Western Graduate&PostdoctoralStudies

Western University
Scholarship@Western

Electronic Thesis and Dissertation Repository

April 2011

Heuristics, Concepts, and Cognitive Architecture: Toward Understanding How The Mind Works

Sheldon J. Chow
The University of Western Ontario

Supervisor
Dr. Christopher Viger
The University of Western Ontario

Graduate Program in Philosophy

A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy

© Sheldon J. Chow 2011

Follow this and additional works at: https://ir.lib.uwo.ca/etd

Part of the Philosophy of Mind Commons

Recommended Citation
Chow, Sheldon J., "Heuristics, Concepts, and Cognitive Architecture: Toward Understanding How The Mind Works" (2011). *Electronic Thesis and Dissertation Repository*. 128.
https://ir.lib.uwo.ca/etd/128

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact tadam@uwo.ca.

HEURISTICS, CONCEPTS, AND COGNITIVE ARCHITECTURE: TOWARD UNDERSTANDING HOW THE MIND WORKS

(Spine title: Heuristics, Concepts, and Cognitive Architecture)
(Thesis format: Monograph)

by

Sheldon Chow

Graduate Program in Philosophy

A thesis submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

© Sheldon Joseph Chow 2011

THE UNIVERSITY OF WESTERN ONTARIO
School of Graduate and Postdoctoral Studies

CERTIFICATE OF EXAMINATION

Supervisor:

. . . .
Dr. Christopher Viger

Supervisory Committee:

. . . .
Dr. Robert Stainton

Examiners:

. . . .
Dr. John Nicholas

. . . .
Dr. Benjamin Hill

. . . .
Dr. Bertram Gawronski

. . . .
Dr. Richard Samuels

The thesis by

Sheldon Joseph Chow

entitled:

Heuristics, Concepts, and Cognitive Architecture: Toward Understanding How The Mind Works

is accepted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

. . . .
Date

. . . .
Chair of the Thesis Examination Board

Abstract

Heuristics are often invoked in the philosophical, psychological, and cognitive science literatures to describe or explain methodological techniques or "shortcut" mental operations that help in inference, decision-making, and problem-solving. Yet there has been surprisingly little philosophical work done on the nature of heuristics and heuristic reasoning, and a close inspection of the way(s) in which "heuristic" is used throughout the literature reveals a vagueness and uncertainty with respect to what heuristics are and their role in cognition. This dissertation seeks to remedy this situation by motivating philosophical inquiry into heuristics and heuristic reasoning, and then advancing a theory of how heuristics operate in cognition. I develop a positive working characterization of heuristics that is coherent and robust enough to account for a broad range of phenomena in reasoning and inference, and makes sense of empirical data in a systematic way. I then illustrate the work this characterization does by considering the sorts of problems that many philosophers believe heuristics solve, namely those resulting from the so-called frame problem. Considering the frame problem motivates the need to gain a better understanding of how heuristics work and the cognitive structures over which they operate. I develop a general theory of cognition which I argue underwrites the heuristic operations that concern this dissertation. I argue that heuristics operate over highly organized systems of knowledge, and I offer a cognitive architecture to accommodate this view. I then provide an account of the systems of knowledge that heuristics are supposed to operate over, in which I suggest that such systems of knowledge are concepts. The upshot, then, is that heuristics operate over concepts. I argue, however, that heuristics do not operate over conceptual content, but over metainformational relations between activated and primed concepts and their contents. Finally, to show that my thesis is empirically adequate, I consider empirical evidence on heuristic reasoning and argue that my account of heuristics explains the data.

Keywords: Heuristics; concepts; mental representation; cognition; cognitive architecture; relevance; the frame problem; reasoning; inference

Acknowlegements

I would like to thank Chris Viger, my supervisor, for all our discussions and for all his comments, suggestions, and general guidance as I wrote and worked through this dissertation. Because of him I was able to improve and sharpen my ideas and writing, and he was a tremendous help in turning a rough and vague idea about heuristics into the present work. I would also like to thank Rob Stainton for his help throughout the writing process. Further, although he probably doesn't know it, John Woods was an influential figure in my undergrad, and he got me interested in heuristics and humans as satisficing agents. So I extend sincere thanks to him for in some sense initiating my work. Barry Hoffmaster was also instrumental in initiating some of my ideas—it was after taking a class of his in my PhD program that I first realized that there wasn't an agreed-upon account of heuristics. Many thanks also go to my examination committee for their insightful questions, suggestions, and general input: John Nicholas, Ben Hill, Bertram Gawronski, and Richard Samuels. Finally, but not least, I thank my loving partner, Janet, who continues to support me and put up with me, even as I went through grad school and wrote this dissertation, and spent long days in front of my computer instead of by her side. ∞

Contents

List of Figures

Chapter 1

Introduction

1.1 Dissertation outline

As its subtitle suggests, this dissertation is an investigation into how the mind works. However, its goal is not as lofty as attempting to describe how the mind works *in toto*. Rather, this dissertation investigates only a specific aspect of how the mind works, namely how heuristics operate in cognition. Heuristics are generally understood as effective shortcut rules or procedures that require little cognitive resources to be deployed. In this dissertation, I argue that heuristics are ways to exploit the organization of mental concepts, and it is this that makes heuristics powerful shortcut procedures. More specifically, I do four things in this dissertation: (i) I motivate philosophical inquiry into heuristic reasoning and develop a characterization of "heuristic" that serves the practical interests of philosophy and cognitive science; (ii) I describe a cognitive architecture that I argue facilitates heuristic operations; (iii) I draw implications for a related and important philosophical problem, namely the so-called frame problem; and (iv) I illustrate how my thesis is empirically adequate.

Chapter 2 is solely devoted to task (i). Interest in heuristics in cognitive science exploded in the advent of the computational theory of mind, which encouraged a specific understanding of what is required of computational cognition: Feasible computation requires computationally tractable operations, and heuristics appear to be the perfect candidates to provide them. Yet, a survey of contemporary philosophical and cognitive science literatures reveals that different authors employ "heuristic" in ways that deviate from one another, and are sometimes inconsistent with one another. Given its widespread use in cognitive science, it is striking that there does not appear to be a consensus on what phenomena "heuristics" or "heuristic processes" refer to. In response, this dissertation considers a number of accounts found in the literature.

The most common accounts of "heuristic" assert something to the effect that "heuristics are processes that do not guarantee correct outcomes", or "heuristics are operations that do not optimize". These accounts, however, are merely perfunctory, and I find them to be unsatisfactory since the contexts in which there exist processes or operations that guarantee correct outcomes or that optimize are not typical of real-world situations. I go on to consider some positive accounts of “heuristic”, including the influential accounts of George Pólya (1957), Herbert Simon (e.g., Simon, 1957; Simon & Newell, 1958; Simon, Newell, Minsky, & Feigenbaum, 1967), Daniel Kahneman and Amos Tversky (e.g., Kahneman, Slovic, & Tversky, 1982a), and Gerd Gigerenzer (e.g., Gigerenzer, Todd, & the ABC Research Group, 1999). These researchers are not univocal with respect to how they characterize “heuristic”, but considering them allows me to draw a number of distinctions between different kinds of heuristics. Specifically, I distinguish between computational, perceptual, and cognitive heuristics; and between methodological and inferential heuristics. These distinctions allow me to hone in on a cluster of notions that is consistent with the influential positive accounts found in the literature, and which puts me in a position to develop a working characterization of “heuristic” of my own: *Heuristics are cognitive procedures that satisfice (i.e., set a reachable aspiration level relative to the goals of the agent), and that require little cognitive resources for their recruitment and execution; they operate by exploiting informational structures.* This working definition is developed with the aim of being coherent and robust enough to account for a broad range of phenomena in reasoning and inference. To the extent that my account is successful, we will better understand how the mind works and how heuristics fit within the conceptual structure of cognitive science.

Chapter 3 addresses tasks (ii) and (iii). Having a working definition of heuristics, I illustrate the work it does for us. I begin by considering a problem that many philosophers believe heuristics solve, namely the frame problem. However, in order to appreciate the role heuristics are supposed to have in solving the frame problem, we must gain an understanding of how heuristics work in cognition and the structures they operate over.

Generally understood, the frame problem is the problem of determining which of a potentially infinite number of beliefs are relevant to the task at hand. However, the frame problem actually constitutes a set of closely related problems. I expound various aspects of the frame problem and discuss their bearing on this dissertation. We will see that the controversy has to do with the idea that the frame problem poses a tractability problem for computational cognition.

Jerry Fodor (1983, 1987, 2000) believes that the frame problem (generally understood) is intractable, and therefore detrimental to computational cognitive science. Yet, since heuristics are generally understood to ensure computational tractability, heuristics are often invoked to circumvent the frame problem. Philosophers such as Peter Carruthers (2006a, 2006c) and Richard Samuels (2005, forthcoming) suggest that heuristics can serve as techniques that pick out computationally feasible subsets of information to be brought to bear on cognitive tasks. I contend, however, that merely pointing to heuristics to circumvent the frame problem reveals a misunderstanding of the problem. For the frame problem is not only a problem of explaining how computationally feasible subsets of information can be brought to bear on cognitive tasks, it is also a problem of explaining the reasonable levels of success that humans enjoy in determining what information is relevant to cognitive tasks; circumscribing what information is to be considered in a computationally tractable way does not necessarily ensure that what gets considered is in fact relevant.

I then proceed to investigate the kind of cognitive architecture that is able to facilitate heuristic processes. I argue that heuristics operate within domain-general cognition by exploiting *informationally rich*, specialized systems of knowledge. Heuristics can thus be fast and frugal processes in virtue of these epistemic structures (what I call *k-systems*), since the latter shoulder much of the informational burden of computational cognition. If it turns out that the assumptions about cognitive architecture which underlie my account of heuristics help in solving, or otherwise circumventing, the frame problem, that is reason to suppose that those assumptions are plausible.

In chapter 4, I continue with task (iii). An analysis reveals that concepts are the cognitive structures that fulfill the role of k-systems. Nevertheless, the understanding of the nature of concepts I adopt is unlike many of the common philosophical theories. Rather, it is an adaptation of a theory envisioned by psychologist Lawrence Barsalou (1999, 2008b, 2009). Barsalou's account is unlike many of the leading philosophical accounts insofar as it is grounded in perception rather than nonperceptual cognition or a "language of thought" hypothesis. The language of thought hypothesis claims that thought consists of the tokening of representations in a mental language ("Mentalese"), which possesses a productive and combinatorial syntax (as well as a semantics) to which cognitive operations are sensitive. Thus, following Alan Turing's (1936-7) idea of computation, the language of thought hypothesis postulates that thinking consists in syntactic operations defined over Mentalese representations. Barsalou, on the other hand, departs from this hypothesis by claiming that concepts are not amodal, formal representations, but are partly constituted by collections of neural patterns in the perceptual centres of the brain that were originally activated upon perceiving instances of referents picked out by the concepts.

I show how Barsalou's theory of concepts is closely related to a view that many philosophers, psychologists, and cognitive scientists have converged on, namely the file model of cognition. According to the file model, concepts are mental "files" that contain collections of information about the entities in their extension. Barsalou's theory of concepts may be understood as a psychologically and biologically plausible account of how concepts exhibit the kind of structure and organization envisioned by the file model. As I show, however, Barsalou's theory is not wholly compatible with the file model. I therefore offer a critical assessment of each account, which leads to certain modifications and qualifications. The result is a reconciled and more robust theory of concepts: *Concepts are highly organized collections of linguistically coded and perceptual information*. This theory of concepts will not be fully defended, for it is beyond the scope of this dissertation to offer such a defense. However, since my proposed cognitive architecture enables us to model heuristic cognition, my reliance on Barsalou's model is justified insofar as we are able to explain a wide range of phenomena (which, again, is a general aim of this dissertation).

What my theory of concepts and cognitive architecture implies for heuristic cognition is that the informational content of our concepts is organized in such a way that there exist patterns of relations among our activated and primed concepts. I contend that these informational relations guide and constrain heuristic operations. More precisely, I argue that heuristics do not operate over conceptual content; rather, they operate over the higher-order structural relations that exist between activated and primed concepts and their content. Such higher-order structural relations embody metainformation about the concepts in question, and this is what enables heuristics to be computationally frugal.

In chapter 5 I turn to task (iv). I illustrate how the account of heuristics developed to this point explains some of the empirical data in the psychological literature. Although the literature on heuristic reasoning is vast, I take up some of the evidence provided by the leaders in this kind of research, namely that provided by Kahneman and Tversky, and Gigerenzer. Specifically, I discuss two of Kahneman and Tversky's most developed heuristics (Representativeness and Availability) and two of Gigerenzer's most developed heuristics (the Recognition heuristic and Take the Best). I illustrate that the empirical data for each of these purported heuristics can be understood as arising from cognitive operations over inherent relations within and between one's concepts and their content. This serves as additional support for my thesis. In addition, this suggests that further empirical investigation can reveal the extent to which my thesis is correct: Once the content of and relations between one's concepts is determined, we can make predictions about what heuristics one will employ and what inferences or judgments one will thereby make.

Chapter 6 is the concluding chapter in which I return to an aspect of task (iii) that I have left unaddressed, namely how relevance is determined in cognition. In addressing this matter, I discuss what *relevance* is, and expound an influential theory developed by Daniel Sperber and Deirdre Wilson (1982, 1986/1995). I offer a reinterpretation of Sperber and Wilson's theory within the framework of cognition developed in this dissertation. This enables me to assess the role of heuristics in making determinations of relevance, as well as to explain the reasonable levels of success exhibited by humans in judging what is and is not relevant in their cognitive tasks. I then argue that our conceptual wherewithal predetermines what I call *de facto* relevance (cf. Gabbay & Woods, 2003). *De facto* relevance, as I explain, arises from the extant relations between activated and primed concepts and their content. I conclude that humans do not actually solve the problem of *determining* relevance; and so heuristic solutions as traditionally conceived to this problem *ipso facto* are empty. However, I go on to suggest that my thesis offers a more substantial understanding of how we pick out what is relevant in our cognitive tasks: *de facto* relevance happens to reflect, in robust but fallible ways, what really is relevant to the task at hand. Heuristics, operating over the informational relations among concepts, thus bring to bear an appropriate subset of (mostly) relevant beliefs on our cognitive tasks.

1.2 Computational cognition

I should take the time here to point out that throughout this dissertation I will be assuming, minimally, that the mind is an information processor. More specifically, I will be assuming that the mind is a computer, though I do not commit to classical computationalism. Indeed, the theory of concepts I advance in this dissertation runs contrary to the classical computationalism of the language of thought hypothesis. Nevertheless, I intend to leave open the matter of whether and the extent to which classical computationalism holds for cognition generally.

