of the innateness controversy. In *Representations: Philosophical Essays on the Foundations of Cognitive Science* (pp. 257-317). Cambridge, MA: MIT Press.

Fodor, Jerry A. (1987). *Psychosemantics*. Cambridge, MA.: MIT Press.

Fodor, Jerry A. (1990). A theory of content II: The theory. In *A Theory of Content and Other Essays* (pp. 89-136). Cambridge, MA.: The MIT Press.

Fodor, Jerry A. (1994). *The Elm and the Expert*. Cambridge, MA: MIT Press.

Fodor, Jerry A. (1998). *Concepts: Where Cognitive Science Went Wrong*. Oxford, England: Clarendon Press.

Fodor, Jerry A., Garrett, Merrill F., Walker, E. & Parkes, C. (1980). Against definitions. *Cognition*, 8, 263-367.

Fodor, Jerry A. & Lepore, Ernie. (1996). The red herring and the pet fish: Why concepts still can't be prototypes. *Cognition*, 58, 253-270.

Frege, Gottlob. (1892). Comments on sinn and bedeutung. Reprinted in Michael Beaney (Ed.) (1997) *The Frege Reader*. Oxford, England: Blackwell.

Gärdenfors, Peter. (2000). *Conceptual Spaces: The Geometry of Thought*. Cambridge, MA: The MIT Press.

Glass, Arnold L., & Meany, Peter J. (1978). Evidence for two kinds of low-typical instances in a categorization task. *Memory and Cognition*, 6(6), 622-628.

Goldstone, Robert L. (1994). The role of similarity in categorization: providing a groundwork. *Cognition*, 52, 125-157.

Goldstone, Robert L., & Rogosky, Brian J. (2002). Using relations within conceptual systems to translate across conceptual systems. *Cognition*, 84, 295-320.

Gopnik, Alison & Meltzoff, A. (1997). *Words, Thoughts, and Theories*. Cambridge, MA: MIT Press.

Grice, H. Paul. (1975). Logic and conversation. In Peter Cole & Jerry L. Morgan (Eds.) *Syntax and Semantics, Vol. 3* (pp. 41-58). New York: Academic Press.

Guildford, J. P., & Fruchter, B. (1973). *Fundamental Statistics in Psychology and Education*. New York: McGraw Hill.

Hampton, James A. (1979). Polymorphous concepts in semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 18, 441-461.

Hampton, James A. (1981). An investigation of the nature of abstract concepts. *Memory and Cognition*, 9(2), 149-156.

Hampton, James A. (1983). Measures of internal category structure: A correlational analysis of normative data. *British Journal of Psychology*, 74, 491-516.

Hampton, James A. (1993). Prototype models of concept representation. In Iven Van Mechelen, James Hampton, Ryszard S. Michalski & Peter Theuns (Ed.) *Categories and Concepts: Theoretical Views and Inductive Data Analysis* (pp. 67-96). London, England: Academic Press.

Hampton, James A. (1995). Testing the prototype theory of concepts. *Journal of Memory and Language*, 34, 686-708.

Hampton, James A. (1998). Similarity-based categorization and the fuzziness of natural categories. *Cognition*, 65, 137-165.

Hampton, James A. (2000). Concepts and prototypes. [Review of *Concepts: Where Cognitive Science Went Wrong*]. *Mind and Language*, 15(2&3), 299-307.

Hampton, James A. (2001). The Role of Similarity in Natural Categories. In Ulrike Hahn & Michael Ramscar (Eds.) *Similarity and Categorization* (pp. 13-28). Oxford, England: Oxford University Press.

Hampton, James A. & Gardiner, Margaret M. (1983). Measures of internal category structure: A correlational analysis of normative data. *British Journal of Psychology*, 74, 491-516.

Harman, Gilbert. (1982). Conceptual role semantics. *Notre Dame Journal of Formal Logic*, 23(2), 242-256

Hays, William L. (1988). *Statistics (4th ed.).* Fort Worth, TX: Harcourt Brace Jovanovich College Publishers.

Heider, Eleanor Rosch. (1971). "Focal" color areas and the development of color names. *Developmental Psychology*, 4, 447-455.

Heider, Eleanor Rosch. (1972). Universals in color naming and memory. *Journal of Experimental Psychology*, 93, 10-20.

Hunt, Karl P. & Milton H. Hodge. (1971). Category-item frequency and category-name meaningfulness (m'): Taxonomic norms for 84 categories. *Psychonomic Monograph Supplements*, 4(6), whole number 54.

Jackendoff, Ray. (1983). *Semantics and Cognition*. Cambridge, MA: MIT Press.

Jackendoff, Ray. (1990). *Semantic Structures*. Cambridge, MA: MIT Press.

Jackendoff, Ray. (1992). *Languages of the Mind*. Cambridge, MA: MIT Press.

Jackendoff, Ray. (in press). Precis of: *Foundations of Language: Brain, Meaning, Grammar, Evolution*. To appear in *Behavioral and Brain Sciences*.

Janczura, Gerson A., & Nelson, Douglas L. (1999). Concept accessibility as the determinant of typicality effects. *American Journal of Psychology*, 112(1), 1-19.

Jolicoeur, Pierre, Gluck, Mark A., & Kosslyn Stephen M. (1984). Pictures and names: Making the connection. *Cognitive Psychology*, 16, 243-275.

Kahneman, D., & Tversky, Amos. (1973). On the psychology of prediction. *Psychological Review*, 80, 237-251.

Katz, Jerrold J. (1966). *The Philosophy of Language*. New York: Harper & Row.

Kay, Paul, & Berlin, Brent. (1997). Science ≠ imperialism: There are nontrivial constraints on color naming. *Behavioral and Brain Sciences*, 20, 196-201.

Keil, Frank C. (1989). *Concepts, Kinds and Cognitive Development*. Cambridge, MA: MIT Press.

Keil, Frank C., & Wilson, Robert A. (2000). The concept concept: The wayward path of cognitive science. [Review of *Concepts: Where Cognitive Science Went Wrong*].. *Mind and Language,* 15(2&3), 309-317.

Kelly, Michael H., Bock, J. Kathryn, & Keil, Frank C. (1986). Prototypicality in a linguistic context: Effects on sentence structure. *Journal of Memory and Language*, 25, 59-74.

Kintsch, Walter. (1974). *The Representation of Meaning in Memory*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences (3rd ed.).* Pacific Grove, CA: Brooks/Cole Publishing Company.

Komatsu, Lloyd K. (1992). Recent views of conceptual structure. *Psychological Bulletin*, 112(3), 500-526.

Kripke, Saul A. (1980). *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Kučera, Henry & Francis, W. Nelson. (1967). *Computational Analysis of Present-Day American English*. Providence, RI: Brown University Press.

Lakoff, George. (1987). *Women, Fire and Dangerous Things*. Chicago: University of Chicago Press.

Laurence, Stephen, & Margolis, Eric. (1999). Concepts and cognitive science. In Eric Margolis & Stephen Laurence (Eds.) *Concepts: Core Readings* (pp. 3 – 81.) Cambridge, MA: MIT Press.

Laurence, Stephen, & Margolis, Eric. (2002). Radical concept nativism. *Cognition*, 86, 25-55.

Lewis, David. (1983). Truth in fiction. In *Philosophical Papers, Volume 1* (pp. 261-280). Oxford: Oxford University Press.

Loess, Henry, Brown, Alan, & Campbell, John. (1969). Cultural norms for items in 30 taxonomic categories. *Psychonomic Monograph Supplements*, 3(7), whole number 39.

Lorch Jr., Robert F., & Myers, Jerome L. (1990). Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(1), 149-157.

Lycan, William G. (2000). *Philosophy of Language: A Contemporary Introduction*. New York: Routledge.

McCloskey, Michael E. (1980). The stimulus familiarity problem in semantic memory research. *Journal of Verbal Learning and Verbal Behavior*, 19, 485-502.

McCloskey, Michael E. & Glucksberg, Sam. (1978). Natural categories: Well defined or fuzzy sets? *Memory and Cognition*, 6(4), 462-472.

McEvoy, Cathy L., & Nelson, Douglas L. (1982). Category name and instance norms for 106 categories of various sizes. *American Journal of Psychology*, 95(4), 581-634.

Malt, Barbara C. (1994). Water is not H_2O. *Cognitive Psychology*, 27, 41-70.

Malt, Barbara C., & Smith, Edward E. (1982). The role of familiarity in determining typicality. *Memory and Cognition*, 10(1), 69-75.