I will be advancing a central systems view of cognition which understands the mind to possess a holistic system, wherein there is free flow of conceptual information and wherein any conceptual information can bear on any other (to be discussed in chapter 3). Although I will not be arguing the matter here, I assume that the massive modularity hypothesis—i.e., the hypothesis that the mind is largely or completely composed of a large collection of individual, domain-specific computational systems (modules)—is incompatible with my thesis (and moreover, that it is in fact wrong). However, my thesis is amenable to Fodor's (1983) view that the mind is comprised of a central system in addition to a number of domain-specific, informationally encapsulated, input-output modules dedicated to peripheral functions, such as vision and language comprehension.

1.3 Philosophy in cognitive science

Some philosophers might consider some of the work done in this dissertation as not falling under the heading of philosophy. Of course, I do not believe that I am guilty of not doing philosophy. I do, however, believe that the kind of work undertaken in this dissertation may be mistakenly thought of as not being philosophy because the role of philosophy vis-à-vis cognitive science is not well understood. The confusion about the role of philosophy vis-à-vis cognitive science is no doubt partly attributable to cognitive science being a relatively young field of study, still without a standard doctrine. Yet it seems that philosophy has always had a rather uncertain role as a discipline of cognitive science (Bechtel, 2010; Brook, 2004).

To clarify the role of philosophy vis-à-vis cognitive science, Andrew Brook (2004) distinguishes philosophy *in* cognitive science from philosophy *of* cognitive science. According to Brook, philosophy of cognitive science is a branch of philosophy of science proper, as it seeks to investigate how scientists do cognitive science. As such, philosophy of cognitive science is better understood (by philosophers and cognitive scientists) than philosophy in cognitive science. Philosophy in cognitive science, on the other hand, is the philosophical investigation of issues concerning human cognition, such as mind and language, which are also studied empirically by psychology, linguistics, and other sciences. It is thus philosophy in cognitive science that is poorly understood.

However, William Bechtel (2010) has identified two ways in which philosophers have contributed to (and continue to contribute to) what Brook calls philosophy in cognitive science: Philosophers have informed and advanced work on the mind-body problem, or in its most contemporary incarnation, the problem of consciousness; and philosophers have offered philosophical treatments of mental representation, in terms of intentionality as well as representational function. In the present dissertation, I believe I have contributed to philosophy in cognitive science in the latter respect. Although I largely ignore issues pertaining to intentionality (although see chapter 4, section 4.2.2.1), I discuss the nature of concepts and representations in order to underwrite my theory of how heuristics work. Nevertheless, I believe that this dissertation contributes to philosophy in cognitive science in more ways than that. The work I do in this dissertation is philosophy, and so does not aim to offer new data, but it is an empirically informed philosophy, drawing on work from empirical disciplines in cognitive science (especially psychology, but also research from artificial intelligence and computer programming, as well as some from evolutionary anthropology). And I believe that this is how philosophy in cognitive science must be undertaken. So-called “armchair” philosophy is now found to be inadequate in many respects. We are at a time in which there are plenty of new and exciting advances in the brain sciences. If anything, empirical data must be understood to constrain philosophical inquiry. Indeed, what is found empirically to be the case should take precedence over conceptual consistency argued from the armchair. But a more positive consequence is that the constraining-relation can guide philosophical investigations. In this latter respect, however, there is potential for a mutual constraining-relation—as empirical work guides philosophical investigation, philosophical investigation can in turn potentially constrain empirical investigation. This, according to Daniel Dennett (2009), is how cognitive science can progress.

In light of these considerations, one might say that in this dissertation I adopt a *naturalized* philosophy of mind, which in many ways departs from the traditional approach to the philosophy of mind. Philosophers of mind have always discussed the nature of the mind and its functions, but as the special sciences of cognition developed (psychology, linguistics, neuroscience, etc.), the topics that were/are of interest to philosophers were/are seen by scientists to bear little relevance to their empirical investigations. Indeed, “classic” philosophical problems about the mind are often met today with a dismissive attitude (Bechtel, 2010; Brook, 2004; Dennett, 2009). For example, philosophers of mind have often appealed to drawn out thought experiments to argue for substantive conclusions about our concepts. Yet as Bechtel (2010) remarks, “In the context of cognitive science, with its ingenious research programs challenging the notion of fixed concepts, the ensuing philosophical arguments often seem beside the point” (p. 358).

Nevertheless, a naturalized philosophy of mind offers a way for philosophy to be accepted and have a central role in cognitive science. The approach that a naturalized philosophy of mind takes is inspired by W. V. O. Quine's (1969) idea of a naturalized epistemology. Quine's proposal was to situate epistemology within the empirical program of psychology, while still addressing fundamental questions about knowledge and how it is acquired. In his own words from a widely quoted passage: "Epistemology, or something like it, simply falls into place as a chapter of psychology and hence of natural science. ... But a conspicuous difference between old epistemology and the epistemological enterprise in this new psychological setting is that we can now make free use of empirical psychology" (pp. 82-83).

Quine's view that we should replace traditional epistemology with psychology is not widely accepted by contemporary epistemologists,[1] and even Quine's (1990) later views were more moderate (R. Feldman, 2008). Nevertheless, there is at least a kernel of wisdom in Quine's naturalized epistemology that can be applied to how we approach the philosophy of mind. A naturalized philosophy of mind benefits from the empirical research undertaken by the special sciences by aiming to offer conclusions that have a direct bearing on the practice of cognitive science. As Bechtel (2010) puts it,

> a naturalized philosophy of mind would become the philosophy of cognitive science and draw upon results in cognitive science as well as its own inquiries to probe the nature of mental phenomena. ... One might recognize philosophical discussions conducted in a naturalistic manner not as arcane intellectual exercises, but as theoretical and methodological contributions to cognitive science. As a bonus, unlike other sciences subjected to philosophical inquiry, cognitive science offers a two-way street: Its theoretical frameworks and findings are a resource philosophers can draw upon to add nuance and techniques to their own subsequent inquiries. (pp. 358-359)[2]

In this light, philosophy can contribute to and advance cognitive science by developing theoretical frameworks for investigating and explaining specific cognitive phenomena. This is

[1]In his famous criticism of Quine, for instance, Jaegwon Kim (1988) points out that to adopt Quine's naturalized epistemology would be to abandon traditional epistemology altogether. For, as Kim argues, traditional epistemology is interested in the epistemic support between our sensory experiences (our basic evidence) and our beliefs about the world, and this involves questions pertaining to rationality, justification, and knowledge. What Quine recommends, however, is to ignore such matters of epistemic support and investigate instead the causal connections between our sensory experiences and our beliefs.

[2]We should take Bechtel's use of "philosophy of cognitive science" as referring to "philosophy in cognitive science" in Brook's sense.

what I aimed to achieve through this dissertation, as I offer a theoretical framework concerning the architecture of cognition and the nature of concepts in order to explain how heuristics work. In this way, I take Dennett's advice: "If philosophers aren't to do the sort of refutation by counter-example and reductio ad absurdum that their training prepares them for, what is the alternative? To take on the burden of working out an implication or two that is actually testable, if not today, in the near future" (Dennett, 2009, p. 233).

Chapter 2

Characterizing “Heuristic”

A quick survey of contemporary philosophical and cognitive science literatures reveals that "heuristic" and its cognates are often used to label certain methods or procedures that guide inference and judgment in problem-solving and decision-making. In more specific terms, heuristics are commonly understood to be shortcut procedures that lead to results that may not be optimal but are in some sense "good enough". However, taking shortcuts carries risks. Though heuristics are understood to generally produce satisfactory outcomes, they are not failsafe procedures, and thus they can result in errors or mistakes. But it is a trade-off: heuristics forego optimal outcomes in favour of computational economy. Procedures that ensure correct or optimal outcomes are typically complex and difficult to carry out, whereas heuristics offer a way to avoid such a computational burden—they are, after all, supposed to be shortcuts.

Since heuristics are believed to economize, and since heuristics generally produce satisfactory outcomes, it is received wisdom that heuristics are a stock commodity in human cognition. Yet a close inspection of the literature reveals that different authors employ "heuristic" in ways that deviate from one another, and are sometimes inconsistent with one another. Given its widespread use in philosophy and cognitive science, it is striking to see that there does not appear to be a consensus on what phenomena "heuristics" or "heuristic processes" refer to. At the same time, however, defining precisely what heuristics are is a difficult task; as Jonathan St.B.T. Evans (2009) remarks, "The term 'heuristic' is deeply ambiguous" (p. 36). Since the term is used in diverse ways in different disciplines, and its meaning has evolved over the years (Gigerenzer & Todd, 1999), it is hard to pinpoint what "heuristic" exactly refers to. Indeed, it may very well not express a single, unified concept. Nevertheless, we might settle for a characterization broad enough to satisfy most of its uses, or at least some of the interesting ones.

In this chapter, we will be introduced to a number of different meanings assigned to "heuristic". I will examine these meanings with an aim to develop a working characterization of my own. I begin, in section 1, by briefly motivating the need for an appropriate understanding of "heuristic". In section 2, I discuss some negative characterizations that are often found in the literature. It is important that we begin by considering these negative characterizations, since many authors often appear to place too much emphasis on what heuristics fail to do. The most common definitions assert something to the effect of "heuristics are processes that do not guarantee correct outcomes" or "heuristics are operations that do not optimize". These types of definitions tell us what heuristics *do not* do, but they do not really tell us anything about what heuristics *do* do for us; in fact, as I shall argue, in certain contexts these definitions do not tell us anything interesting at all. By analyzing these negative definitions, my aim is to get clear on why we need a positive characterization of heuristics, as well as what an adequate, positive characterization needs to be in order to get a decent grasp on what heuristics are and their role in cognition.

Before offering my own positive characterization, however, I discuss some influential accounts of heuristics in section 3. In particular, I discuss those offered by George Pólya, Herbert Simon, Daniel Kahneman and Amos Tversky, and Gerd Gigerenzer. These authors are not univocal with respect to how they characterize "heuristic", but considering their accounts will allow me to hone in on the positive elements offered by them. In section 4, I synthesize these positive contributions. I draw a number of distinctions between different kinds of heuristics, and then develop my working definition of the sorts of heuristics that are believed to be employed in human cognition. My goal is to offer an account of heuristics that is coherent and robust enough to account for a broad range of phenomena in reasoning and inference,[1] and thus makes sense of empirical data in a systematic way.[2] As we work through this chapter, it will become obvious that the other accounts I survey—both positive and negative—do not provide these benefits.

---

[1]In light of certain ambiguities with the terms "coherent" and "robust", it should be understood that when I employ these terms throughout this dissertation I mean them, as I say, in the sense of consistently accounting for, or being applicable to, a broad range of phenomena.

[2]In chapter 5, I illustrate how the account of heuristics developed in this chapter, and elaborated upon in following chapters, accounts for empirical data.

2.1 Why we need an appropriate characterization of "heuristic"

Given its widespread use within (and without) philosophy and cognitive science, it is intriguing that there is little concern for a fully developed characterization of "heuristic". The lack of a developed characterization of the notion is disconcerting in two ways. First, since heuristics have rapidly made their way as central tenets in theories in cognitive science, we might expect something that approaches a refined and robust explication of what is meant by "heuristic". But, as far as I can tell, no one has produced such an explication. If cognitive science is indeed a science, then careful consideration should be taken to delimit heuristics *qua* entities of scientific theories. To put it another way, if cognitive science is in the business of discovering and studying the kinds within its domain, pains should be taken to appropriately characterize these kinds (Carruthers, 2006a; Pylyshyn, 1999), among which heuristics are included. For example, the dual-processes theory of cognition, which lately has been gaining popularity (and to be briefly discussed below), invokes heuristics to explain certain kinds of cognitive processes, namely what is known as "type 1" processes (see e.g., Wason & Evans, 1975; J. Evans & Over, 1996; J. Evans, 2009). Other psychological theories, such as that propounded by the heuristics and biases program (also to be discussed below), are based on the presumption that many of our judgments follow from a small set of basic heuristics (e.g., Kahneman et al., 1982a). Furthermore, heuristics are sometimes invoked to suggest ways to circumvent problems of determining what, of all the available information, should be considered in a cognitive task (Carruthers, 2006a, 2006c; Samuels, 2005, forthcoming)—a problem known as the frame problem (to be discussed in chapter 3). Being clear on what phenomena are picked out by "heuristic" should be of concern to both scientists and philosophers who wish to make sense of, and even adopt, such theories of cognition.

Secondly, many arguments—descriptive and normative—are made on the premise that heuristics are ubiquitous in human cognition. Such arguments try to decide the extent to which humans rely on heuristics, what heuristics people employ in what circumstances, and whether it is rational to do so. But vagueness (or, in the best case, ambiguity) in the concepts employed by a theory will be inherited by any arguments based on the theory. Unclarity in what heuristics are in the first place therefore impedes progress on matters involving the purported uses and abuses of heuristics in judgment, inference, and decision-making. In the face of vagueness (or ambiguity) in concepts, the arguments concerning human reasoning may very well lack a sufficient basis to be decisive. For instance, both the heuristics and biases research program and the fast and frugal heuristics research program (again, to be discussed below) assert that heuristics are pervasive in human cognition and have a substantial role in our judgment and decision-making. But if we do not know what cognitive processes "heuristics" refer to, then it is unclear whether and to what extent purported heuristics have a role in our cognitive lives. Indeed, it is sometimes hard to understand what claims are exactly being made by these research programs (cf. Samuels, Stich, & Faucher, 2004). Furthermore, there has been a long-standing debate between those who believe that heuristic strategies are or can be rational, and those who believe that rationality resides in the dictates of some normative theory which heuristic strategies do not meet (see e.g., Simon, 1957; Gigerenzer, 1996; Kahneman & Tversky, 1996; Samuels, Stich, & Bishop, 2002). But once again, if it is unclear what either party means by "heuristic" or "heuristic process", then it is equally unclear what the force of the arguments from either side of the debate is, or precisely what is being claimed.[3]

These two problems act as barriers that inhibit the advancement of research on cognition. To overcome these barriers, we need an account of heuristics that is coherent and robust enough to plausibly fit a natural kind,[4] on the one hand, as well as to provide clarity in the objects of

[3]It is instructive to note that Samuels et al. (2002) argue that the debate—which they dub the "rationality wars"—is not really much of a debate at all, since upon close inspection both parties actually agree: heuristics can sometime produce good outcomes, but can also lead to systematic errors, depending on the context. Thus, it must be recognized, I think, that much of the debate may very well stem from both sides not being clear on what heuristics are in the first place.