Malt, Barbara C., & Smith, Edward E. (1984). Correlated properties in natural categories. *Journal of Verbal Learning and Verbal Behavior*, 23, 250-269.

Margolis, Eric. (1998). How to acquire a concept. *Mind and Language*, 13, 347-369.

Medin, Douglas L. (1989). Concepts and conceptual structure. *American Psychologist*, 44(12), 1469-1481.

Medin, Douglas L., Lynch, Elizabeth B., & Solomon, Karen O. (2000). Are there kinds of concepts? *Annual Review of Psychology*, 51, 121-147.

Medin, Douglas L., & Ortony, Andrew. (1989). Psychological essentialism. In Stella Vosniadou, & Andrew Ortony (Eds.) *Similarity and Analogical Reasoning* (pp. 179-195). Cambridge, England: Cambridge University Press.

Medin, Douglas L., & Smith, Edward E. (1984). Concepts and concept formation. *Annual Review of Psychology*, 35, 113-138.

Medin, Douglas L., Wattenmaker, William D., & Hampson, Sarah E. (1987). Family resemblance, conceptual cohesiveness, and category construction. *Cognitive Psychology*, 19, 242-279.

Meinong, Alexius. (1904). The theory of objects. In Roderick M. Chisholm (Ed.) (1960). *Realism and the Background of Phenomenology* (pp. 76-117). New York: The Free Press.

Mervis, Carolyn B., Caitlin, Jack, & Rosch, Eleanor. (1976). Relationships among goodness-of-example, category norms, and word frequency. *Bulletin of the Psychonomic Society*, 7(3), 283-284.

Michalski, Ryszard S. (1993). Beyond prototypes and frames: The two-tiered concept representation. In Iven Van Mechelen, James Hampton, Ryszard S. Michalski & Peter Theuns. *Categories and Concepts: Theoretical Views and Inductive Data Analysis* (pp. 145-172). New York: Academic Press.

Miller, George A., & Johnson-Laird, Phillip. N. (1976). *Language and Perception*. Cambridge, MA: Harvard University Press.

Millikan, Ruth Garrett. (1984). *Language, Thought and other Biological Categories*. Cambridge, MA.: MIT Press.

Millikan, Ruth Garrett. (1989). Biosemantics. *Journal of Philosophy*, 86(6), 281-297.

Millikan, Ruth Garrett. (1990). Compare and contrast Dretske, Fodor, and Millikan on teleosemantics. *Philosophical Topics*, 18(2), 151-161.

Millikan, Ruth Garrett. (1998). A common structure for concepts of individuals, stuffs, and real kinds: More mama, more milk, and more mouse. *Behavioral and Brain Sciences*, 21, 55-65.

Morison, P., & Gardner, H. (1978). Dragons and dinosaurs: The child's capacity to differentiate fantasy from reality. *Child Development*, 49, 642-648.

Murphy, Gregory L. (2002). *The Big Book of Concepts*. Cambridge, MA: MIT Press.

Murphy, Gregory L., & Brownell, Hiram H.. (1985). Category differentiation in object recognition: Typicality constraints on the basic category advantage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11(1), 70-84.

Murphy, Gregory L., & Medin, Douglas L. (1985) The role of theories in conceptual coherence. *Psychological Review*, 92(3), 289-316.

Murphy, Gregory L., & Wisniewski, Edward J.. (1989). Categorizing objects in isolation and in scenes: What a superordinate level is good for. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(4), 572-586.

Neumann, Paul G. (1974). An attribute frequency model for the abstraction of prototypes. *Memory and Cognition*, 2(2), 241-248.

Nosofsky, Robert M. (1988). Exemplar-based Accounts of Relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 4, 700-708.

Nosofsky, Robert M. (1992). Exemplars, prototypes and similarity rules. In Alice F. Healy, Stephen M. Kosslyn, & Richard M. Shiffrin (Eds.) *Essays in Honor of William K. Estes, Volume 1* (pp. 149-168). Hillsdale, NJ: Lawrence Erlbaum Associates.

Osherson, Daniel N., & Smith, Edward E. (1981). On the adequacy of prototype theory as a theory of concepts. *Cognition*, 9, 35-58.

Parsons, Terrance. (1980). *Nonexistent Objects*. New Haven, CT: Yale University Press.

Pedhazur, Elazar J. (1997). *Multiple Regression in Behavioral Research: Explanation and Prediction, Third Edition*. Fort Worth, TX: Harcourt Brace College Publishers.

Pothos, Emmanuel M., & Hahn, Ulrike. (2000). So concepts aren't definitions, but do they have necessary *or* sufficient conditions? *British Journal of Psychology*, 91, 439-450.

Prinz, Jesse J. (2002). *Furnishing the Mind: Concepts and their Conceptual Basis*. Cambridge, MA: MIT Press.

Putnam, Hilary. (1975). The meaning of 'meaning'. *Language, Mind, and Knowledge*. Minnesota Studies in the Philosophy of Science, VII, 131-193.

Quine, Willard Van Orman. 1953. Two dogmas of empiricism. In *From a Logical Point of View* (pp. 20-46). Cambridge, MA: Harvard University Press.

Recanati, François. (1993). *Direct Reference*. Oxford, England: Blackwell.

Reed, Stephen K. (1972). Pattern recognition and categorization. *Cognitive Psychology*, 3, 382-407.

Rey, Georges. (1983). Concepts and stereotypes. [Review of *Categories and Concepts*]. *Cognition*, 15, 237-262.

Richter, Martin L., & Seay, Mary B. (1987). ANOVA designs with subjects and stimuli as random effects: Applications to prototype effects on recognition memory. *Journal of Personality and Social Psychology*, 53, 470-480.

Rips, Lance J. (1975). Inductive Judgments about natural categories. *Journal of Verbal Learning and Verbal Behavior*, 14, 665-681.

Rips, Lance J. (1989). Similarity, typicality, and categorization. In Stella Vosniadou & Andrew Ortony. *Similarity and Analogical Reasoning* (pp. 21-59). Cambridge, England: Cambridge University Press.

Rips, Lance J. (1995). The current status of research on concept combination. *Mind and Language*, 10, 72-104.

Rips, Lance J., Shoben, Edward J., & Smith, Edward. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior*, 12, 1-20.

Rosch, Eleanor H. (1973a). Natural categories. *Cognitive Psychology*, 4, 328-350.

Rosch, Eleanor H. (1973b). On the internal structure of perceptual and semantic categories. In Timothy E. Moore (Ed.) *Cognitive Development and the Acquisition of Language* (pp. 111-144). New York: Academic Press.

Rosch, Eleanor. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, 104(3), 192-233.

Rosch, Eleanor. (1978). Principles of Categorization. In Eleanor Rosch & Barbara B. Lloyd (Ed.) *Cognition and Categorization* (pp. 27-48). Hillsdale, NJ: Lawrence Erlbaum Associates.

Rosch, Eleanor, & Mervis, Carolyn B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7, 573-605.

Rosch, Eleanor, Mervis, Carolyn B., Gray, Wayne D., Johnson, David M., & Boyes-Braem, Penny. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8, 382-439.

Rosch, Eleanor, Simpson, Carol, & Miller, R. Scott. (1976). Structural bases of typicality effects. *Journal of Experimental Psychology: Human Perception and Performance*, 2(4), 491-502.

Rosengren, Karl S., Johnson, Carl N., & Harris, Paul L. (2000). *Imagining the Impossible: Magical, Scientific, and Religious Thinking in Children*. New York: Cambridge University Press.

Roth, Emilie M., & Shoben, Edward J. (1983). The effect of context on the structure of categories. *Cognitive Psychology*, 15, 346-378.

Russell, Bertrand. (1905). On denoting. Reprinted in Robert Charles Marsh (Ed.) (1956). *Logic and Knowledge: Essays, 1901-1950* (pp. 41-56). London, England: George Allen & Unwin.

Russell, Bertrand. (1919). *Introduction to Mathematical Philosophy*. London, England: George Allen & Unwin.

Saunders, B. A. C., & van Brakel, J. (1997). Are there nontrivial constraints on colour categorization? *Behavioral and Brain Sciences*, 20, 167-228.

Schank, Roger. (1973). Identification and conceptualizations underlying natural language. In Roger Schank & Kenneth Mark Colby (Eds.) *Computer Models of Thought and Language* (pp. 187-248). San Francisco, CA: Freeman.

Schiffer, Stephen. (2000). Pleonastic Fregeanism and empty names. In Anthony Everett & Thomas Hofweber (Eds.) *Empty Names, Fiction and the Puzzles of Non-existence* (pp. 3-16). Stanford, CA: CSLI Publications.

Scott, Sam. (2000). On Building a Denotation-Based Semantics of Fiction. *Carleton University Student Journal of Philosophy*, 19(1). http://rideau.carleton.ca/philosophy/cusjp

Scott, Sam. (2002a). The Psychological Implausibility of Naturalized Content. In Wayne D. Gray & Christian D. Schunn (Eds.) *Proceedings of the Twenty-fourth Annual Conference of the Cognitive Science Society*, (pp. 822-827). Mahwah, NJ: Lawrence Erlbaum Associates.

Scott, Sam. (2002b). *Saving Conceptual Atomism from Itself.* Paper presented at the 28th meeting of the Society for Philosophy and Psychology, Edmonton, Alberta, Canada.

Shapiro, S. I., & Palermo, David S. (1970). Conceptual organization and class membership: Normative data for representatives of 100 categories. *Psychonomic Monograph Supplements*, 3(11), whole number 43.

Smith, Edward E., Balzano, G. J., & Walker, J. H. (1978). Nominal and semantic processes in picture categorization. In J. Cotton and R. Klatzky (Eds.) *Semantic Factors in Cognition* (pp. 137-168). Potomac, MD: Lawrence Erlbaum Associates.

Smith, Edward E., & Medin, Douglas L. (1981). *Categories and Concepts*. Cambridge, MA: Harvard University Press.

Smith, Edward E., Patalano, Andrea L., & Jonides, John. (1998). Alternative strategies of categorization. *Cognition*, 65, 167-196.

Smith, Edward E, Shoben, Edward J. and Rips, Lance J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, 81, 214-241.

Smith, Edward E., & Sloman, Steven A. (1994). Similarity- versus rule-based categorization. *Memory and Cognition*, 22(4), 377-386.

Sperber, Dan, & Wilson, Deirdre. (1995). *Relevance: Communication and Cognition, Second Edition*. Cambridge, MA: Blackwell.

Stainton, Robert J. (1996). *Philosophical Perspectives on Language*. Peterborough, ON: Broadview Press.

Stanley, Jason. (2000). Context and logical form. *Linguistics and Philosophy*, 23, 391-434.

Tanaka, James W., & Taylor, Marjorie. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology*, 23, 457-482.

Taylor, Kenneth. (2000). Emptiness without compromise: A referentialist semantics for empty names. In Anthony Everett & Thomas Hofweber (Eds.) *Empty Names, Fiction and the Puzzles of Non-existence* (pp. 17-36). Stanford, CA: CSLI Publications.

Terry, Pamela, Samuels, S. Jay, & LaBerge, David. (1976). The effects of letter degradation and letter spacing on word recognition. *Journal of Verbal Learning and Verbal Behavior*, 15, 577-585.

Usher, Marius. (2001). A statistical referential theory of content: Using information theory to account for misrepresentation. *Mind and Language*, 16(3), 311-334.

van Inwagen, Peter. (2000). Quantification and fictional discourse. In Anthony Everett & Thomas Hofweber (Eds.) *Empty Names, Fiction and the Puzzles of Non-existence* (pp. 235-248). Stanford, CA: CSLI Publications.

Whaley, C. P. (1978). Word-nonword classification time. *Journal of Verbal Learning and Verbal Behavior*, 17, 143-154.

Wilks, Yorick A., Slator, Brian M., & Guthrie, Louise M. (1996). *Electric Words: Dictionaries, Computers and Meanings*. Cambridge, MA: MIT Press.

Wittgenstein, Ludwig. (1953). *Philosophical Investigations, Third Edition*. Anscombe (Tr.) Oxford: Blackwell.

Woolley, J. D. (1997). Thinking about fantasy: Are children fundamentally different thinkers and believers from adults? *Child Development*, 68(6), 991-1011.

Zalta, Edward N. (2000). Pretense theory and abstract object theory. In Anthony Everett & Thomas Hofweber (Eds.) *Empty Names, Fiction and the Puzzles of Non-existence* (pp. 117-148). Stanford, CA: CSLI Publications.

Appendices

Reality is a crutch for people who can't handle drugs.

– Lily Tomlin

A: Associative Frequency

The following tables show the entire set of items reported by the participants in Experiment 1 for all 10 categories. For each category, items with an associative frequency (AF) of 3 or more are presented in a table, while items with an AF of 2 or 1 are presented in list form after the table. The table columns are as follows: "Item" = item name (spelling corrected), "AF" = associative frequency (max 50), and "#1" = number of participants listing the item first (max 50). The tables are sorted by column "AF", then by "#1", and finally by "item". The low-AF lists are alphabetized with the "#1" data shown in brackets after each item that was listed first by at least one participant.

Aliens

Item	AF	#1	Item	AF	#1
E.T.	33	25	U.F.O.	4	0
Alf	14	6	mars	3	1
Martian	11	5	the movies ("Alien", etc.)	3	1
3rd Rock from the Sun characters	6	0	Marvin the Martian	3	0
green	4	0	Spock	3	0

AF = 2: big eyes, green men(1), illegal alien, Klingon(1), little green man(1), me, outer space, Star Trek, X-files.

AF = 1: alien, alien citizens, aliens from the Simpsons(1), anyone from "X-Files", Astro, big heads and eyes, body jumper, Casey, Dick ("3rd Rock from the Sun"), Ferengi, Fox, Grey 5(1), Harry ("3rd Rock from the Sun"), illegal immigrant, John Lithgow, K-Pax, little blue man, mac, man from mars, man on the moon, Men in Black, Mork, mystery, one eye, pasty white skin, power rangers(1), Predator, professor(1), Romulan, Roswell, shape shifter(1), small, small body, small mouth, Solomon ("3rd Rock from the Sun"),

space, space ship, Star Wars, Star Wars creature, stars, stranger, Superman, teacher, the ones from “The Abyss”, Tom (“3rd Rock from the Sun”), unknown, Zeus.

Animals

Item	AF	#1	Item	AF	#1
dog	49	28	rabbit	6	1
cat	49	8	beaver	6	0
bird	30	4	lizard	6	0
cow	27	3	alligator	5	0
fish	26	0	crocodile	5	0
horse	23	0	duck	5	0
elephant	20	0	fox	5	0
bear	19	2	turtle	5	0
lion	19	0	dolphin	4	0
mouse	18	0	hippopotamus	4	0
pig	17	0	shark	4	0
tiger	16	0	squirrel	4	0
deer	14	0	canary	3	1
moose	14	0	bee	3	0
monkey	13	1	chimp	3	0
chicken	11	1	coyote	3	0
snake	11	0	gerbil	3	0
giraffe	10	0	goat	3	0
zebra	10	0	goose	3	0
hamster	9	0	iguana	3	0
rat	9	0	kangaroo	3	0
whale	9	0	lamb	3	0
frog	8	0	panda	3	0
sheep	8	0	raccoon	3	0
wolf	8	0	reptile	3	0
donkey	7	0	turkey	3	0
gorilla	7	0			

AF = 2: buffalo, bunny, camel, caribou, cheetah, crow, dinosaur, eagle, ferret, flamingo, groundhog, guinea pig, leopard, parrot, penguin(1), rhinoceros, skunk, tuna, worm.

AF = 1: albatross, amphibian, ant, anteater, armadillo, bass, bili, bobcat, bonobo, bug, bull, calf, chipmunk, cod, elk, finch, gazelle, golden retriever, goldfish, gopher, hawk, hen, human, jaguar, k-9, mammal, mosquito, nut, octopus, opossum, ostrich, ox, pigeon, polar bear, rodent, rooster, salmon, seahorse, seal, serpent, siomeas, snail, sparrow, spider, toucan, walrus, whooping crane, woodpecker, zoo.

Dinosaurs

Item	AF	#1	Item	AF	#1
t-rex	28	19	triceratops	12	0
brontosaurus	27	7	velociraptor	9	0
pterodactyl	20	1	raptor	7	1
tyrannosaurus rex	17	11	tyrannosaurus	5	3
stegosaurus	13	3	brachiosaurus	5	0

AF = 2: albertosaurus, alligator, allosaurus, ankylosaurus, diplodocus, long-neck, platysaurus.

AF = 1: 3 horn, apatosaurus, archaeopteryx, Barney, dimetrodon, Drumheller Alberta(1), duckbilled, edmontosaurus, encephlosaurus, flying one, Jurassic Park(1), little spindly leg one, mastodon, meat-eater, megasaurus, paradon, porkosaurus, pterosaurs, saber-toothed tiger, staphylosaurus, stethosaurus, trinoceroses(1), woolly mammoth.