[4]The way in which "natural kind" is construed can vary quite widely. For the purposes of this dissertation, I follow Richard Samuels (2009) in adopting Richard Boyd's (1991) notion of "homeostatic property clusters". As Samuels explains, on this account a kind is natural if (i) it is associated with a range of particular characteristics which tend to be co-instantiated by instances of the kind (though these characteristics are not necessary conditions); (ii) there is some underlying causal mechanisms and constraints whose operations explain the co-instantiation of the characteristics of (i); and (iii) it is not the characteristics of (i) that define the kind, but the underlying causal mechanisms and constraints of (ii). For the remainder of this dissertation, I should be taken to

study, on the other. To the extent that an account of heuristics is successful with respect to these tasks, we will better understand how the mind works and how heuristics fit within the conceptual structure of cognitive science. As we shall see, the working definition of "heuristic" offered in this chapter will be developed such that these conditions are satisfied.

2.2 Abuses of "heuristic"

Some authors attempt to make clear their intended meaning of the word "heuristic". Other researchers, however, do not even bother to provide a simple definition, taking its meaning to be intuitive or obvious, when in fact it is not (cf. Simon's remarks below). Indeed, it is not uncommon for a book to be missing "heuristic" from its index, despite of the fact that the term was mentioned throughout.[5] Nevertheless, those authors who do provide definitions almost always offer simple, merely perfunctory definitions, without any concern for whether the implicated concept is coherent or robust. What I intend to do now is critically assess these definitions of "heuristic". My target will be negative definitions that characterize heuristics in terms of what they fail to do, while ignoring any detailed explication of their positive aspects. I will argue that such negative definitions do not provide an adequate understanding of what heuristics are, especially within the context of practical, real-world problems—problems that are of interest to cognitive science. More specifically, I will show that the definitions do not apply sensibly to a range of important cases. Considering these negative definitions will thus motivate the need for a positive characterization of heuristics. We will also get a clearer idea of what an adequate, positive characterization needs to be if heuristics are to be interesting and meaningful kinds within our theories of cognition.

refer to this view when I speak of natural kinds. According to Samuels, this is the most plausible view of natural kinds, and the view that comports best with cognitive science research on cognitive architecture, and I tend to agree (at least with the last part).

[5] One example is Michalewicz and Fogel's (2000) book *How to Solve It*. Michalewicz and Fogel's lack of concern for a clear understanding of what heuristics are is especially alarming; the subtitle of their book is "Modern Heuristics"!

2.2.1 Heuristics vs. guaranteed correct outcomes

A common feature among almost all perfunctory accounts of heuristics is that they are negatively defined (cf. Fodor, 2008, p. 116). A heuristic is thus understood to be deficient or inadequate in some respect. Indeed, it is not uncommon to happen upon a definition that states that heuristics are procedures that can produce good outcomes but do not guarantee correct solutions (e.g., Dunbar, 1998) or do not always work (e.g., Fodor, 2008). At other times one finds heuristics contrasted with algorithms (e.g., Richardson, 1998). Algorithm here is not meant as a "procedure determined by a set of instructions that specify, at each moment, precisely what is to be done next" (Simon et al., 1967, p. 18). If the latter were the case then the heuristic-algorithm distinction[6] would be meaningless for computer programming and artificial intelligence (AI) research, as well as for human cognition provided that one subscribes to a version of the computational theory of mind (CTM). Rather, in this context, algorithms are understood as procedures that, when applied properly, invariably produce correct outcomes; hence, heuristics are conceived simply as procedures that *do not* invariably produce correct outcomes. In one textbook on cognition, it is claimed that a "hallmark" of a heuristic approach to solving a problem is that "It will work part of the time or for part of the answer, but it is very unlikely to furnish the complete, accurate answer" (Ashcraft, 2002, p. 465). Thus, in these sorts of definitions we find a distinction between heuristic processes and processes that are guaranteed to produce correct answers or outcomes. I will refer to these latter as guaranteed-correct-outcome (GCO) procedures.

Consider an example. In *The Language of Thought*, Jerry Fodor (1975) observes that a failsafe method of understanding the message of a given utterance is to compute the grammatical relations exhibited by the sentence expressed by the utterance. He observes, however, that although people presumably can infer messages in this way, they demonstrably often do not. "What apparently happens," Fodor claims, "is that grammatical relations are computed only when all else fails. There exist heuristic procedures for sentence recognition which, in

[6]See Simon et al. (1967) for an informative discussion on the heuristic-algorithm distinction.

effect, ignore grammatical relations and infer messages directly from lexical content, accepting, thereby, the penalties of fallibility" (pp. 167-168). He asks us to consider sentences with self-embeddings, such as "The boy the girl the man knew wanted to marry left in a huff". With enough time, patience, and insight into its grammatical structure, one is able to work out the meaning of this sentence: The girl, whom the man knew, wanted to marry the boy who left in a huff. We might notice, however, that this meaning is not easy to come to. Yet if we compare "The boy the girl the man knew wanted to marry left in a huff" with the similarly structured sentence "The boat the sailor the dog bit built sank", we find the meaning of latter sentence more transparent. According to Fodor, what seems to be going on is that the message intended by "The boat the sailor the dog bit built sank" is inferred from background knowledge considerations, such as that boats typically sink (but not dogs or sailors), that sailors typically build boats (but not dogs), that it is more common to hear about dogs biting (as opposed to sailors), and that it is more likely that a sailor would be bitten than a boat. This kind of analysis suggests the plausibility of assigning a meaning to a sentence (like the one in question) without engaging a syntactic structural description.

What Fodor's remarks illustrate is a distinction between the sorts of mechanical computations that guarantee right answers (i.e., GCO procedures) from the sorts of shortcut procedures (i.e., heuristics) that avoid having to undertake those mechanical computations.

> In short, the computational load associated with the solution of a class of problems can sometimes be reduced by opting for problem-solving procedures that work only *most* of the time. Reliability is wagered for efficiency in such cases, but there are usually ways of hedging the bet. Typically, heuristic procedures are tried first; relatively slower, but relatively algorithmic, procedures are 'called' when the heuristics fail. The way of marshaling the available computational resources can often provide the optimal trade-off between the speed of computation and the probability of getting the right results. (pp. 166-167)

This sort of characterization of heuristics is not surprising in light of the fact that there is typically a risk of suffering negative consequences for cutting corners or taking shortcuts. As Fodor remarks, "one can often get away with less than strict compliance with such requirements [that guarantee a correct solution] so long as one is willing to tolerate occasional mistakes" (p.

166). Yet it is not the intended purpose of heuristics to make mistakes, but rather to offer a method of solving a problem that achieves computational economy. As Fodor understands them, heuristics avoid certain computations within the representational or conceptual structures of cognition (at least with respect to understanding sentences). In later chapters, we shall see how important it is for heuristics to avoid computations within conceptual structures. The present point, however, is that heuristics offer a way to cut down on the computations needed to arrive at a desired outcome. In addition, as intimated by Fodor and as it is often claimed, heuristics are able to deliver reasonable answers or solutions by exploiting the right information in the right circumstances.

Acknowledging the benefits afforded by heuristic procedures is very important in understanding the nature of heuristics. For we would hesitate to call a procedure that invariably leads to disaster a heuristic procedure; nor would we want to call a procedure that is irrelevant to the task at hand a heuristic procedure. Yet such procedures very certainly would not guarantee correct outcomes. Therefore, what heuristics fail to achieve (viz. reliably getting the right results) should rightly be thought of as a *corollary* of a definition of "heuristic", rather than made to be an essential feature.

Nevertheless, notice from Fodor's remarks that the benefits offered by heuristics are always contrasted with associated "penalties" or risks—whatever good heuristics do, they do not always get things right. In this way, Fodor provides a particularly clear example of the attitude toward heuristics that typically leads to negatively characterizing them in terms of what they fail to do, or what they do not achieve. This type of definition can be summed up as follows.

> $\mathbf{H}_1$ Although heuristics can provide certain benefits, they are procedures that do not guarantee correct outcomes.

I will discuss how heuristics "can provide certain benefits" in more detail when I offer my positive characterization below. For now, however, I will focus on the "do not guarantee correct outcomes" part of this definition. In particular, I shall address the extent to which the claim that heuristics are procedures that do not guarantee correct outcomes does useful work for us as a central part of a definition of heuristics. My argument will be that such a claim, by itself, does not do the work that we ought to expect from a definition of heuristics, hence affirming the comment recently made that what heuristics fail to achieve should be understood as a corollary of our definition rather than as an essential feature.

GCO procedures are essentially formal operations. This is exhibited, for instance, by the grammatical computations that generate the meaning of sentences. But the paradigm of GCO procedures are the formal operations of mathematics or logic. For example, applying the rules of division will guarantee that you will arrive at the correct answer to a given division problem of arithmetic; constructing truth-tables will guarantee a correct determination of whether a given set of propositions is consistent; applying the basic axioms of probability theory to a given chance set-up will guarantee a correct determination of chances or objective probabilities.[7] In general, GCO procedures necessitate the right answers, when applied correctly. Again, in contrast to these GCO procedures, heuristics are supposed to be procedures that efficiently produce outcomes or answers that may be good enough for one's purposes, and that may very well be correct, but will not *guarantee* that they will be correct. For example, randomly selecting a limited subset of a set of propositions and simply looking at its members to see if any contradicts any other is a heuristic process to check the consistency of the superset of propositions; another heuristic process is to roughly estimate probabilities based on remembered past experiences.

GCO procedures exist for problem tasks other than those proposed in logic or mathematics, as we saw with respect to Fodor's example. But there are other problem tasks in addition to these that admit GCO procedures. Suppose your lamp does not light when you switch it on, and you want to determine whether it needs to be replaced. Here is an algorithm that will guarantee a correct determination: First, (i) check to see if the lamp is plugged in; if so, then (ii) check to see if the bulb needs replacing (e.g., by testing it in another lamp that is known to be working);

[7] I say in each of these examples "a given ..." because certain initial conditions have to be met if the procedure is to produce a *correct* outcome: the numbers in the division problem must belong to the set of rational numbers; one must have the right initial probability assignments in the chance set-up problem.

if not, then (iii) ensure that the power outlet is working properly; if so, then (iv) infer that the lamp needs to be replaced. Of course, it might be easier and less time consuming to simply rely on a heuristic that assumes that the lamp needs replacing if it is plugged in but does not light when switched on. There are even algorithms for playing tic-tac-toe which guarantee either a win or draw (Crowley & Siegler, 1993). Yet most people do not know of these GCO tic-tac-toe algorithms, and use instead certain heuristics, such as "Always start in the centre square" (Dunbar, 1998).

So far we have considered a small number of contexts in which there exist GCO algorithms, as well as procedures that do not guarantee correct outcomes but may approximate good outcomes, or at least produce satisfactory outcomes insofar as certain goals are met (e.g., computational efficiency); and these latter procedures we are calling "heuristic". Within these contexts H_1 makes perfect sense. However, these contexts do not represent all possible, or indeed common, circumstances. Of concern is the extent to which H_1 is usefully applicable within contexts that are of practical interest to cognitive science, namely typical real-world contexts.

As we have observed, H_1 is applicable in the paradigm case when there exists a known, possible, feasible GCO procedure for a given problem. But H_1 is also applicable when there exists a known, possible GCO procedure that is not feasible, as when the GCO procedure cannot be practically performed by a human (or even machine) because it is too complex, or too time-consuming, or too taxing on cognitive (computational) resources. In this situation, there can very well be procedures that do not guarantee correct outcomes, but can lead to good or useful outcomes, and we might call such procedures heuristic. H_1 can remain applicable even when there is not a known GCO procedure for a given problem task, so long as there can exist a GCO procedure *in principle*. This is the case for more complex tasks such as making a good move in chess. Chess is a finite game and has a finite number of possible positions, and there will be in principle a GCO procedure to determine the best move(s) at a given point in a game. But there are far too many possibilities to compute for there to exist a known algorithm that would guarantee the best outcome. Heuristics are therefore the *only* viable procedures for chess (Richardson, 1998). Again, these heuristics do not guarantee the best outcome, or even a good outcome—one may lose the game in the end—but they are believed to generally result in good outcomes.

Notice, however, that the problems for which H_1 is usefully applicable are all *well-defined* problems. Well-defined problems are problems that have discrete states and goals that are, or can be in principle, known and understood. In other words, well-defined problems have well-defined problem-spaces.[8] Chess, tic-tac-toe, and problems in math and logic, all have well-defined problem-spaces. This feature is especially conducive to generating algorithms that guarantee correct outcomes. For not only are the goals of these problems understood and known *to be correct outcomes*—that is, we know precisely what a correct outcome is—but one can identify definite and discrete steps by which to navigate the problem space, getting from one state to the next, so as to ultimately reach the goal state. This is what is meant by a problem's having a GCO procedure *in principle*—if the states and goals of a problem are well-defined, then there will exist some procedure that will guarantee a correct outcome, even if this requires exhaustive search of the problem-space.[9]

Nevertheless, even when we are engaged in solving well-defined problems, it is almost always the case that the only alternative for us is to use heuristics in the sense of H_1. The reasons are epistemological: cognitive limitations or time constraints prevent us from carrying out a GCO procedure, or there is just no known correct solution or no known GCO procedure (notwithstanding that these may exist in principle, which is an ontological issue). Indeed, the well-defined problems for which there are known GCO procedures are very few as it is (Richardson, 1998). In this light, then, we can understand heuristics as procedures that do not

[8]A problem-space is an abstract representation of a problem consisting of discrete states, a goal state, and can include operators that determine how to get from one state to another. See below for further discussion.

[9]Of course, the goal must be possible. For example, if in a game of chess there is no possibility for black to win, then *ipso facto* there cannot be a correct outcome with respect to black winning the game. In these sorts of situations, the goal may be shifted from winning the game to, perhaps, forcing a stalemate. If a stalemate is also not possible, then there very well may not be a procedure that will guarantee a correct outcome for black. Indeed, it is uncertain what is meant by "correct outcome" in this situation.

guarantee correct solutions, but it seems to allow too much in. It is sort of like defining *terminal velocity* as "A velocity slower than that of light". Everything of interest with respect to terminal velocities (i.e., macro-sized objects) travels slower than the speed of light. Similarly, (nearly) everything of interest with respect to heuristic procedures (i.e., real-world cognition) does not guarantee correct solutions. Thus, on this reading, we come to an important problem for H_1: the widespread claim that much of human cognition is heuristic appears uninteresting if not trivially true, since much of our cognition cannot be anything but heuristic. I will call this the *triviality problem*.