Fish

Item	AF	#1	Item	AF	#1
goldfish	36	13	rainbow	6	0
salmon	31	7	angel fish	5	2
shark	26	4	blowfish	5	0
trout	23	1	minnow	5	0
tuna	18	0	rainbow trout	5	0
bass	16	1	carp	4	2
catfish	14	1	haddock	4	0
pike	13	5	halibut	4	0
swordfish	12	0	jellyfish	4	0
sunfish	9	0	beta fish	3	2
whale	8	1	mackerel	3	1
pickerel	7	1	eel	3	0
cod	7	0	fighting fish	3	0
perch	7	0	guppy	3	0
dolphin	6	1	large mouth bass	3	0
starfish	6	1			

AF = 2: arctic char, bowl, herring, Japanese fighting fish, manta ray, muskie, piranha, pond(1), puffer, Siamese fighting fish, small mouth bass, stingray, tropical fish(1), water.

AF = 1: achigan, barbotte, barracuda(1), big fish, big head, bluefish, bottom feeder, brochet(1), card game, chinook, chips, cichlid, clown fish, crayfish, dogfish, dorée, fish tank, Fridays, fried, heron(1), hilsa(1), Kraft Dinner, lakes, lobster, macero, neon, neon tetra, octopus(1), parrot fish, pearl, pisces, poisson rouge, redbelly, remora, salt water fish, sardine, scaly, sergeant fish, shellfish, shrimp, silver dollar, smells grass, sockeye,

sole, speckle, sucker, sushi, swim, tank, tilapia, tetra, the band, turbot, ugly, walleye, wet, whitefish.

Gods

Item	AF	#1	Item	AF	#1
Buddha	21	6	Athena	4	0
Zeus	21	4	Hera	4	0
Jesus	18	8	Muhammad	4	0
Allah	16	2	Thor	4	0
God	13	7	Shiva	3	2
Aphrodite	13	1	sun god	3	2
Hercules	8	2	Ganesh	3	1
Apollo	8	1	God (Christian)	3	1
Christ	5	2	Brahma	3	0
Vishnu	5	2	Krishna	3	0
Jupiter	4	2	Poseidon	3	0
Venus	4	1			

AF = 2: Achilles, Catholic God(1), church, Dionysus, Greek god, holy spirit, Jove, moon god, religion, Yahweh.

AF = 1: Abraham, Areas, Ares, Atlas, Buddhism, Catholic, Catholicism, clothes(1), Cupid, Diana, earth god, elephant god, Father, food god, Frigga, Gene Simmons, Greek god of Love, Greek god of Power, Greek god of Strength, Hephaestus, Hermes, Hinduism, India, invisible, Jebus ("Simpsons"), Jehovah , Jehovah God(1), Jehovah Witness, Jesus Christ(1), Jesus' Father, Jewish god, lord, Madame Blavatsky, Mars, Mary, me, Mercury, messiah, mon dieu(1), Morpheus, Neptune, Odin, pagan(1), Parvati, Persephone, plain spirituality, Pluto, Pygmalion, Ra, spiritual, Tiamat, water god.

Imaginary or Mythical Animals

Item	AF	#1	Item	AF	#1
unicorn	42	30	half man, half horse	6	1
dragon	13	4	fairy	4	0
centaur	9	0	phoenix	4	0
Loch Ness Monster	8	1	Snuffle-upagus	3	1
mermaid	8	1	sphinx	3	0
Pegasus	6	2	troll	3	0

AF = 2: Big Bird, Bigfoot, dragon (fire breathing), giant, griffin(2), merman, Minotaur(1), nymph, ogre, sea monster, vampire, werewolf.

AF = 1: 7-legged ostrich, big blue ox thing, Bugs Bunny, cherub, crows with huge wings, Cyclops(1), Daffy Duck, dinosaur(1), dinosaur that blows fire, dryad, dwarf, Easter Bunny, elf, Elmo, Fluffy ("Harry Potter"), flying dog ("Never Ending Story"), flying

horse, flying turkeys, fraggle, ghost, giant snake, gnome, goblin, heffalump(1), hippogriff, hobbit, horse with a horn, Jabba the Hutt, leprechaun, lion with two heads(1), muse, narasimha, Newt (on "Hercules"), Ogopogo, Pan(1), Pip (from "Hercules"), pixie, Polkaroo, Pticru, rabbit with antlers(1), reindeer, Road Runner, Sasquatch, siren, Snoopy(1), talking animal, trammar, Winnie the Pooh, woozle, Yoda, zombie.

Monsters

Item	AF	#1	Item	AF	#1
boogeyman	19	9	the one in the closet	5	1
Frankenstein	10	5	bigfoot	5	0
my pet monster	8	6	Dracula	4	1
Cookie Monster	8	3	Godzilla	4	0
Loch Ness Monster	8	2	Freddy Krueger	3	2
vampire	7	4	dragon	3	1
"Monsters Inc." monsters	7	3	Harry (Harry & the Hendersons)	3	0
ghost	7	3	King Kong	3	0
the one under the bed	7	1			

AF = 2: Cyclops, ogre, Sasquatch, Sesame Street characters, Shrek(1).

AF = 1: 3 eyes, Abominable Snowman, alien, Beetlejuice, boogie monster(1), Bush(1), creature from the black lagoon, Darth Vader, demon, devil, drunken monkey, E.T., Elmo, evil, Freddy, furry men, gargoyle, ghoul, green, green flying thing ("Ghostbusters"), Harry Potter books, Headless Horseman, Hitler, Hydra, Jabba the Hutt, jarasandh(1), Jason, Lorelei, Lucifer, marshmallow man, me(1), Medusa, Mike, moth men, Mr. Boogady, mummy, Ogopogo, Oscar the Grouch, parents, purple people eater, rakshas, Sixth Sense, Slimer, slimy, Staypuft Marshmallow Man, Sully, Swamp Thing, The Count, the guys in "Lord of the Rings", troll, werewolf, witch, wolf man, you(1).

Pets

Item	AF	#1	Item	AF	#1
dog	50	24	parrot	10	0
cat	50	18	horse	9	0
fish	37	2	monkey	9	0
hamster	36	1	pig	9	0
snake	27	0	spider	9	0
bird	22	2	goldfish	8	3
rat	18	0	ferret	7	0
mouse	16	0	budgie	6	0
iguana	15	0	canary	6	0
lizard	14	0	frog	6	0
turtle	14	0	tarantula	5	0
gerbil	13	0	chinchilla	4	0
guinea pig	12	0	cow	3	0
rabbit	12	0			

AF = 2: ant, bug, bunny, hedgehog, lovebird, parakeet, pot-bellied pig, sea monkeys, shark.

AF = 1: alligator (baby), bat, black lab, bunny rabbit, butterfly, chameleon, Chihuahua, elephant (in "The Simpsons"), flying squirrel, Japanese fighting fish, keeshond, newt, nut, peroquet, pet sister, plant, python, reptile, rock, salamander, scorpion, shark (baby), snail, tiger, weasel.

Pet Fish

Item	AF	#1	Item	AF	#1
goldfish	47	35	Japanese fighting fish	4	0
piranha	13	2	Siamese fighting fish	4	0
angel fish	11	1	tropical fish	4	0
guppy	8	3	neon	3	1
shark	8	0	tetra	3	1
fighting fish	6	0	mini-shark	3	0
minnow	5	0	sucker fish	3	0
beta fish	4	1			

AF = 2: algae eater, carp, catfish, fish food, neon tetra, net, tank(1), ugly.

AF = 1: a dead one, Arnold's fish from "Different Strokes"(1), blackfish, blowfish, bluefish, bottom feeder, bowl, Brian Mulroney fish, bubbles, castle, Chinese fighting fish, cleaning, clown fish, comet fish, dogfish(1), dolphin, dragon fish, fake plants, fan tail, feed, feeder fish, figurines, fish, food, heater, in a bowl(1), Junkyard George(1), Kraft Dinner, lab experiment, large fish, marshmallows, mini-piranha, oskar, parrot fish, pet

store, pike, puffer, pump, rainbow fish, red fish(1), salmon, samurai fighter fish, sand, silver dollar, small fish, snail, starfish, stingray, swim, tiger, toilet, tuna, turtle, water, whale, whitefish.