What makes matters worse, real-world problems are rarely well-defined. Instead, we usually find *ill-defined* problems in the real-world, which have undefined or indeterminate states and/or goals; the problem-spaces for ill-defined problems therefore cannot be completely, or sometimes even adequately, known, understood, or characterized. Ill-defined problems include problems such as deciding whether to take a job offer, coming up with a feasible policy to reduce carbon emissions, and solving a murder case. Because of their very nature, ill-defined problems preclude the possibility of GCO procedures: The goals of these problems are often sufficiently undefined such that it is unknown what a "correct" outcome would be, and this *ipso facto* makes it impossible to generate any kind of process that would guarantee a correct outcome. Moreover, even if an ill-defined problem has a sufficiently defined goal, such as finding the murderer in a murder case, the means by which one would proceed to achieve this goal would still be uncertain since the mediating states can still remain indeterminate, or we would be unable to evaluate the steps taken. Solving a murder case thus has an ill-defined solution space since it is often difficult, if not impossible, to tell whether the work one does on the case brings one closer to or further from solving the case. It is important to note here that the matter is not epistemological like the concerns mentioned above of not knowing or not being able to perform GCO procedures. Rather, the matter here is ontological—there simply does not exist a GCO for many ill-defined problems.

Since it does not make much sense to speak of correct outcomes for many ill-defined problems, much less processes that guarantee correct outcomes, what sense can be made of the claim made by H_1, viz. that heuristics are procedures that do not guarantee a correct solution? If ill-defined problems do not allow for the evaluation of intermediate steps toward a goal, and therefore in principle cannot have GCO procedures, then any and all procedures for ill-defined problems are heuristic by definition. This is just another version of the triviality problem identified above. The situation would not be so bad if it were not the case that the majority of real-world problems humans face are ill-defined and characteristically lack certainty. However, we do not live in a tic-tac-toe world—for most of our problems, it is impossible to determine what would count as steps toward a correct outcome, and nothing is guaranteed to produce any given outcome. The strategies we employ when dealing with these real-world, ill-defined problems will necessarily be heuristic. And, in fact, the situation appears to be worse. For if there is no way to evaluate intermediate steps toward a goal, then no sense can be made of heuristics being satisfactory, reasonable, or "good enough" solutions. That is, we would not know what would make a heuristic solution satisfactory, reasonable, or "good enough" in ill-defined contexts.

A possibility worth considering is that a deeper functional analysis of heuristics might have it such that certain heuristics represent ill-defined problems as well-defined ones. That is, a heuristic process might simplify a complex, ill-defined problem to a manageable and well-defined problem. If this is the case, then it would make sense to claim that heuristics are processes that do not guarantee correct outcomes, for they would be working within the confines of well-defined problems for which there are in principle GCO procedures. This is certainly a possibility, and many contemporary decision researchers agree that much of human reasoning involves search and manipulation of problem-spaces (Dunbar, 1998).[10] However, problems with this functional analysis arise when we consider the fact that there are no such things as correct outcomes, and *ipso facto* no GCO procedures, for simplifying ill-defined problems or representing them as well-defined problems. Again, this is a fact about ill-defined

[10]Such contemporary research is inspired by Herbert Simon who understood heuristic processes for problem-solving as procedures that reduce the complexity of the problem. I discuss Simon's work in more detail below.

problems generally. That is, representing an ill-defined problem as a well-defined problem is itself an ill-defined problem. Thus, the H_1 definition is rendered effectively useless with respect to this functional analysis since there is no GCO algorithm with which to contrast heuristic procedures.

I suppose that it may be possible that some independent, non-heuristic process is responsible for representing ill-defined problems as well-defined problems, and then heuristics are employed within these well-defined problems (hence, maintaining the useful applicability of H_1). But the problem of simplifying ill-defined problems goes deeper. Given the detail and complexity of ill-defined problems, there will be any number of ways to carve them up into problem-spaces consisting of discrete states and goals. Consider, for instance, a simple, though ill-defined, problem of deciding whether to go for a picnic. What states should the problem-space consist of, and what would be the goal state? Should there only be two states, (i) *go for a picnic* and (ii) *not go for a picnic*? Or should the problem-space consist of (i) *go for a picnic*, (ii) *not go for a picnic, but go to a restaurant instead*, and (iii) *not go for a picnic, but stay in instead*? I need not belabour the issue here of how difficult it is to formulate an ill-defined problem as a well-defined one.[11] The point is that the difficulties in representing ill-defined problems as well-defined problems suggest that there is not a clear-cut case for simplifying ill-defined problems. This is not to say that we never simplify ill-defined problems; in fact, we probably do this often with certain circumscribed ill-defined problems. Of course, whether and to what extent people actually represent ill-defined problems as well-defined problems is an empirical issue. But unless evidence would show that representing ill-defined problems as well-defined ones is universal, or nearly so, there is still a large class of problems for which it makes it trivial to speak of heuristics as processes that do not guarantee correct outcomes. However, as I want to maintain and as certain researchers demonstrate, heuristics are interesting strategies within this large class of problems. Thus, the issue here is not really the triviality

[11] This is similar to the what is known as "the problem of small worlds", introduced by Savage (1954); see also Shafer (1986). The problem of small worlds has to do with whether expected utility assignments will maintain the same structure, and therefore dictate the same decisions, upon refinements of a decision-problem.

problem, but what I will call a *range problem*—H_1 does not apply sensibly to an interesting range of cases.

The range problem indicates that we must seek a different understanding of heuristics than what is given by H_1. This is not to say, however, that heuristics are not procedures that do not guarantee correct outcomes. As we shall see, there is a cogent understanding of "heuristic" from which this follows (i.e., as a corollary). But an adequate and robust characterization of "heuristic" must offer more than a negative definition with respect to GCO algorithms, and indeed more than what H_1 provides.

2.2.2 Heuristics vs. optimization

There is another negative definition of "heuristic" that is not as explicitly stated in the literature as H_1, but is implicitly suggested by some authors (e.g., Carruthers, 2006a; Fodor, 2008; Gigerenzer & Goldstein, 1996; Simon, 1957). It is this:

> **H_2** Although heuristics can provide certain benefits, they are procedures that do not optimize/maximize.

Most of what I said against defining heuristics as H_1 can be applied *mutatis mutandis* to defining heuristics as H_2. But some further remarks are in order.[12]

The idea that heuristics are to be contradistinguished from procedures that optimize/maximize (herein simply as optimize) is associated with the rational decision theory that was developed in the mid-twentieth century by economists and other theorists such as John von Neumann and Oskar Morgenstern (1944), and L. Jimmie Savage (1954). According to this theory, a rational agent will optimize, that is find and perform the best act available. The "best" act here traditionally refers to that which maximizes expected utility. If an agent is faced with a problem for

[12]It is important to keep in mind here a distinction between *optimal strategies* and *optimization strategies*. A heuristic may very well be an optimal strategy, given certain constraints. That is, a heuristic strategy may be optimal in the sense that it is the best that a system can do, given certain constraints (cf. Samuels et al., 2002); for example, maximizing the cost-benefit ratio of a process. An optimization strategy, on the other hand, specifies a procedure to be followed in order to fully optimize or maximize some unit of desire (typically expected utility; see below) regardless of constraints. Although a heuristic can arrive at an optimal outcome in the right circumstances, on the foregoing definition it cannot be an optimization strategy since under different constraints it will not be the best strategy to adopt.

which there is a GCO, optimizing will just be carrying out the GCO procedure.[13] And this will be the case even if there is no known GCO procedure (though there exists one in principle) or it is impossible to perform the GCO procedure. This is because the rational decision theory idealizes decision problems and agents, and as such real-life human agents are either believed to strive to act as an ideal agent, or nonetheless placed under normative demands to do so.

What makes defining heuristics as procedures that do not optimize different from defining them as procedures that do not guarantee correct outcomes—what makes H_2 different from H_1—has to do with the procedures from which they are being negatively contradistinguished. Optimization procedures do not aim at guaranteeing correct outcomes, but at maximizing some value, traditionally utility. Optimization procedures take into account problems for which there are in principle *no* GCOs or GCO procedures, and thereby provide richer decision methods. For problems that do not afford correct solutions, there may nonetheless be some unit of desire that one wants to optimize, and so there may be an optimization strategy that can be applied. The upshot, then, is that if it is only maintained that heuristics are procedures that do not guarantee correct outcomes (H_1), then an optimizing procedure can qualify as heuristic, so long as this is not a GCO procedure. But this possibility is blocked by H_2.

Understanding heuristics to be procedures that do not optimize is a view inspired by Herbert Simon's notion of *satisficing*. Satisficing procedures are procedures that do not aim to meet the standards of rational decision theory (or any theory of rationality that is governed by the dictates of logic and probability theory), but instead aim to meet some aspiration level that is in some sense "good enough" relative to the desires, interests, and goals of the agent. (More on this below.) As Simon initially made apparent, satisficing procedures reduce problem-spaces and make decisions or choices without complicated computations. Thus, as procedures that do not optimize, heuristics are supposed to pick out a more general class of procedures. More specifically, the class of procedures picked out is supposed to consist of those that generally conserve time and cognitive resources at the expense of an optimized outcome. As we saw

[13]This comes close to what Savage (1954) called "The Sure Thing Principle".

above, this is typically understood as a good trade-off—foregoing an optimized outcome is not despaired since what is saved in terms of time and cognitive resources make up for it, and one typically gets an outcome that is "good enough". Thus, H_2 essentially takes heuristics to be just satisficing procedures.

It is empirically well-established that humans generally do not optimize in their decision-making (Gigerenzer & Todd, 1999; Simon, 1956, 1990). Rather, humans are essentially satisficers by nature. There are, of course, evolutionary explanations for this: survival requires that decisions be made in an efficient and timely manner at the expense of optimizing. But more than this, optimizing places far too many demands on our computational resources to feasibly arrive at solutions, even if we are at leisure to devote time and cognition to making a decision. This point was made by Simon. Depending on the manner in which one tries to optimize, one might have to carry out a number of calculations such as those involved in subjective expected utility, multiple regression, and other weighted linear models. Not only are these calculations often very complex—too complex for the lone person to compute—but they may also be aggregately far too many to perform within reasonable time frames. Consider, for instance, a problem of buying a car. Optimizing may entail computing expected utility for *every* possible consequence of buying any car under consideration. This is quite a cumbersome task that cannot possibly be completed within a lifetime. One may try to ensure that only a manageably few calculations are required to be carried out by abstracting away from the details of the problem, so to speak, and stipulating that only certain consequences are relevant and worth computing utilities for. However, similar to the problem discussed above with respect to GCO procedures, it is often difficult to determine which consequences are indeed relevant and to what degree, and whether the decision based on what few consequences are believed relevant does in fact recapitulate *in abstracto* a decision based on the consideration of all consequences, as the theory demands.[14] What is more, optimizing would require that the costs associated with determining which consequences are relevant be themselves subject to a cost-benefit analysis. And this is

[14]This is Savage's "small worlds" problem. See footnote 11 above.

just the beginning of an infinite regress of calculations.

There are further reasons why optimization procedures are not generally used by humans. For example, there are notorious problems having to do with assigning probabilities and utilities to events. Humans do not have the cognitive wherewithal to form precise degrees of belief (i.e., subjective probabilities) as required to carry out many of the calculations for optimizing. It is also difficult to assign numerical utilities to objects or events. What is the numerical utility for buying a car? Or for getting married? Or for having your favourite cereal in the morning? And this leads to problems of comparisons. Gigerenzer and Todd (1999), for instance, argue that there are no common grounds—no "common currency"—by which we might judge and compare many of our objects of desire. They claim that love, friendship, and PhDs, for example, are "priceless", and therefore cannot be compared to the value of items for sale in a shopping mall (Todd & Gigerenzer, 2000). Gigerenzer and Todd therefore throw into doubt the very idea of expected utilities. Regardless of whether they are right about this, however, assigning numerical utilities is certainly a cognitively taxing task, if not impossible altogether.

Suffice it to say that humans typically do not employ optimizing strategies in their judgments and decision-making. But if this is the case, then the triviality problem cited against H_1 applies to H_2—defining heuristics as procedures that do not optimize trivializes the fact that much of our cognition is guided by heuristics. My general point is that H_2, like H_1, makes our general reliance on heuristics uninteresting, since all too often the only alternative is to use heuristics. To put it another way, H_2, like H_1, does not provide a contrast class that interestingly distinguishes heuristics from other types of inferential or problem-solving procedures in most situations. Certainly, there are situations in which such characterizations are usefully applicable, such as in certain economic or mathematical situations where there are known feasible optimizing strategies (or GCO procedures); heuristics can here be sensibly understood as procedures that do not optimize (or do not guarantee correct outcomes), but can be beneficial and lead to good or useful outcomes. However, it is not the norm for there to be optimizing (or GCO) procedures, and so the decision-problems of mathematics and economics do not represent the problems typically confronted in general cognition. And again, since general cognition is what we are interested in here, we therefore need a more robust understanding of heuristics for the general case.

2.2.3 Inherent normativity?

The characterizations that I have been considering are negative characterizations. However, what heuristics are defined not to be are what are (or were) believed to be normative procedures of reason. Both GCO and optimization procedures are supposed to be prescriptions for how to reason best, which are believed to ensure that either the correct or optimizing solution would be arrived at. Thus, characterized as a procedure that does not guarantee a correct outcome or that does not optimize, as according to H_1 and H_2 respectively, "heuristic" appears to be inherently normative.

To be sure, it seems perfectly fine to describe a procedure as heuristic without asserting anything normative (I will return to this point below). And we can even speak of procedures that do not guarantee correct outcomes or that do not optimize without asserting anything normative. Nevertheless, GCO and optimization procedures have a theory of rationality built in, as they pick out a class of procedures that is normatively appropriate.

GCO or optimization procedures are, of course, ideals. Ideals are not always meant to be achieved, but can serve simply as things to aim at or aspire to. And, as discussed, there are many instances where GCO or optimization procedures are not practically possible for humans (or sometimes even computers). Nevertheless, ideals are still normative, whether in the capacity of requirements to be fulfilled or in the capacity of things to aspire to. If heuristics are procedures that do not guarantee a correct outcome or do not optimize, they will neither fulfill the given normative requirements nor aspire to do so. Thus, H_1 and H_2 overtly cast heuristics as normatively inappropriate procedures, notwithstanding that some heuristic will often be the only alternative.

Contradistinguishing heuristics from GCO or optimization procedures may be a symptom of the hangover from taking algorithmic rules to dictate the normatively appropriate way to reason. Algorithmic rules are proven to necessitate the right answers, when applied correctly. Moreover, algorithmic rules are meant to be universal in the sense that they are domain-general and apply to every person in every situation. According to the traditional theory of rationality, these two features—necessity and universality—are believed to dictate the normativity of reasoning procedures (Brown, 1988). It is *because* algorithmic rules are necessary and universal that they are traditionally considered to constitute rationality. Heuristics, on the other hand, neither necessitate right answers nor are universal in the said sense. Instead, the success of heuristics is variable and depends on a number of contextual factors. Thus, heuristics do not embody the key features of rationality, and therefore they cannot be normatively appropriate procedures.

Yet there is something wrong with the idea that "heuristic" is inherently normative. Certainly there is a normative dimension to heuristics and their deployment. We can, for instance, evaluate whether a given heuristic should have been used, or whether it was used in the right way. But "heuristic" is not a normative concept *per se*. The normative appropriateness of employing a given heuristic is an issue that may be addressed only in concert with a normative account of reasoning and decision-making—i.e., a theory of rationality. In the absence of a theory of rationality, however, we can still discuss and analyze heuristics and their deployment in a descriptive manner. I therefore take this as motivation to divorce "heuristic" from the inherent normativity implied by the characterizations of H_1 and H_2. In effect, I believe that this is further reason to reject the negative characterizations of heuristics as procedures that do not guarantee correct outcomes or as procedures that do not optimize, or more generally, as procedures that do not meet some normative standard. "Heuristic" is not an inherently normative concept, and need not be characterized negatively. Positively characterizing "heuristic" as a kind of procedure used in reasoning, judgment, inference, and decision-making will be my next task.

2.3 Uses of "heuristic"

What has been discussed so far should not be taken to imply that no one in the literature ever gives some definition of "heuristic" as they intend to employ the term beyond perfunctory remarks. Some authors attempt to define what they take "heuristic" to mean, while others include discussion from which we might infer their intended meaning. These characterizations differ from those considered so far in that they are positive definitions, unlike the negative definitions of H_1 and H_2. There are some key researchers who have offered various positive characterizations of "heuristic", from which we can gain an initial grasp of the term. I will review these here.

Let us begin by briefly considering one example of an attempt at positively defining "heuristic", which illustrates the difficulty of capturing the nature of heuristics. In a book entitled *Heuristics*, Judea Pearl (1984) expounds and discusses several heuristics used in computation and computer science. Throughout the entire book, however, there are only two brief passages in which he attempts to characterize heuristics:

> This book is about *heuristics*, popularly known as rules of thumb, educated guesses, intuitive judgments or simply *common sense*. In more precise terms, heuristics stand for strategies using readily accessible though loosely applicable information to control problem-solving processes in human beings and machine[s]. (p. vii)

> Heuristics are criteria, methods, or principles for deciding which among several alternative courses of action promises to be the most effective in order to achieve some goal. They represent compromises between two requirements: the need to make such criteria simple and, at the same time, the desire to see them discriminate between good and bad choices. (p. 3)

There are a few problems with these characterizations. First, pointing out synonyms provides little insight if the synonyms themselves are just as vague. Claiming that heuristics are "rules of thumb, educated guesses, intuitive judgments or simply *common sense*" gives no clarification since each of these terms is likewise undefined and unclear (although we will see below how "rules of thumb" can offer insight into the nature of heuristics when closely analyzed). It is also unclear what "readily accessible though loosely applicable information" means, or what "to control problem-solving processes" entails. Pearl does not comment on these remarks at any other point in his book. Finally, with respect to the second passage, although heuristics can sometimes be employed to make decisions "among several alternative courses of action", there are certainly heuristics that perform other functions, such as making inferences (examples of which we shall see below). It therefore seems as if Pearl does not fully or adequately capture the nature of heuristics. This is made apparent by Pearl's example of a heuristic that can be used in common life:

> (1) To choose a ripe cantaloupe, press the spot on the candidate cantaloupe where it was attached to the plant and smell it; if the spot smells like the inside of a cantaloupe, it's probably ripe. (p. 3)

In terms of alternative courses of action, however, it would seem that the "most effective", and perhaps most "simple", way of choosing a ripe cantaloupe is to cut it open, and perhaps taste it. Of course this is not a practical or realistic solution when purchasing a cantaloupe from a grocery store. Maybe there will be trade-offs and compromises between what is desired and what is practical, which may very well be a crucial aspect of heuristic strategies, but Pearl does not mention these; he only mentions the compromise heuristics represent between simplicity and success.

Pearl's attempt at defining heuristics illustrates not only the difficulty in capturing the nature of heuristics, but also a common lack of concern among researchers to adequately characterize heuristic processes. Many instead rely on intuitive understandings. This may serve us well for empirical study of the kinds of processes humans engage in, but it will not do for the philosophical project of developing a positive characterization that is coherent and robust enough to account for a wide range of phenomena, as well as to find its place within the conceptual structure of cognitive science (see section 2.1 above).

I will now proceed to present other, more influential characterizations through a short semi-historical account of the meaning of "heuristic". I trace its meaning by focusing mainly on four research programs. As we shall see, the way in which "heuristic" is used varies between programs and researchers, and yet there are some basic connections between them all. After expounding these disparate but connected accounts of "heuristic", I will explain some of the connections between them as I develop a more precise conception which is meant to apply to general cognition, and which will figure in the rest of this dissertation.

2.3.1 Processes of discovery: Pólya

The term "heuristic" has its roots in the Greek word *heuriskein*, which means "to find" or "to discover" (Moustakas, 1990). Throughout history, however, different meanings and connotations have been attributed to the word. George Pólya (1957) was perhaps the first to attempt to develop a coherent account of "heuristic".[15] His interest was in the methods and strategies in solving mathematical problems, although he indicated a wider applicability of and interest in heuristic[16] to include other areas of study such as education, logic, psychology, and philosophy. In the same vein of the word's etymology, Pólya claimed that "The aim of heuristic is to study the methods and rules of discovery and invention" (p. 112). At the same time, however, he believed that a proper notion of "heuristic" is really concerned with the *process* of problem-solving, and so Pólya maintained that heuristic seeks to investigate the psychological processes of discovery and invention. Pólya therefore defined what he called "modern heuristic" as the endeavour to understand "the *mental operations* typically useful in solving problems" (p. 130, emphasis added).

It is important to notice here what might be considered nowadays to be an idiosyncratic use of "heuristic". Pólya employs the term to refer to a branch of study or a kind of research program, rather than to the (mental) processes or operations themselves that are to be studied; that is, rather than the means by which problems are solved. This understanding of "heuristic" is distinct from its common use as an adjective to describe certain kinds of (mental) processes or operations. To be sure, Pólya spoke of "heuristic reasoning" which is supposed to refer to merely provisional, though plausible, reasoning, as opposed to rigorous proof that is both

[15]Though see Pólya's remarks about Bernard Bolzano (pp. 57-58).

[16]Notice here Pólya's special use of the word "heuristic": he uses the term in the singular to describe a research program. More on this below.

certain and final. Yet heuristic reasoning is not what Pólya had in mind when he spoke of "heuristic" *simpliciter*. These observations will be useful for the discussion below.

In any event, Pólya devised a number of techniques to aid in problem-solving, many of which may be characterized in contemporary parlance as heuristics *qua* means by which problems are solved. Indeed, many contemporary authors refer to just such techniques when discussing heuristics (e.g., Michalewicz & Fogel, 2000). Some of Pólya's techniques include:

(2) Draw a diagram when trying to solve a problem.
(3) If you can't solve a problem right away, try an indirect proof.
(4) Try to restate the problem in different terms.
(5) Assume you have a solution and work backwards.
(6) Find a related problem that has been solved before, and try to use its result or method.

2.3.2 Information processing theory: Simon

Interestingly, around the same time Pólya published his account of heuristic, a different but very influential notion began to emerge. It was then that research in computer programming was quickly advancing, and a task that occupied many researchers was programming a computer to simulate human intelligence—this was the birth of information-processing theory and the research program of AI. Of course, optimization strategies such as exhaustive search and maximizing utility, though theoretically sound, were for the most part impractically burdensome in terms of time and other computational resources. If the chief interest in AI is *actual* rather than *in principle* or *theoretical* computing, then it may very well be irrelevant to employ optimizing strategies, or to investigate whether such strategies even exist (Simon et al., 1967). Thus, what was needed were strategies that would cut down on the computational burden that come with optimization strategies, but that would also perform relatively well by finding satisfactory solutions, although perhaps not providing the best solutions as optimization strategies do. What was needed, in other words, were heuristic strategies.

Perhaps the most notable of pioneers of early AI is Herbert Simon (although his then-colleague, Allen Newell, may be as famous). Simon exalted the employment of heuristics in computing. He believed that the ability of heuristics to find solutions rapidly with relatively little computational costs not only made their use indispensable to computer programming, but made most of them better than optimizing strategies, even when the latter are available and practicable. In fact, Simon formulated the term "heuristic power" as a relative measure of a computer program's capacity to find solutions in "reasonable" computation times (Simon et al., 1967). More importantly, however, he believed that "we can use this theory [of heuristic computational problem solving] both to understand human heuristic processes and to simulate such processes with digital computers" (Simon & Newell, 1958, p. 6). That is, Simon observed that there are mutual ramifications of practical computing or AI on the one hand, and human decision-making, or cognition more generally, on the other.

Arguably, the most important contribution Simon made to the study of human decision-making was his notion of *bounded rationality*. Simon despaired of the formal model of rationality of his time which prescribed optimization strategies, including maximizing expected utility and abiding by the basic axioms of probability. Just as with the demands it puts on computation systems, optimization, as he made clear, requires agents to know more than a human can ever know (e.g., all the possible states of a problem and possible outcomes) and do more than a human can ever do (e.g., perform numerous calculations over the sets of possibilities and outcomes). Bounded rationality, however, takes into account empirical limits on human cognition, and offers a more plausible and psychologically realistic notion of rationality. According to Simon (1957), what is fundamental to bounded rationality are special simplifications of complex problems in order to render them realistically solvable by humans. Rather than maximize or optimize, agents can employ procedures that exploit the simplifications to solve the problem in a manner that is in some sense "good enough", or *satisficing* (pp. 204-205). As the lessons from computer programming indicate, satisficing procedures are commonly understood to be heuristics.

Simon endeavoured to clarify the term "heuristic", since he thought that, as it is commonly used, it "denotes an intuitive notion usually misunderstood when oversimplified" (Simon et al., 1967, p. 10). And yet the notion to which the term refers can be quite complicated (p. 18). He explicitly rejects the characterization of a heuristic as "a technique that yields improved overall performance at the expense of failure on some problems", claiming that "the usage of 'heuristic' does not in itself imply any *necessity* of failure on some problems" (p. 11, emphasis in original; cf. my remarks in section 2.2.1 of the present chapter). Simon goes on to note that many heuristics merely adjust the way a problem-space is searched thereby making it easier to find a solution, or by eliminating possibilities that are guaranteed not to be solutions. He offers instead the following as a definition: "Any component of a program that contributes to its heuristic power we may call a *heuristic process*, or a *heuristic*" (p. 17), where heuristic power is as characterized above, viz. the capacity to find solutions within certain time-frames.[17]

However, it is important to note that, although Simon defines "heuristic" squarely from the perspective of computation, he expressly intended heuristics to play a central role in discovery (Kulkarni & Simon, 1988; Langley, Bradshaw, Simon, & Zytgow, 1987; Simon, 1973, 1977). That is, he did not disavow "heuristic" of its etymological roots. Indeed, since Simon vehemently believed that human psychological processes can be modeled by computers, heuristics take centre-stage in accounting for how humans discover (scientific) theories or, more generally, discover solutions to problems. Hence, he comments, "if 'heuristic' is to be used as a noun, it is best employed as a synonym for 'heuristic process'—a process that aids in the discovery of a problem solution" (Simon et al., 1967, p. 17). As processes for discovery, the following are examples of heuristics given by Simon:

(7) If S_1 is your current state, S_2 is your goal state, d a difference between S_1 and S_2, and O an operator that, in past experience, affects differences of the type to which d belongs, try to apply O to transform S_1. (Simon et al., 1967, p. 17)

(8) If the problem is to prove that two triangles are equal, and you are given information about the equality of angles, search for a theorem whose premises refer to equality of angles and whose conclusion refers to the equality of triangles. (ibid.)

[17] More precisely, "A heuristic is an addition, rearrangement, or other modification to a program that yields an improvement in its heuristic power" (Simon et al., 1967, p. 18). Simon offers this alternative formulation since he notes that it is usually difficult to identify which parts of a program contribute to its heuristic power. But for our purposes, the simplified formulation given in the main text will suffice.

(9) In chess, consider first those moves that remove or protect pieces under attack. (ibid.)

(10) To detect regularities in numerical data, if the values of two numerical terms increase together, then consider their ratio. (Langley et al., 1987, p. 66)[18]

2.3.3 Heuristics and biases: Kahneman and Tversky

Simon's work had a significant influence on psychologists Daniel Kahneman and Amos Tversky (Gilovich & Griffin, 2002; Kahneman, Slovic, & Tversky, 1982b). In the 1970s, Kahneman and Tversky began studying people's judgments under uncertainty, and thereby established a research program that has come to be known as "heuristics and biases". Though Simon had rejected the formal model of rationality as psychologically unrealistic due to the computational demands it imposes, Kahneman and Tversky were interested in the extent to which people follow the elementary rules of probability in their reasoning, and the psychological implications that follow. Through their research they discovered that people tend to rely on a limited number of "heuristics that generally govern judgment and inference" (Tversky & Kahneman, 1983, p. 313). Furthermore, they demonstrate that these heuristics are generally "quite useful" (ibid.) and "sometimes yield reasonable judgments" (Kahneman & Tversky, 1973b, p. 237), notwithstanding that they can lead to judgments that depart from the basic axioms of probability and result in systematic, predictable biases.