Superheroes

Item	AF	#1	Item	AF	#1
Superman	49	30	Catwoman	5	0
Spiderman	40	4	Green Lantern	5	0
Batman	30	7	X-Man	5	0
Wonder Woman	18	2	Captain America	4	0
Superwoman	12	1	Flash	4	0
Wolverine	11	1	Hercules	4	0
Incredible Hulk	11	0	The Tick	4	0
Robin	10	0	Cyclops	3	0
She-Ra	9	0	Mighty Mouse	3	0
He-Man	7	1	Power Ranger	3	0
Storm	6	0	Silver Surfer	3	0

AF = 2: Aquaman, Batgirl, Harry Potter(1), Prof. X , Rogue(1), Sailor Moon, Spawn, Supergirl.

AF = 1: Bionic Woman, Captain Kangaroo, Captain Kirk, Captain Marvel(1), Captain Planet, Cheetara, Clark Kent, Darkwing Duck, Dr. Freeze, Earthworm Jim, Flash Gordon, G.I. Joe, Gene Simmons, Ice Man, Iron Man, Joker, Jubilee, Liono, Mr. Freeze, My father to me, Mystique, Phoenix, Punisher, Radioactive Man, Rage, Rainbow Brite, Rubberman, Sakh Muzib(1), Tarzan, Teenage Mutant Ninja Turtle, Thing, Thor, Xena, Zeus, Zorro.

B: Typicality (Experiment 2)

The following tables show the average typicality ratings for every item in study 2, as well as the total numbers of participants who did not recognize the item and who indicated it did not belong in the category. The table columns are as follows: "Item" = item name; "Typ" = mean typicality (N=49); "?" = number of participants circling the question mark, indicating they were not familiar with the item (max 49); and "×" = number of participants circling the "×" symbol, indicating they did not consider the item a member of the category (max 49). An asterisk indicates the item did not reach significance on a χ^2 test against the flat distribution. A superscript "+" indicates the item was not carried forward to Experiment 3.

An Animal

Item	Typ	?	×	Item	Typ	?	×
a bear	6.78	0	0	a hippopotamus	6.04	0	1
a dog	6.73	0	0	a kangaroo	5.78	0	0
a tiger	6.73	0	1	a chicken	5.69	0	1
a horse	6.63	0	0	a ferret	5.45	1	1
a monkey	6.58	0	1	a penguin	4.84	0	0
a cow	6.49	0	0	a canary	4.65	0	1
an elephant	6.47	0	0	an octopus*	4.15	0	2
a pig	6.41	0	0	a phoenix*$^{+}$	3.91	13	13
a sheep	6.40	0	0	a unicorn*$^{+}$	3.60	0	17
a moose	6.39	0	0	a fraggle^{+}	2.43	18	17
a donkey	6.29	0	0	a centaur^{+}	2.38	17	10
a giraffe	6.21	0	1	a mermaid^{+}	2.12	1	22
a beaver	6.15	0	0				

Mean typicality for items: `5.41` (`6.04` for referring terms only)
Standard Deviation: `1.49` (`0.74` for referring terms only)

Percentage "?" responses: `4.1` (`0.1` for referring terms only)
Percentage "×" responses: `7.2` (`0.9` for referring terms only)

A Monster

Item	Typ	?	×	Item	Typ	?	×
Godzilla	6.31	1	0	a gargoyle	4.86	3	4
a cyclops	6.09	3	2	Abominable Snowman	4.84	0	5
Frankenstein	6.04	0	1	a troll	4.42	1	3
a werewolf	5.67	0	1	a ghost*	3.78	0	4
Loch Ness Monster	5.49	2	2	an alien*	3.70	0	6
a demon	5.40	0	2	a witch*	3.50	0	7
Medusa	5.23	5	5	Cookie Monster	3.34	0	8
a vampire	5.19	0	6	an alligator^{+}	2.41	0	22
an ogre	5.15	2	1	a shark^{+}	2.28	0	22
a dragon	5.13	0	1	a gorilla^{+}	2.15	0	22
a mummy	5.05	0	8	a rhinoceros^{+}	1.95	1	27
a boogeyman	5.02	1	1	a snake^{+}	1.92	0	25
Bigfoot	4.96	1	2				

Mean typicality for items: 4.39 (4.96 for non-referring terms only)
Standard Deviation: 1.38 (0.84 for non-referring terms only)

Percentage "?" responses: 1.9 (1.6 for non-referring terms only)
Percentage "×" responses: 15.3 (7.0 for non-referring terms only)

A Fish

Item	Typ	?	×	Item	Typ	?	×
a trout	6.74	1	1	a swordfish	5.47	0	0
a salmon	6.65	0	0	a guppy	5.22	6	2
a goldfish	6.45	0	0	a puffer fish	5.17	6	1
a pickerel^{+}	6.45	9	1	a fighting fish	5.14	5	2
a carp	6.12	3	2	a neon tetra^{+}	5.00	25	5
a catfish	6.10	0	0	a suckerfish*	4.95	6	3
a sunfish	6.05	5	0	a sardine	4.83	0	1
a piranha	5.82	1	3	a stingray	3.93	4	3
an angel fish	5.76	5	2	an eel*	3.13	0	3
a minnow	5.54	5	3	a shrimp	2.55	0	11

Mean typicality for items: 5.35
Standard Deviation: 1.11

Percentage "?" responses: 8.6
Percentage "×" responses: 4.8

A Superhero

Item	Typ	?	×	Item	Typ	?	×
Superman	6.78	0	0	Supergirl^{+}	5.28	5	3
Spiderman	6.71	0	0	Zorro	5.19	0	2
Batman	6.45	0	0	Bionic Woman	5.11	8	3
He-Man	6.38	3	1	Radioactive Man	4.93	8	1
Wonder Woman	6.32	1	1	Green Lantern^{+}	4.81	18	4
She-Ra	6.08	9	2	Xena	4.81	0	1
Hercules	5.62	1	3	Tarzan*	4.11	0	3
Captain America	5.59	8	0	Mighty Mouse*	3.93	1	2
Batgirl	5.42	2	2	Sailor Moon*	3.62	0	2
Incredible Hulk	5.35	3	3	Captain Kirk*	3.51	4	10

Mean typicality for items: 5.30
Standard Deviation: 0.99

Percentage “?” responses: 7.5
Percentage “×“ responses: 4.7

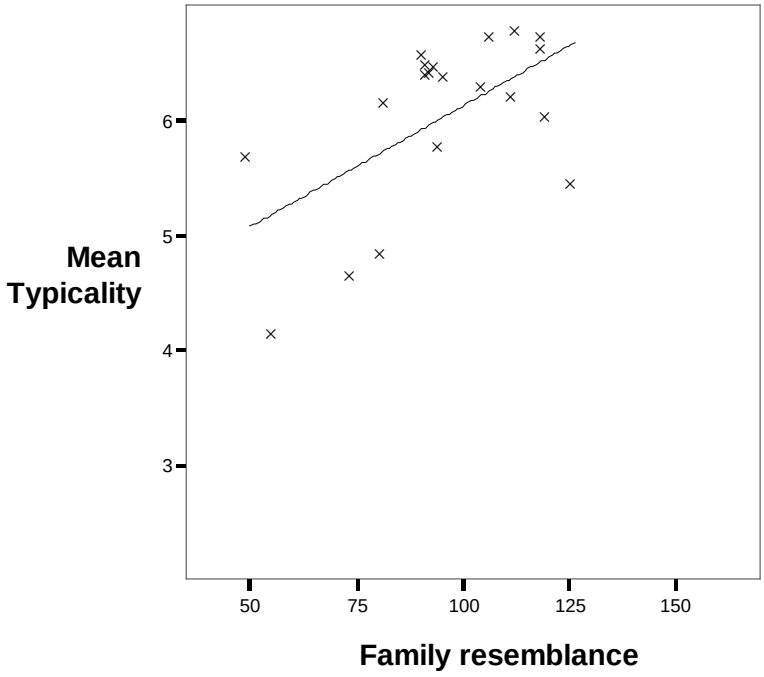
C: Family Resemblance

The following tables show the family resemblance scores computed for all 20 items in all 4 categories. Mean typicality scores from Appendix B are also included for easy comparison. Items are sorted by typicality. "Typ" = mean typicality, "FR" = Family Resemblance. Following the tables are scatter plots for each category showing typicality vs. family resemblance and with outlying points labeled where appropriate.