Although the heuristics and biases program has had considerable influence in contemporary psychological research, precisely defining the cognitive processes of interest—i.e., precisely defining "heuristic"—has evaded Kahneman and Tversky, as well as those who follow them. They do, however, offer brief definitions. In their initial work they claim, "heuristic principles … reduce the complex tasks of assessing probabilities and predicting values to simpler judgmental operations" (Tversky & Kahneman, 1974, p. 1124). Almost ten years later, they asserted, "The term *judgmental heuristic* refers to a strategy—whether deliberate or not—that relies on a natural assessment to produce an estimation or a prediction" (Tversky & Kahne-

[18] This heuristic purportedly aided Simon and Newell's computer program in the "discovery" of Kepler's Third Law.

man, 1983, p. 294). Unsurprisingly, Kahneman and Tversky are not entirely clear on what they mean by "natural assessment"; they offer only the following: "*natural assessments* ... are routinely carried out as part of the perception of events and the comprehension of messages. Such natural assessments include computations of similarity and representativeness, attributions of causality, and evaluations of the availability of associations and exemplars" (Tversky & Kahneman, 1983, p. 294). Throughout the body of written work produced by the heuristics and biases program, the term "heuristic" is used as a noun referring vaguely to the said principles or strategies (although the reliance on natural assessments seem to have been largely neglected; Gilovich & Griffin, 2002).

We might note here three things. First, Kahneman and Tversky's characterization of heuristics (or heuristic processes; I will be using these terms interchangeably) moves away from Pólya's account and from the word's etymology more generally. It seems Kahneman and Tversky are not strictly interested in the processes of discovery. Rather, they are interested more specifically in certain specific processes of judgment, inference, and prediction, and a part of their research program is to specify the conditions under which such processes depart from the basic axioms of probability theory.[19]

The second thing to note is that Kahneman and Tversky's characterization agrees with Simon's definition of "heuristic" insofar as both indicate that heuristics simplify complex tasks to enable simpler operations. And although Kahneman and Tversky are not explicit about computational implications, simplifying complex tasks to enable simpler operations reduces computational burden, and this is precisely Simon's central feature of heuristic processes.

However, the third thing to note is that, unlike Simon, Kahneman and Tversky are not concerned about heuristics *qua* computational operations. Instead, the heuristics they are interested in are restricted to humans. Although these heuristics are proclaimed to be cognitive processes, Kahneman and Tversky make their claims in the absence of a theory of mind, and

[19]In the words of Gilovich and Griffin (2002), the heuristics and biases research program has two agendas, one positive and one negative. The positive agenda is to elucidate the processes that people rely on to make intuitive judgments; the negative agenda is to elucidate the conditions under which intuitive judgments will likely disagree with the rules of probability.

are therefore not committed to the position that cognition is computation in the sense of CTM. Hence, their account of heuristics (and natural assessments) are not, and need not be, specified as classical computations.

Kahneman and Tversky's work initiated a deluge of research devoted to investigating the inferential and judgmental processes employed by humans. Subsequent work in the heuristics and biases tradition has established the pervasiveness of the heuristics hypothesized by Kahneman and Tversky. Indeed, it has been convincingly shown that people generally are guided by such heuristics regardless of incentive or motivation, and regardless of increased attention and effort by devoting their full cognitive resources to the task (Gilovich & Griffin, 2002).[20] The three heuristics most studied by Kahneman and Tversky are:

(11) Probabilities are evaluated by the degree to which one thing or event is representative of (resembles) another; the higher the representativeness (resemblance) the higher the probability estimation. (Representativeness)

(12) The frequency of a class or the probability of an event is assessed according to the ease with which instances or associations can be brought to mind. (Availability)

(13) Estimates are made by starting with an initial value (the "anchor") and making adjustments to it; the anchor can be influenced arbitrarily through suggestive techniques. (Anchoring and Adjustment) (Tversky & Kahneman, 1974)

2.3.4 Fast and frugal heuristics: Gigerenzer

Gerd Gigerenzer has also been influenced by Simon's seminal works, and Gigerenzer also propounds the ubiquity of heuristics in human cognition. However, Gigerenzer accuses Kahneman and Tversky of offering "nebulous", "one-word explanations" to be passed off as heuristics (Gigerenzer, 1996; Gigerenzer & Todd, 1999). Gigerenzer argues furthermore that Kahneman and Tversky's proposed heuristics at once account for everything and nothing: they explain everything because they are so vague that any one of them "can be fit to almost any empirical result post hoc" (Gigerenzer & Todd, 1999, p. 28); they explain nothing because Kahneman

[20]This is not to say that anyone in the heuristics and biases tradition has suggested that people always rely on heuristics and never abide by the rules of probability. Rather, what has been shown is that, motivation, attention, effort, etc., can decrease the reliance on heuristics, and thereby decrease biasing effects. But such biasing effects never completely disappear.

and Tversky do not specify any underlying cognitive processes.[21] In return, Gigerenzer offers a more refined and synthesized account of "heuristic". He takes his cue from the word's etymology as processes for discovery, where "heuristic" refers to "useful, even indispensable cognitive processes for solving problems that *cannot* be handled by logic and probability theory" (Gigerenzer & Todd, 1999, p. 25). But Gigerenzer is also inspired by Simon and Newell's models of heuristic processes for guiding search through problem-spaces. In this respect, the term refers to "a useful shortcut, an approximation, or a rule of thumb for guiding search" (p. 26), and of which precise computational models can be made. Building upon these traditions, Gigerenzer (and his colleagues) established the "fast and frugal heuristics" research program.

Gigerenzer explains that a fast and frugal heuristic is generally a rule according to which an organism reasons and acts (Gigerenzer, 2004, 2006, 2008b); they are fast and frugal because they operate quickly and with little information. But such a rule is not a heuristic unless it embodies two fundamental features: First, heuristics must *exploit the evolved capacities of an organism*, such as those of perception or memory (cf. Kahneman and Tversky's notion of natural assessments).[22] Gigerenzer believes that it is this feature that allows heuristics to be simple, and that simplicity results in "*fast*, *frugal*, *transparent*, and *robust* judgments" (Gigerenzer, 2004, p. 64). Second, heuristics must *exploit structures of environments*.[23] What is meant by structures of environments is the structures of information found in environments. The idea is that the success of heuristics partially lies in the way information in the environment is structured, and how heuristics use such structures to their advantage. This might consist

[21]Kahneman and Tversky have replied by claiming that "This objection misses the point that [our heuristics] can be assessed experimentally; hence it need not be defined a priori" (Kahneman & Tversky, 1996, p. 585). I believe that this is an important point, since heuristics are not the kind of thing that should be defined or analyzed a priori. Rather a more sensible, and interesting, approach is to give an empirically informed account of heuristics—to see what processes are actually employed in human cognition, and then determining which of them we want to call heuristic. This is the approach that I attempt below.

[22]In more recent works, Gigerenzer expresses his belief that heuristics also exploit learned capacities. He states, for instance, that "a heuristic exploits hard-wired *or learned* cognitive and motor processes" (2006, p. 121; 2008b, p. 23; my emphasis). This is a more encompassing view, but I believe that it leads to a trivial account of how heuristics operate. For if a heuristic is conceived to exploit evolved or learned capacities—whether they be cognitive or motor processes—there would seem to be nothing left to distinguish capacities that heuristics exploit from those they do not, for it appears as if evolved or learned capacities exhaust the possibilities.

[23]Gigerenzer (2008a) refers to exploiting evolved or learned capacities and exploiting environmental structures as *embodiment* and *situatedness*, respectively.

in recognizing how certain information is distributed in the environment, or seeing how certain information is correlated with other information. Heuristics attuned to such structures will therefore make good predictions. Exploiting environmental structures contributes to the robustness of heuristics, since different environments can share a similar informational structure, and a heuristic that can exploit such a structure can therefore be successful in either environment.

The fast and frugal heuristics program takes both descriptive and normative analyses of heuristics as essential. The program adopts a normative "analysis of heuristic accuracy, speed, and frugality in real-world environments" (Gigerenzer & Todd, 1999, p. 29). It may be observed that this is similar to, and appears to be an adaptation of, Simon's (descriptive) notion of heuristic power. The normative analysis given by the fast and frugal heuristics program consists in what Gigerenzer calls ecological rationality.[24] According to Gigerenzer, ecological rationality eschews the standard normative assumption "that formal axioms and rules of choice can define rational behavior without referring to factors external to choice behavior" (Gigerenzer, 2000, p. 202). Hence, Gigerenzer attacks the interest of the heuristics and biases approach in studying how the use of heuristics yields judgments that disagree with the basic rules of probability and lead to biases. What is of better interest, according to ecological rationality, is studying the environments in which given heuristics will succeed or fail. What follows is that one cannot speak of a given heuristic being universally ecologically rational: "Ecological rationality implies that a heuristic is not good or bad, rational or irrational per se, but only relative to an environment" (Gigerenzer, 2006, p. 121). Given this characterization, Gigerenzer argues that once we find the environments in which given heuristics perform well, employing heuristics constitutes good reasoning and that their use is justified because they result in adaptively useful outcomes.

For a descriptive analysis, the fast and frugal heuristics program aims to "specify the function and role of fast and frugal heuristics more precisely than has been done in the past, by building computational models with specific principles of information search, stopping, and

[24]The notion of ecological rationality is inspired by the ideas of both Simon and Egon Brunswik.

decision making" (Gigerenzer & Todd, 1999, p. 29). Such computational modeling makes transparent the mechanics of heuristics and the ways in which information (in the environment) is used. This facilitates the investigation of ecological rationality. The computational modeling is made apparent by looking at the heuristics that Gigerenzer commonly discusses:

(14) (i) Search among alternatives; (ii) stop search when one alternative is recognized; (iii) choose the recognized object. (Recognition Heuristic)

(15) (i) Search among alternatives; (ii) stop search when one alternative scores higher on some predetermined criterion; (iii) choose the object identified in step ii. (Take the Best)

(16) Use the cue that discriminated on the most recent problem to choose an object; if the cue does not discriminate use the cue that discriminated the time before last; and so on. (Take the Last)

(17) (i) Fixate your eyes on the moving object; (ii) start running; (iii) adjust your running speed such that the angle of your gaze remains constant. (Gaze Heuristic)

The fast and frugal heuristics program represents probably the most earnest attempt to offer a detailed and precise account of heuristics. However, the extent to which their account is useful as an appropriate and adequate characterization of "heuristic" will be assessed below. For now, let us observe that the heuristics of the fast and frugal program are not as general as the ones Pólya and Kahneman and Tversky introduced, but at the same time they are not as specific as the ones Simon offered. Rather, the fast and frugal heuristics are domain-specific in the sense that they are meant to be deployed in certain circumscribed domains. As we shall later see, this feature has important implications for cognitive architecture.

2.3.5 Other uses

In other areas of the literature, we find different and less developed accounts of heuristics. Most of these have roots in the manner in which they are characterized by the heuristics and biases tradition. Nisbett, Krantz, Jepson, and Kunda (1983), for instance, raise some concerns similar to those raised by Gigerenzer. They accept Kahneman and Tversky's position that the employment of heuristics can lead to judgments that depart from the basic rules of probability. However, they also argue that people often use heuristics that produce intuitions that agree with probability theory. Nisbett et al. therefore label the heuristics of the heuristics and biases tradition "nonstatistical heuristics", and distinguish them from "statistical heuristics", which are "intuitive, rule-of-thumb inferential procedures that resemble formal statistical procedures" (p. 345). Other than this sentence, however, Nisbett et al. provide no fuller characterization of statistical heuristics.

Another use of the term "heuristic" worth mentioning is Jonathan St.B.T. Evans', which is idiosyncratic. Evans is a pioneer of the contemporary *dual-process theory*. Dual-process theory hypothesizes that cognitive processes come in two kinds. On the one hand, some of our cognition, sometimes called *type 1*, is unconscious or preconscious, automatic, fast, parallel, high capacity, and operates with implicit knowledge. On the other hand, some of our cognition, sometimes called *type 2*, is conscious, controlled, slow, serial, low capacity, and operates with explicit knowledge. (For extensive discussion see J. Evans & Frankish, 2009). Evans holds the view that type 1 processes are "heuristic processes". What he means by this, however, is not precisely what anyone who we have discussed so far meant by the term. According to Evans, heuristic processes "focus our attention on selective aspects of presented information and rapidly retrieve and apply relevant prior knowledge and belief" (J. Evans, 2006, p. 392; cf. J. Evans & Over, 1996; J. Evans, 2009). Although such a role can be circumscribed for some of the characterizations of "heuristic" given above, it would take some work and might appear unnatural. Unfortunately, Evans leaves the details of such heuristic processes largely unspecified.

Some researchers who employ the term "heuristic" are not concerned with the higher cognitive processes of thought and reason, but rather with the low-level cognition that underwrites perception. There is a camp of perceptual psychologists, for instance, who understand and model visual perception as an inferential process (Braunstein, 1994). The idea is that the eye is usually presented with visual information that is either incomplete or very complex, and all but the simplest events require that many dimensions of information be integrated to produce a complete and coherent scene. Such an integration is too demanding of the human visual system, so it has to make "judgments", so to speak, about patterns, shapes, and so on. Since visual perception is thus conceived as judgmental in nature, a line of argument pursued is that certain operations of the visual system are heuristic in the sense of producing systematic outcomes that can result in error (e.g., J. Feldman, 1999; Kaiser, 1998).[25] Perceptual linguists similarly invoke heuristics to explain various operations involved in language perception (e.g., Fodor, Bever, & Garrett, 1974; Frazier & Fodor, 1978; Pericliev, 1990). The same reasoning of the visual perceptual psychologists applies here. Indeed, all perceptual cognition, according to this view, involves heuristics in one way or another. However, a deeper account of such heuristic processes is not generally pursued.

2.4 A positive characterization

Recall that in section 2.2 I argued that the two most common perfunctory characterizations of "heuristic" are at best trivial or uninteresting, and at worst inapplicable in the face of real-world problems and the characteristics of actual human cognition. It should not come as news that there are not too many real-world problems that admit GCO or optimization procedures. In fact, this is a rather boring truth. But calling strategies "heuristic" just because they do not generate correct or optimal solutions does not make the situation any more interesting. Since every problem-solving or inferential procedure is heuristic in real-world contexts, there seems to be no point at all in invoking the term "heuristic" to refer to these procedures, for any and all of them just are, simply and plainly, cognitive procedures applied in the real-world. It thereby becomes trivially true that we rely on heuristics, and it is empirically vacuous to state as much. These problems are what I have called the range problem and the triviality problem,

[25]Perceptual psychologists often assume that the perceptual system attempts to maintain veridical representations of the world; and so those who sit in the perceptual heuristics camp believe that heuristics approximate, as close as they can, veridicality. However, there is a compelling case for the position that the aim of perception is not always to provide accurate representations. According to Christopher Viger (2006b), "The aim of perception is to help guide action via planning and directly responding to a dynamic and often hostile environment; sometimes that requires accurate representations and ... sometimes it does not" (p. 275). In cases when our perceptions are inaccurate, "our perceptual systems [can] make salient important environmental differences, thereby sacrificing accuracy for usefulness" (p. 279). If this is right, then perceptual heuristics do not necessarily approximate reality or veridicality, but can serve as economical processes for making important environmental differences salient.

respectively. Moreover, I have suggested that contradistinguishing heuristics from GCO or optimizing procedures implies an inherent normativity that should be resisted, and that the negative characterization of heuristics as procedures that do not guarantee correct outcomes, or as procedures that do not optimize, should be rejected. If we are to have a robust and usefully applicable notion of "heuristic", and a notion that applies to general cognition, we are going to have to say more about what heuristics *are*, as opposed to what they *are not*.

Now that we have surveyed a range of uses of "heuristic", we are in a position to characterize more precisely the nature of the concept. There are certainly many other instances than what I have given here where "heuristic" is invoked to describe some aspect of cognition. But I believe that this initial exposition gives a sufficient feel for the various and disparate ways that the term has been and is employed.

Before we get started, let us briefly look at some of the heuristics to which we have been introduced. If we examine some of them, we find that it can be very difficult to identify common characteristics. Consider for instance

(4) Try to restate the problem in different terms,

and

(14) (i) Search among alternatives; (ii) stop search when one alternative is recognized; (iii) choose the recognized object.

(4) commands someone to do something rather vague and uncertain—*how* or *in what other terms* is one to restate the problem? And even after one has done as (4) commands, it is still uncertain whether one will have a favourable outcome, or indeed whether one will be able to solve the problem at all in such other terms. On the other hand, (14) lays out what is to be done in definite and discrete steps—no doubt an artefact of its being a computational model. Moreover, a definite outcome is guaranteed by (14), so long as one comes across an alternative that is recognized. And whereas (4) can be applied to just about any problem whatsoever, (14)

can be applied only when the problem consists in choosing an object among alternatives. Thus, these examples of heuristics do not appear to be of a piece. Even considering Gigerenzer's heuristics on their own, they do not seem to pick out a natural kind—(14)-(16) appear to be cognitive heuristics while (17) appears to be a heuristic for perception and motor function. It therefore remains to be seen what properties, if any, are shared by the given examples that make them heuristics.

The remainder of this chapter will be devoted to drawing out some fundamental similarities and differences among the various accounts of heuristics we have thus far considered. In undertaking this project, I will make a number of distinctions. These distinctions will serve to classify different kinds of heuristics. They will also serve to hone in on a cluster of properties from which we might establish a sensible and robust characterization of "heuristic", and thereby better understand the cognitive processes that are of concern to this dissertation.

2.4.1 Some distinctions

2.4.1.1 Stimulus-response vs. heuristics

The first distinction I will draw is between *stimulus-response* behaviour and *heuristic-produced* behaviour.[26] By stimulus-response behaviour I do not intend to refer to the behaviourist idea that takes the mind to be a black box to remain unanalyzed. Rather, I mean to refer to some unspecified theory that takes a stimulus to be the sole cause, in a course-grained way, of a response. By heuristic-produced behaviour I mean behaviour (including cognitive behaviour) that is brought about or caused by the operation of some heuristic. For example, consider any of the proposed heuristics mentioned above, such as:

[26] It may also be worthwhile to distinguish between stimulus-response (SR) and stimulus-stimulus (SS) theories of classical conditioning. According to SR, classical conditioning occurs via association and without conceptualization. According to SS, conditioning involves a conceptualization of the object of association. On the SS model, Pavolv's dog salivates because the bell evokes the concept FOOD. On the SR model, no such conceptualization occurs—salivating is simply associated with the ringing of the bell. Whether the SR or SS theory is correct is still debated. I suspect that both may be true, depending on the context and organism in question. To avoid complications, however, my remarks can be understood to refer to the SR theory.

(9) In chess, consider first those moves that remove or protect pieces under attack.

(11) Probabilities are evaluated by the degree to which one thing or event is representative of (resembles) another; the higher the representativeness (resemblance) the higher the probability estimation.

These heuristics are not mere stimulus-response operations. Rather, they are relatively more complex operations involving conceptual information. In contrast, stimulus-response behaviour does not implicate any conceptual information, since the stimulus, not anything having to do with concepts, is supposed to be the cause of the response. As a paradigm example, we might take a case of shuttering upon hearing an unexpected loud noise—the loud noise is not conceptualized, and thus concepts have no role in producing the shutter response; it is simply a brute reaction or reflex.

The distinction between heuristic-produced and stimulus-response behaviour may seem obvious. It is an important distinction nonetheless, especially given some recent work of Gigerenzer's. Gigerenzer (2007, 2008c) has speculated that some nonhuman animal behaviour can be understood in terms of heuristics. According to Hutchinson and Gigerenzer (2005), behavioural biologists have studied "rules of thumb" since the middle of the twentieth century. Rules of thumb describe some animal behaviour through simple rules which appear to be largely innate. So, for instance, the manner in which the wasp *Polistes dominulus* constructs its nest dictates that there are a possible 155 different hexagonal arrangements. However, only 18 arrangements are ever observed. And the 18 arrangements appear to follow a rule of thumb "in which the wasp places each cell at the site where the sum of the ages of the existing walls is greatest" (p. 103). Similarly, if there are two light sources in a copepod's environment, it will follow "a trajectory as if it were pulled toward each source with a force proportional to $\frac{\text{source intensity}}{\text{distance}^2}$" (ibid.). Yet behavioural biologists explain this seemingly complex behaviour by a simple rule of thumb that says that the copepod "adjusts its orientation so as to maximise the amount of light falling on [its] flat eye" (ibid.). Hutchinson and Gigerenzer claim that such rules of thumb "correspond roughly" (p. 98) to heuristics, though Gigerenzer (2007, 2008c) appears to make the stronger claim that "rule of thumb" is synonymous with "heuristic".

Gigerenzer (2007) goes on to explain that many other animal behaviours can be understood as being the result of following rules of thumb or heuristics.[27]

However, given the distinction made here between stimulus-response and heuristic-produced behaviour, Gigerenzer's belief that such animal behaviours are the result of following heuristics may be misguided. This certainly appears to be the case with respect to the copepod example—it is more likely that copepods are guided by stimulus-response mechanisms rather than heuristics; it is doubtful that they have the wherewithal to possess or develop any concepts.[28] It seems, rather, that stimulus-response behaviour is best characterized as merely satisfying a rule, or conforming behaviour to a rule. As Gigerenzer's examples illustrate, satisfying a rule may give the appearance that the behaviour is produced by a heuristic, but such stimulus-response behaviours do not involve involve a rule as a causal ingredient. We might say that merely satisfying a rule is not sufficient for heuristic-produced behaviour. (More on satisfying rules below.)

To avoid mere rule-satisfying behaviour as being considered heuristic-produced behaviour, I will offer the following constraint on what heuristics are supposed to be:

C: Heuristics (in some way) utilize conceptual information.[29]

I do not wish to put any stronger constraints on what heuristics are supposed to be at the moment. Nevertheless, with this weak constraint we can rule out stimulus-response behaviour as being heuristic-produced insofar as the former is an instance of merely satisfying a rule, where rule-satisfying behaviour does not implicate the involvement of any conceptual information. As we shall see presently, however, condition C has a bigger role to play in developing a robust characterization of heuristics.

[27]For example, Gigerenzer believes that birds of paradise mate selection behaviour can be explained as the females following the heuristic *Look over a sample of males, and go for the one with the longest tail*; Gigerenzer also claims that aggressor-assessment of deer stags follow a pattern of sequential reasoning, which can be described by a heuristic to the effect of *Use deepness of roar to estimate the size of rival first; use appearance afterward*.

[28]Many of Gigerenzer's other assertions about animals following heuristics (see previous footnote) are likewise thrown into doubt, for the behaviour in question is probably the result of stimulus-response.

[29]I hedge by saying "in some way" because, as we shall see in chapter 4, heuristics do not directly utilize conceptual information but metainformation embodied by relations within and between conceptual content.

2.4.1.2 Computational vs. cognitive heuristics

Some may object to condition C, stating that it is not necessary that heuristics utilize conceptual information. Many heuristics developed by computer scientists, for instance, operate within systems that (arguably) do not possess, and (arguably) are not capable of possessing, concepts. A case in point are heuristics, with varying complexities, that facilitate economical search within decision-trees. Nevertheless, such heuristics are not the same kind of heuristics that are of interest to this dissertation. Let me therefore draw a further distinction between what I will call *computational heuristics* and *cognitive heuristics*. It is the latter that concerns this dissertation.

Although cognitive heuristics may very well be computational in nature insofar as CTM holds, the way I am using the term "computational" here is to refer simply to non-conceptual processes that do not invoke a conceptual cognitive architecture. It is an open matter whether and to what extent conceptual, non-heuristic cognitive processes involve computational heuristics. However, cognitive heuristics, as I propose to understand them, are processes that utilize conceptual information. Thus, condition C holds for cognitive heuristics but not computational heuristics. To maintain generality, we might claim that heuristics involve relatively complex integrations of information. But since cognitive heuristics are the interest of this dissertation, I will keep the formulation of the condition given by C; it will have a role to play in the following chapters as I advance my thesis on how cognitive heuristics work.

2.4.1.3 Perceptual vs. cognitive heuristics

Another important distinction is between *perceptual heuristics* and *cognitive heuristics*. As mentioned above, perceptual psychologists who subscribe to the perceptual heuristics approach to vision or speech perception believe the human perceptual system to be too deficient to fully compute perceptual information coming in from the environment. And so, it is claimed, the human perceptual system relies on heuristics to make approximations. An important aspect of this view is that perception is understood to be computational in nature. However, what perceptual heuristics compute over is not conceptual information, but perceptual information. Thus, in a way, perceptual heuristics are akin to computational heuristics insofar as both are concerned with non-cognitive—that is, non-conceptual—computation. At least this is the feature that is important for present purposes. It is also important to note that perceptual heuristics and cognitive heuristics are instantiated within two different, functionally distinct systems, the perceptual system and the conceptual system respectively.

An example of a visual perceptual heuristic is:

> (18) If a repetitive texture gradually and uniformly changes scale, then interpret the scale change as a depth change. (J. Feldman, 1999, p. 215)

Although such perceptual heuristics are often expressed in conceptual terms, they are computational and non-conceptual in operation. And, as I had discussed with respect to computational heuristics, condition C does not hold.

In a certain sense, Gigerenzer's Gaze Heuristic,

> (17) (i) Fixate your eyes on the moving object; (ii) start running; (iii) adjust your running speed such that the angle of your gaze remains constant,

is not so much a cognitive heuristic as it is a perceptual heuristic. We might characterize the Gaze Heuristic more precisely as a percept-motor heuristic, since it also involves online adjustments to bodily motion. In any case, as a heuristic for object tracking, no conceptualization is required. So, at the very least, we can safely assert that the Gaze Heuristic is not a cognitive heuristic. Given the interests of this dissertation, this particular heuristic, along with other perceptual and percept-motor heuristics, will not be a subject of inquiry and will be left to one side.

2.4.1.4 Methodological vs. inferential heuristics

Finally, I make a distinction between what I will call *methodological heuristics* and *inferential heuristics*. This distinction falls within the category of cognitive heuristics. We saw that the etymological understanding of the term "heuristic" has to do with methods of discovery. It was in this sense that Pólya employed the term to refer to the endeavour to study the strategies and processes typically useful for solving problems. Using the term as a noun, heuristics in this tradition are methodological devices for learning and problem-solving. It is in this sense that the use of models and analogy are heuristic devices to help us learn or understand something about our world, and the techniques offered by Pólya, such as those given in (2)-(6), are heuristic aids in solving problems. This is what I intend to refer to by methodological heuristics. Methodological heuristics are what concern philosophers of science who are interested in creative thought, the logic of discovery, and the construction and improvement of theories in science (e.g., Lakatos, 1963-64; Popper, 1959).

However, a different meaning of "heuristic" emerged in the psychology literature, inspired (perhaps christened?) by the work of Kahneman and Tversky. As I had described above, the heuristics studied by Kahneman and Tversky (and others such as Gigerenzer, and Simon to some extent) are not precisely the methods of discovery and invention intimated by the term's etymology. The important distinction to recognize here is between the respective domains of use, and their respective functions in our cognitive lives. As tools or devices for discovery and problem-solving, methodological heuristics are aids to learning and understanding. In contrast, the heuristics that are of interest to Kahneman and Tversky (and other psychologists) are principles that guide judgment and inference. Thus, these latter are what I call inferential heuristics. Inferential heuristics are not about learning or understanding *per se*, but serve to facilitate judgments, inferences, and decision-making.

In a certain sense, inferential heuristics may be understood as special cases of methodological heuristics. This is suggested by Simon's remarks given above regarding using the term "heuristic" as a noun; in this sense, inferential heuristics aid in the discovery of a solution to a problem by providing the appropriate judgments and inferences. Nevertheless, inferential heuristics remain distinct in kind from methodological heuristics insofar as either serves distinct cognitive functions.

Moreover, we can distinguish inferential heuristics from methodological heuristics by making some generalizations. Inferential heuristics are often epistemically opaque—people often employ these heuristics without *knowing* that they do so, and without knowing the nature of these heuristics (that is, absent of a psychologist informing one of such things). Methodological heuristics, on the other hand, are generally epistemically transparent—these methods are more or less easily identified; we often consciously and deliberately employ them; their usefulness is usually known; and, because of this, an individual is able to compare and manipulate them (and all this without a psychologist informing one of such things). Moreover, methodological heuristics are typically cultivated from experience and therefore vary between individuals, whereas inferential heuristics can be to some extent immune to experience and very common among everyone, and some may even be innate.[30] For example,

(6) Find a related problem that has been solved before, and try to use its result or method,

is a technique that one acquires by working through many problems and drawing abstract principles, whereas

(11) Probabilities are evaluated by the degree to which one thing or event is representative of (resembles) another; the higher the representativeness (resemblance) the higher the probability estimation.

and

(12) The frequency of a class or the probability of an event is assessed according to the ease with which instances or associations can be brought to mind.

are pervasive, as those in the heuristics and biases tradition have shown, and are not cultivated or developed through experience.

[30] I use the qualifiers "often" and "generally" and "typically" because there will undoubtedly be exceptions. The exceptions, I suppose, can contribute to our understanding of the natures of methodological and inferential heuristics.

Since methodological and inferential heuristics are both kinds of cognitive heuristics, condition C applies to both. The epistemic opaqueness of inferential heuristics and the epistemic transparency of methodological heuristics may be owed to the kind of process each heuristic is and the kinds of information each operates over. In terms of dual-processes theory (Wason & Evans, 1975; J. Evans & Over, 1996; Frankish & Evans, 2009), it may be said that inferential heuristics are type 1 processes, and that methodological heuristics are type 2 processes. But this is just speculation.