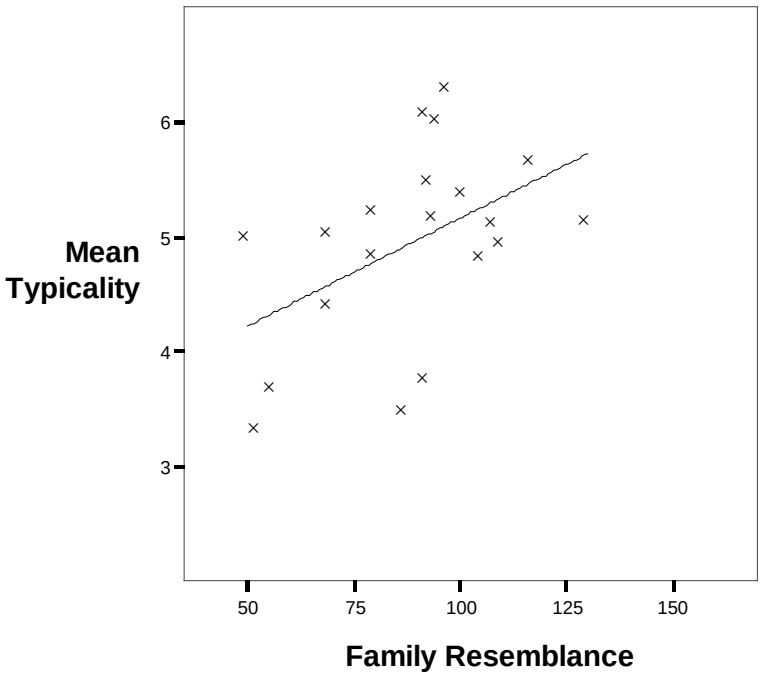
ANIMAL			MONSTER		
Item	Typ	FR	Item	Typ	FR
a bear	6.78	113	Godzilla	6.31	97
a dog	6.73	107	a cyclops	6.09	92
a tiger	6.73	119	Frankenstein	6.04	95
a horse	6.63	119	a werewolf	5.67	117
a monkey	6.58	91	Loch Ness Monster	5.49	93
a cow	6.49	92	a demon	5.40	101
an elephant	6.47	94	Medusa	5.23	80
a pig	6.41	93	a vampire	5.19	94
a sheep	6.40	92	an ogre	5.15	130
a moose	6.39	96	a dragon	5.13	108
a donkey	6.29	105	a mummy	5.05	69
a giraffe	6.21	112	a boogeyman	5.02	50
a beaver	6.15	82	Bigfoot	4.96	110
a hippopotamus	6.04	120	a gargoyle	4.86	80
a kangaroo	5.78	95	Abominable Snowman	4.84	105
a chicken	5.69	50	a troll	4.42	69
a ferret	5.45	126	a ghost	3.78	92
a penguin	4.84	81	an alien	3.70	56
a canary	4.65	74	a witch	3.50	87
an octopus	4.15	56	Cookie Monster	3.34	52

FISH			SUPERHERO		
Item	Typ	FR	Item	Typ	FR
a trout	6.74	128	Superman	6.78	148
a salmon	6.65	130	Spiderman	6.71	126
a goldfish	6.45	141	Batman	6.45	132
a pickerel	6.45		He-Man	6.38	149
a carp	6.12	154	Wonder Woman	6.32	136
a catfish	6.10	166	She-Ra	6.08	150
a sunfish	6.05	150	Hercules	5.62	107
a piranha	5.82	93	Captain America	5.59	106
an angel fish	5.76	122	Batgirl	5.42	124
a minnow	5.54	163	Incredible Hulk	5.35	64
a swordfish	5.47	135	Supergirl	5.28	
a guppy	5.22	117	Zorro	5.19	104
a puffer fish	5.17	110	Bionic Woman	5.11	80
a fighting fish	5.14	105	Radioactive Man	4.93	72
a neon tetra	5.00		Green Lantern	4.81	
a suckerfish	4.95	98	Xena	4.81	125
a sardine	4.83	116	Tarzan	4.11	80
a stingray	3.93	125	Mighty Mouse	3.93	124
an eel	3.13	119	Sailor Moon	3.62	87
a shrimp	2.55	130	Captain Kirk	3.51	38

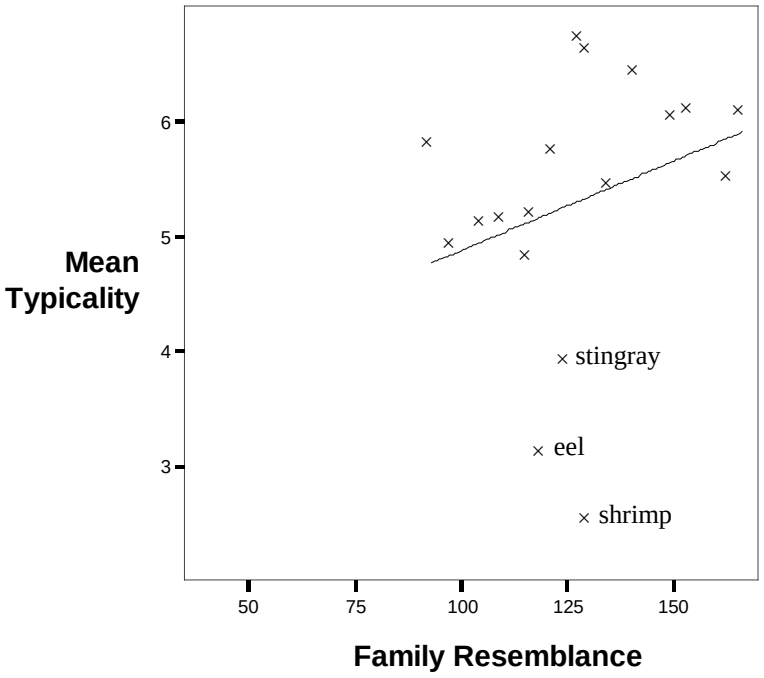
ANIMAL Scatter Plot



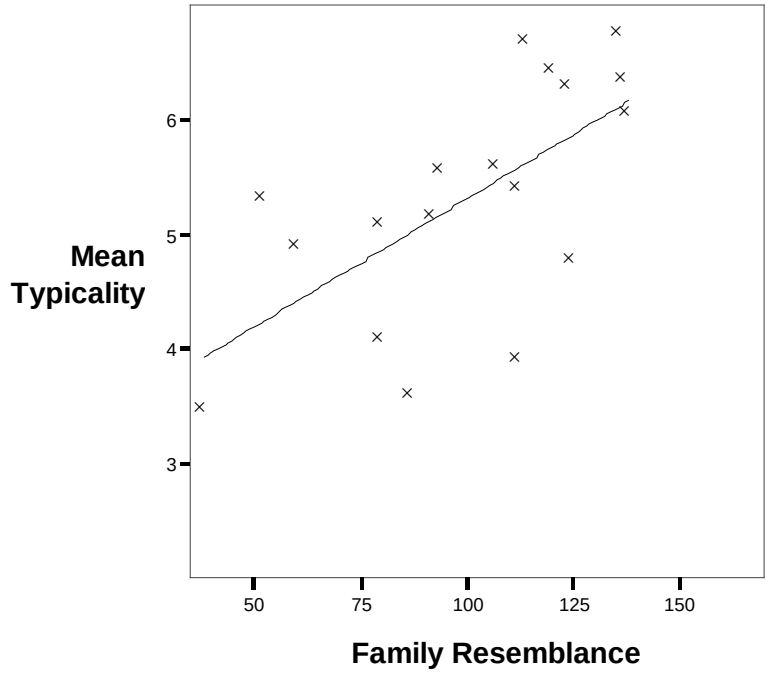
MONSTER Scatter Plot



FISH Scatter Plot



SUPERHERO Scatter Plot



D: Typicality (Experiment 3a)

The first two tables below show the mean typicality scores for Groups A and B of Experiment 3a on the four target categories. For convenience, both the typicality scores from Experiment 2 and the family resemblance scores from Experiment 3 are included in the tables. Items are sorted by typicality from Experiment 2. "E2" = mean typicality from Experiment 2, "A" = mean typicality from Group A of Experiment 3a, "B" = mean typicality from Group B of experiment 3a, "FR" = Family Resemblance. The next two tables show the typicality ratings from Group B compared to those generated in (Rosch, 1975). "R" = mean typicality ratings from participants in (Rosch, 1975). Following the tables are scatter plots of mean typicality vs. family resemblance for the MONSTER category.

ANIMAL					MONSTER				
Item	E2	A	B	FR	Item	E2	A	B	FR
bear	6.78	6.75	6.81	113	Godzilla	6.31	6.20	6.00	97
dog	6.73	6.91	6.87	107	cyclops	6.09	5.58	5.67	92
tiger	6.73	6.72	6.84	119	Frankenstein	6.04	6.26	6.06	95
horse	6.63	6.94	6.77	119	a werewolf	5.67	6.19	5.45	117
monkey	6.58	6.59	6.77	91	Loch Ness Monster	5.49	4.93	5.41	93
cow	6.49	6.63	6.68	92	demon	5.40	5.75	5.66	101
elephant	6.47	6.72	6.68	94	Medusa	5.23	5.54	5.14	80
pig	6.41	6.59	6.77	93	vampire	5.19	5.93	5.61	94
sheep	6.40	6.72	6.50	92	ogre	5.15	6.04	5.34	130
moose	6.39	6.47	6.35	96	dragon	5.13	4.30	5.17	108
donkey	6.29	6.25	6.13	105	mummy	5.05	5.66	4.80	69
giraffe	6.21	6.35	6.68	112	boogeyman	5.02	6.00	6.10	50
beaver	6.15	6.03	6.06	82	Bigfoot	4.96	4.53	5.13	110
hippopotamus	6.04	5.83	6.26	120	gargoyle	4.86	5.26	4.57	80
kangaroo	5.78	6.47	6.39	95	Abominable Snow…	4.84	4.27	4.62	105
chicken	5.69	4.93	6.13	50	troll	4.42	5.44	4.97	69
ferret	5.45	5.72	5.07	126	ghost	3.78	3.83	3.85	92
penguin	4.84	4.73	5.10	81	alien	3.70	3.74	3.43	56
canary	4.65	3.96	3.96	74	witch	3.50	4.21	3.53	87
octopus	4.15	3.31	4.22	56	Cookie Monster	3.34	3.67	3.50	52

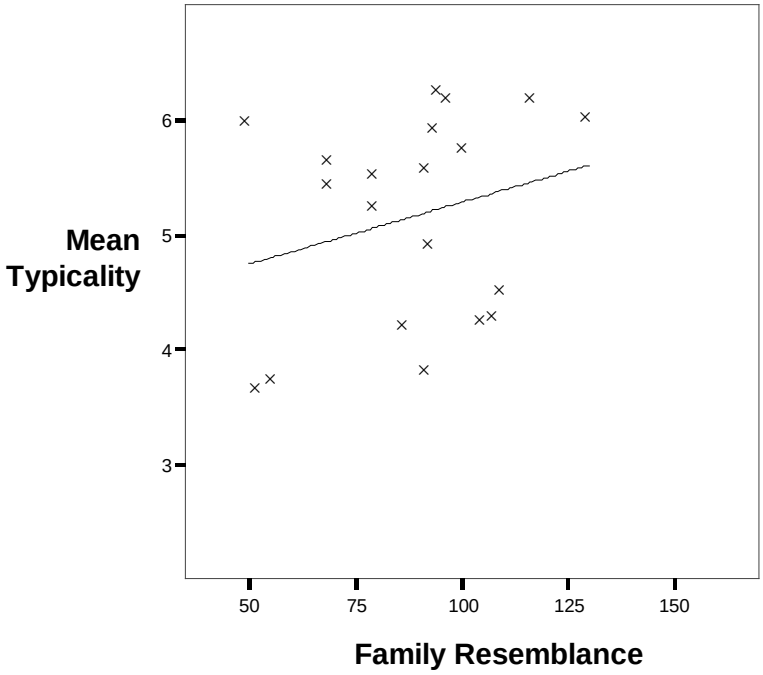
FISH					SUPERHERO				
Item	E2	A	B	FR	Item	E2	A	B	FR
trout	6.74	6.69	6.55	128	Superman	6.78	7.00	6.97	148
salmon	6.65	6.97	6.90	130	Spiderman	6.71	6.78	6.81	126
goldfish	6.45	6.88	6.81	141	Batman	6.45	6.69	6.39	132
pickerel	6.45	6.00	5.57		He-Man	6.38	5.86	5.70	149
carp	6.12	6.68	5.89	154	Wonder Woman	6.32	6.30	5.97	136
catfish	6.10	6.10	6.06	166	She-Ra	6.08	5.58	4.87	150
sunfish	6.05	6.04	5.86	150	Hercules	5.62	6.23	5.90	107
piranha	5.82	5.97	5.80	93	Captain America	5.59	6.28	5.71	106
angel fish	5.76	5.86	5.67	122	Batgirl	5.42	6.19	4.87	124
minnow	5.54	5.96	5.38	163	Incredible Hulk	5.35	5.96	5.57	64
swordfish	5.47	5.84	6.19	135	Supergirl	5.28	6.25	5.50	
guppy	5.22	5.81	5.59	117	Zorro	5.19	5.20	4.10	104
puffer fish	5.17	5.83	5.71	110	Bionic Woman	5.11	5.19	4.42	80
fighting fish	5.14	5.73	5.21	105	Radioactive Man	4.93	5.61	4.81	72
neon tetra	5.00	5.30	5.31		Green Lantern	4.81	5.50	5.00	
suckerfish	4.95	5.55	5.30	98	Xena	4.81	5.61	4.45	125
sardine	4.83	5.61	5.34	116	Tarzan	4.11	4.81	3.47	80
stingray	3.93	4.93	4.36	125	Mighty Mouse	3.93	4.56	3.89	124
eel	3.13	3.65	4.71	119	Sailor Moon	3.62	4.08	3.14	87
shrimp	2.55	3.73	3.74	130	Captain Kirk	3.51	3.00	2.47	38

CLOTHING			FURNITURE			SPORT		
Item	R	B	Item	R	B	Item	R	B
shirt	1.14	6.94	chair	1.04	6.71	football	1.03	7.00
skirt	1.21	6.74	table	1.10	6.65	tennis	1.15	6.58
jacket	1.68	6.29	dresser	1.37	6.39	handball	1.42	4.70
sweater	1.89	6.84	love seat	1.44	6.61	ice hockey	1.45	6.87
underpants	2.01	5.32	bed	1.58	6.61	boxing	1.66	6.10
socks	2.13	5.06	divan	1.70	4.38	skiing	1.76	6.23
pajamas	2.25	5.35	chest	1.98	5.42	surfing	1.84	5.06
raincoat	2.44	4.97	bookcase	2.15	5.90	cricket	1.99	5.10
slip	2.67	4.38	ottoman	2.43	5.30	racing	2.08	5.61
tuxedo	2.76	5.84	bench	2.77	4.90	bowling	2.18	4.38
nylons	2.98	3.97	lamp	2.94	4.52	skating	2.39	5.57
cape	3.38	3.34	drawers	3.63	4.73	sailing	2.44	4.74
sandals	3.56	3.22	magazine rack	4.14	3.41	judo	2.63	4.84
belt	3.93	4.04	stereo	4.32	4.00	rowing	2.82	5.42
slippers	4.08	3.55	television	4.41	4.60	running	3.01	5.16
gloves	4.53	3.86	pillow	5.03	2.71	weight lifting	3.59	3.86
handkerchief	5.87	1.85	sewing machine	5.39	2.27	boating	3.75	3.90
ring	6.11	2.36	clock	5.48	3.59	hunting	4.05	4.16
cuff links	6.18	2.17	picture	5.75	3.04	chess	5.07	2.64
cane	6.25	1.53	ashtray	6.35	1.65	cards	5.79	2.50

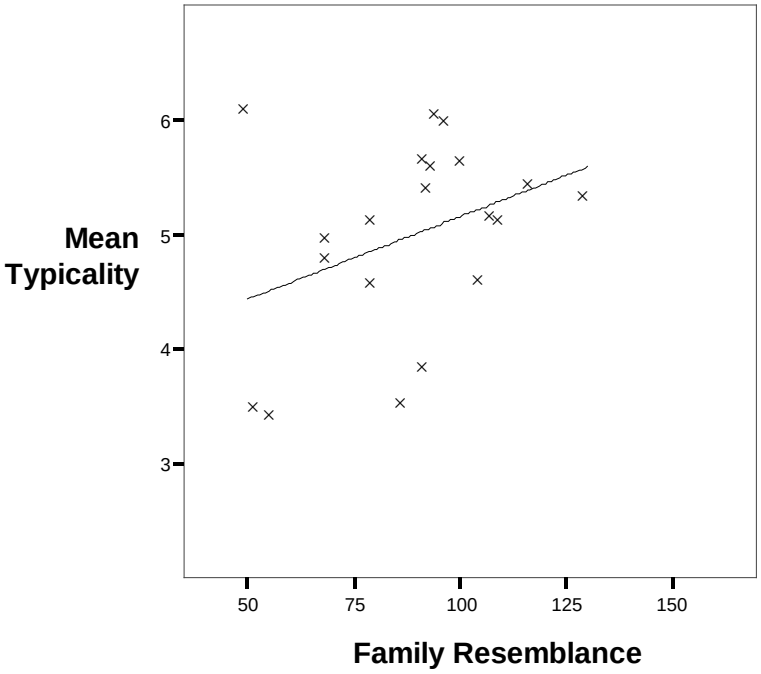
TOY			VEGETABLE			WEAPON		
Item	R	B	Item	R	B	Item	R	B
doll	1.41	6.35	pea	1.07	5.77	gun	1.03	7.00
toy soldier	1.61	6.19	spinach	1.22	5.81	rifle	1.17	6.90
marbles	1.74	4.77	asparagus	1.41	6.17	knife	1.40	6.52
water pistol	1.88	5.30	brussels sprouts	1.72	6.13	sword	1.47	6.39
rocking horse	1.91	5.13	celery	1.90	6.71	a-bomb	1.69	6.84
jacks	2.09	5.14	tomato	2.23	5.68	bazooka	1.76	6.21
jump rope	2.29	4.60	artichokes	2.32	4.74	club	2.09	5.06
wagon	2.49	4.77	eggplant	2.38	5.50	bullet	2.44	4.93
tricycle	2.61	4.24	radishes	2.51	5.74	arrow	2.66	4.84
truck	2.78	5.93	green onion	2.60	6.00	tank	2.74	5.41
balloon	3.07	3.22	parsnip	2.91	4.73	ice pick	3.14	3.80
drum	3.23	4.45	watercress	3.04	4.14	fist	3.31	3.90
swing	3.35	3.83	peppers	3.21	6.17	razor	3.82	3.72
sled	3.43	4.39	parsley	3.32	4.27	stick	4.04	3.03
checkers	3.56	3.83	mushroom	3.56	5.86	rock	4.18	3.20
bicycle	3.68	4.70	kale	3.67	3.80	chain	4.45	3.59
sandbox	3.71	4.29	pickles	4.57	4.37	pitchfork	4.67	2.86
rope	4.20	2.11	seaweed	5.04	2.65	hand	5.01	3.07
cards	4.56	3.29	dandelion	5.20	1.73	airplane	5.09	2.76
tennis racket	5.40	2.73	peanut	5.56	2.42	screwdriver	5.40	2.45

More MONSTER Scatter Plots

Using typicality data from Group A of the Experiment 3a:



Using typicality data from Group B of the Experiment 3a:



E: Speeded Categorization

The following two tables are lists of the stimuli used in the speeded categorization study, organized by category and mean typicality level ("Typ. Level"). Also shown are the mean typicality scores from Experiment 2 ("Typ"), the Kučera and Francis (1967) word frequency ("K-F"), the number of question marks the item received in Experiment 2 ("?"), and the category name used for that item in negative classifications ("Foil").

Non-referring Categories

Item	Typicality	K-F	?	Foil
	MONSTER			
High Typicality				
godzilla	6.31	0	1	superhero
cyclops	6.09	0	3	fish
frankenstein	6.04	0	0	animal
werewolf	5.67	0	0	fish
demon	5.40	9	0	fish
Low Typicality				
boogeyman	5.02	0	1	superhero
bigfoot	4.96	0	1	superhero
gargoyle	4.86	0	3	animal
troll	4.42	0	1	superhero
witch	3.50	5	0	animal
	SUPERHERO			
High Typicality				
superman	6.78	0	0	animal
spiderman	6.71	0	0	fish
batman	6.45	0	0	monster
he-man	6.38	0	3	animal
hercules	5.62	3	1	fish
Low Typicality				
batgirl	5.42	0	2	animal
supergirl	5.28	0	5	monster
zorro	5.19	0	0	fish
xena	4.81	0	0	animal
tarzan	4.11	0	0	monster

Referring Categories

Item	Typicality	K-F	?	Foil
		FISH		
High Typicality				
trout	6.74	4	1	monster
salmon	6.65	3	0	superhero
pickerel	6.45	0	9	monster
carp	6.12	0	3	superhero
piranha	5.82	0	1	superhero
Low Typicality				
guppy	5.22	0	6	monster
sardine	4.83	0	0	monster
stingray	3.93	0	4	superhero
eel	3.13	2	0	superhero
shrimp	2.55	2	0	monster
		ANIMAL		
High Typicality				
tiger	6.73	7	0	fish
monkey	6.58	9	0	monster
elephant	6.47	7	0	superhero
pig	6.41	8	0	fish
moose	6.39	0	0	fish
Low Typicality				
kangaroo	5.78	0	0	fish
ferret	5.45	1	1	superhero
penguin	4.84	0	0	monster
canary	4.65	0	0	monster
octopus	4.15	1	0	superhero

F: Familiarity (Experiment 4a)

These tables show the average familiarity ratings collected in Experiment 4a (N=32), presented along with the mean reaction times by item as measured in Experiment 4, and mean typicality from Appendix B, sorted by reaction time, then mean typicality. "RT" = reaction time, "Fam" = familiarity, "?" = number of participants who circled the question mark on the familiarity task.

ANIMAL

Item	RT	Typ	Fam	?
tiger	769	6.73	6.13	0
moose	782	6.39	5.74	0
monkey	784	6.58	6.00	0
kangaroo	805	5.78	6.06	0
pig	822	6.41	6.39	0
elephant	827	6.47	6.06	0
ferret	876	5.45	4.57	1
penguin	885	4.84	5.71	0
canary	935	4.65	5.23	1
octopus	1069	4.15	5.42	0
bear		6.78	6.26	0
dog		6.73	6.77	0
horse		6.63	6.16	0
cow		6.49	6.68	0
sheep		6.40	6.16	0
donkey		6.29	5.58	0
giraffe		6.21	6.06	0
beaver		6.15	6.19	0
hippopotamus		6.04	5.87	0
chicken		5.69	6.45	0
animal			6.52	0

FISH

Item	RT	Typ	Fam	?
trout	734	6.74	4.94	0
salmon	744	6.65	5.71	0
carp	755	6.12	4.25	3
guppy	769	5.22	3.59	2
eel	775	3.13	4.58	0
sardine	818	4.83	4.90	0
piranha	830	5.82	4.19	0
pickerel	852	6.45	4.00	7
shrimp	869	2.55	5.74	0
stingray	876	3.93	4.16	0
goldfish		6.45	6.32	0
catfish		6.10	4.87	0
sunfish		6.05	3.77	0
angel fish		5.76	3.33	1
minnow		5.54	4.64	3
swordfish		5.47	4.65	0
puffer fish		5.17	2.96	4
fighting fish		5.14	4.22	4
neon tetra		5.00	3.69	15
suckerfish		4.95	3.00	8
fish			6.48	0

MONSTER

Item	RT	Typ	Fam	?
Frankenstein	836	6.04	6.03	0
demon	895	5.40	5.00	0
werewolf	906	5.67	5.71	0
cyclops	938	6.09	4.26	0
witch	948	3.50	5.77	0
Bigfoot	951	4.96	5.13	0
gargoyle	968	4.86	4.74	0
boogeyman	985	5.02	4.61	0
Godzilla	991	6.31	5.39	0
troll	1060	4.42	5.19	0
Loch Ness Monster		5.49	4.77	1
Medusa		5.23	4.55	0
vampire		5.19	6.16	0
ogre		5.15	4.73	1
dragon		5.13	5.52	0
mummy		5.05	5.52	0
Abominable Snowman		4.84	4.67	1
ghost		3.78	5.74	0
alien		3.70	5.42	0
Cookie Monster		3.34	5.81	0
monster			5.71	0

SUPERHERO

Item	RT	Typ	Fam	?
Batman	783	6.45	6.35	0
Supergirl	791	5.28	3.43	3
He-man	801	6.38	4.67	1
Superman	813	6.78	6.74	0
Hercules	842	5.62	5.61	0
Batgirl	849	5.42	3.77	0
Spiderman	853	6.71	6.68	0
Xena	879	4.81	5.00	0
Zorro	914	5.19	5.35	0
Tarzan	946	4.11	5.52	0
Wonder Woman		6.32	4.16	0
She-ra		6.08	4.38	7
Captain American		5.59	3.72	6
Incredible Hulk		5.35	4.07	1
Bionic Woman		5.11	2.30	4
Radioactive Man		4.93	3.41	2
Green Lantern		4.81	3.70	11
Mighty Mouse		3.93	3.83	2
Sailor Moon		3.62	4.65	0
Captain Kirk		3.51	4.54	3
superhero			6.45	0

Index of Names

H

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

Z