Alternatively, this might be cashed out in terms of what representational information figures into the operations of either type of process. As strategies that help us to understand or learn something about our world, methodological heuristics must involve deep conceptual processing. At the very least, the problems that methodological heuristics are recruited to solve (learning about or understanding our world) are problems that appear to require the use and manipulation of conceptual content. This may be a factor in what makes us consciously aware of such problems, and allows us to consciously and deliberately apply methodological heuristics in solving them. On the other hand, inferential heuristics may not have to engage in deep conceptual processing to guide inference and judgment, at least not to the same extent as methodological heuristics. In chapter 4, we shall see that, according to my account of how heuristics work in cognition, it is a feature of inferential heuristics that they do not operate over conceptual content but over conceptual relations. I will briefly return to this matter below, but I save substantial discussion for chapter 4. For now, I will simply speculate that not generally engaging and manipulating conceptual content may contribute to the epistemic opaqueness of inferential heuristics. Admittedly, these are empirical issues that can be confirmed or falsified with the appropriate evidence.

This is not to say that inferential heuristics cannot be instantiated in higher-order cognition. For instance, it is possible that one can learn and consciously employ the Availability heuristic, or Take the Best. Moreover, it may not always be clear whether a given heuristic falls in the methodological or inference category. That there are fuzzy cases, however, does not bear negatively on the methodological-inferential distinction drawn here. Indeed, fuzzy cases should not come as a surprise since it is not uncommon for us to learn something about the world while we make inferences, nor is it uncommon that the very act of making a judgment solves a problem. Nevertheless, the functional distinction between methodological and inferential heuristics can be maintained so long as the functional role of the heuristic can be determined.[31]

To sum up, heuristics are to be distinguished from stimulus-response behaviour. Moreover, there are many kinds of heuristics, including (though perhaps not exhausted by) computational, perceptual, and cognitive. And cognitive heuristics can be subdivided into methodological heuristics and inferential heuristics. This is represented, with examples, in Figure 2.1. (The branch representing cognitive heuristics is highlighted since these will be the main focus from here on in.)

2.4.2 Characterizing cognitive heuristics

Now that I have distinguished between different kinds of heuristics and disambiguated the term, I will sketch a characterization of the cognitive heuristics that are of interest to this dissertation, and, as I believe, to philosophy and cognitive science generally. Although my concern will be for cognitive heuristics in general, the characterization I will give will be most applicable to inferential heuristics. This will become apparent quite quickly, and at the end the characterization I offer is exclusive to inferential heuristics. However, as I proceed I will comment on how the characterization of inferential heuristics I develop relates to methodological heuristics, as well as to computational and perceptual heuristics. In the remainder of this section, I will refer

[31] A question may arise here regarding whether the dual-process theory can allow for such fuzzy cases, since type 1 and type 2 processes are disparate processes. However, as I hope is obvious by now, the methodological-inferential distinction is a distinction primarily of function, whereas type 1 and type 2 processes are best conceived in terms of being distinct cognitive types or kinds (Samuels, 2009). Understood in this way, a heuristic which is fuzzy with respect to its being methodological or inferential can be either a type 1 or type 2 process depending on what kinds of cognitive system(s) is (are) instantiating it. It also follows that it is possible for a methodological heuristic to be a type 1 process that is unconscious, automatic, etc. This should not be too surprising since enough practice at honing a skill generally results in a lot of behaviours that are unconscious, automatic, etc. (i.e., best characterized as type 1), and I do not see why some of these behaviours should not engage methodological heuristics.

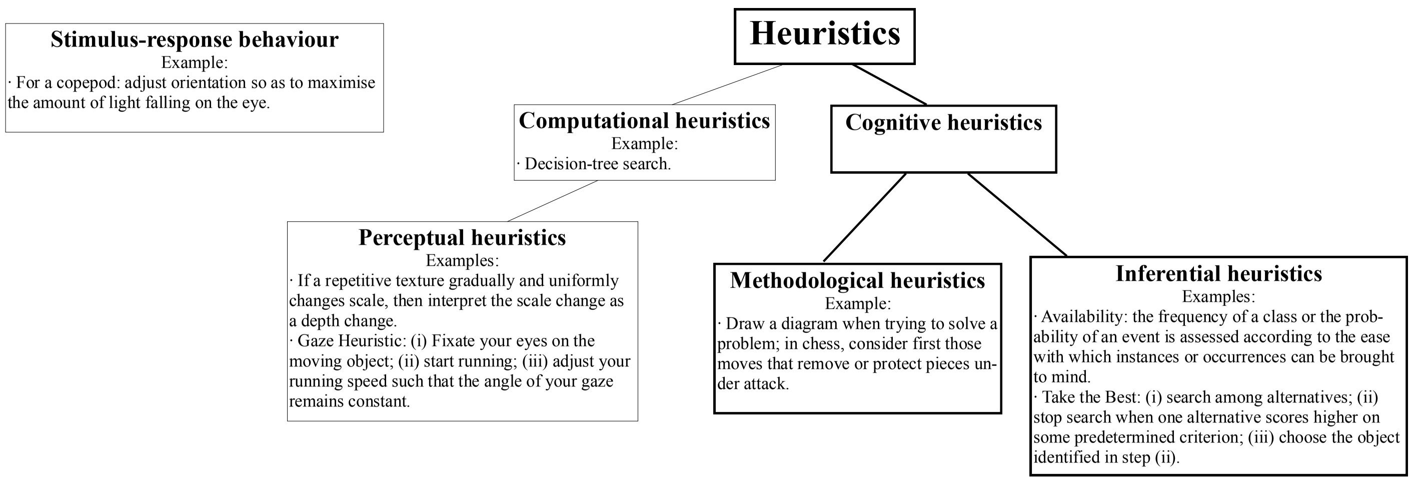


Figure 2.1: Classifying heuristics

to cognitive heuristics simply by "heuristics" unless otherwise specified.

2.4.2.1 Heuristics as rules of thumb

As might be obvious from the remarks at the beginning of this chapter, heuristics are almost invariably characterized as "rules of thumb". So common is this characterization that one may even take "rule of thumb" to be synonymous with "heuristic" (as we saw Gigerenzer to do). Though I do not want to commit to wholly endorsing this synonymity (see below), such a rule-of-thumb-definition illuminates several central qualities of heuristics. One is that heuristics are *rules* of some sort. Notice however that as rules, heuristics need not (and perhaps should not) be understood to be normative.[32] Rather, as rules, heuristics are procedures that can be specified and applied in a given situation. That is, "rule", as it is being used here, refers not to a prescriptive guide for conduct, but to a procedure that (non-prescriptively) regulates or governs some process, or describes a method for performing some operation.[33]

It would be useful here to make some distinctions with respect to the roles rules can play in governing action or behaviour. It is common to differentiate between *satisfying* a rule and *following* a rule (cf. Searle, 1980; Wittgenstein, 1953, §§185-242). To satisfy a rule is simply to behave in such a way that fits the description of the rule—to merely conform behaviour to the rule. It is in this sense that the motion of the planets satisfy the rules embodied by classical physics. On the other hand, following a rule implies a causal link between the rule and some behaviour, and moreover that the rule is an intentional object in cognition. As Fodor (2008) puts it, "following [a rule] R requires that one's behavior have a certain kind of etiology; roughly, that one's intention that one's behavior conform with R explains (or partly explains) why it does conform to R" (pp. 37-38). Thus, merely satisfying a rule is not sufficient for following the rule.

[32]Kahneman and Tversky oppose heuristics to rules (see e.g., their 1982, 1996). But they are contrasting normative rules with heuristics, the latter which I take here to be characterized as descriptive rules.

[33]Of course, there are normative matters with respect to the application of certain heuristics in certain situations. These are matters in which Gigerenzer and his followers are intensely engaged. However, to address them, one must appeal to certain *independent* normative principles that inform us of the (in)appropriateness of the (ab)use of the (descriptive) heuristic rules in question. Gigerenzer is generally uninterested in specifying such independently justified normative principles. But this is a different issue that must be considered some other time.

Our concern with respect to cognitive heuristics is not so much about behaviour as it is about reasoning or thinking (just "reasoning" for short).[34] Let us therefore say that to satisfy a rule in reasoning is to reason in such a way that merely conforms to or fits a description of the rule, whereas to follow a rule in reasoning implies a causal role for the rule in reasoning by its being cognitively represented. I suppose that in both these ways a rule can (non-prescriptively) regulate, govern, or describe a reasoning process.

We saw above that stimulus-response behaviour is best characterized as satisfying a rule; and since rule-satisfying behaviour does not implicate the involvement of any conceptual information, stimulus-response behaviour is distinguished from heuristic-produced behaviour. In terms recently stated, merely satisfying a rule is not sufficient for heuristic-produced behaviour. I believe we can likewise observe that satisfying a rule in reasoning is not sufficient for heuristic reasoning, though not because satisfying a rule in reasoning does not involve conceptual information (indeed, I am not sure how much sense can be made of nonconceptual reasoning). Rather, satisfying a rule in reasoning is not sufficient for heuristic reasoning because some heuristic must have a causal role in heuristic reasoning processes, otherwise we would not be able to say that the reasoning process was in fact heuristic. To paraphrase Fodor, heuristic reasoning requires that it have a certain kind of etiology. Nevertheless, I do not think that heuristics are necessarily rules that are followed in reasoning, for I do not think that heuristics are necessarily represented in cognition. Fodor (1975, 2000, 2008) argues that reasoning requires representing the rules according to which one reasons.[35] But this does not seem to be a necessary requirement of reasoning, especially with respect to heuristics. As we saw, inferential heuristics are often epistemically opaque, and they are pretty much automatic and to a certain extent impervious to improvement. Moreover, Simon emphasizes that inferential

[34]Reasoning might be construed as a special case of behaviour—cognitive behaviour, as it were—in which case there is not much substance to this distinction. But the point to the present discussion will be preserved either way.

[35]More specifically, Fodor believes that acting (intentional acting, that is, as opposed to reflex or thrashing about) requires planning, which requires deciding on a plan, which requires representing plans; on the other hand, he believes that belief fixation requires hypothesis formation and confirmation, which requires representing hypotheses.

heuristics are computationally cheap procedures, and assuming that representing uses computational resources, inferential heuristics would be less computationally cheap if they were required to be represented and consulted before they were employed. In any case, I see no reason why heuristics (generally) must be intentional objects in cognition. It seems entirely plausible that they can be procedures that are programmed into cognition, but are not consulted nor represented when employed; or perhaps they are built into the cognitive architecture of the reasoner as a sort of primitive process. It is certainly not impossible for computational procedures to be executed without being represented; even Fodor (2008, p. 37) admits that there must be at least some rules that are not represented in computations, e.g., those that instruct a Turing machine to move along the tape. Thus, my claim is that heuristics are plausible candidates to be unrepresented procedures in cognition, and hence that heuristics are employed, having a causal role in reasoning, but without being followed.

In light of these considerations, let me make a further distinction between, on the one hand, satisfying a rule and following a rule, and on the other hand, *acting in accordance with* a rule. Acting in accordance with a rule is sometimes meant in the same sense as satisfying a rule, but I will propose a special way to understand it here. By acting in accordance with a rule, I propose to mean that a rule is guiding one's behaviour and that there is a causal link between the rule and the behaviour—in Fodor's terms, the rule is implicated in the etiology of the behaviour that accords with the rule—but the rule is not an intentional object in cognition. In terms of reasoning, reasoning in accordance with a rule means that the rule has a causal role in the reasoning process, but is not represented. Notice that in this way a rule still (non-prescriptively) regulates, governs, or describes a reasoning process.

My suggestion, then, is that heuristics can be rules that one reasons in accordance with in the sense I just outlined. (Henceforth, whenever I speak of reasoning in accordance with a rule, I mean it in my special sense.) For heuristics certainly have a causal role in reasoning, but it does not appear that they must be intentional objects in cognition. I am inclined to believe that inferential heuristics generally are rules that one reasons in accordance with, whereas methodological heuristics are usually (though perhaps not always) followed. Whether this is true, however, does not bear on the present point, which is that heuristics (generally) do not have to be represented when employed. Let us therefore observe that:

> **$\mathbf{H}_3$** Heuristics are cognitive procedures that can be expressed as rules that one reasons in accordance with.

H_3 will serve as our initial, basic understanding of heuristics. However, to keep things simple, I will omit expressly stating that heuristics can be expressed as rules that one reasons in accordance with. Throughout the remainder of the discussion, whenever I state that heuristics are cognitive procedures in all subsequent definitions, this should be taken as shorthand for all that is expressed by H_3.

Another quality of heuristics illuminated by the rule-of-thumb-definition is that a rule of thumb is so called because a thumb can be used as a device for roughly estimating measurements. We might take this to imply that heuristics are imprecise cognitive tools. But this is too broad of a claim. For there are many imprecise inferential procedures that we would hesitate to call heuristic. For instance, procedures that are irrelevant to the task at hand, or that invariantly result in disaster, are (in some sense) imprecise, but such procedures are by no means heuristic (cf. discussion in section 2.2.1 above). Moreover, Gigerenzer has shown that heuristics can sometimes outperform complex operations on certain decision-tasks. For example, in choosing which of two cities is larger, Gigerenzer and Goldstein (1996, 1999) show that (14), the Recognition heuristic, is more accurate than weighted linear models and multiple regression. Such results throw into doubt the claim that heuristics are imprecise.

If we are to maintain the analogy between heuristic and rule of thumb, perhaps it is better that we focus on the fact that rules of thumb provide rough estimations, or in other words, *approximations*. Hence, we might say that:

> **$\mathbf{H}_4$** Heuristics are cognitive procedures that provide approximations.

Understanding heuristics as providing approximations helps our characterization in a number of ways. First, it does justice to the evidence that heuristics can sometimes result in good inferences, since approximation techniques can sometimes hit the mark, and also since what is considered "good" can very well be close to what is ideal. But this is also the sense in which computational heuristics are generally understood in computer programming and AI research—heuristics approximate some presumed normative ideal (such as a correct or an optimized outcome), and some heuristics are considered better than others if they result in better approximations (Horgan & Tienson, 1998). Furthermore, H_4 is roughly the sense in which perceptual heuristics are conceived as providing approximations.

Moreover, although approximations are deficient in the sense that they do not meet some standard, understanding heuristics as providing approximations offers a positive account of what heuristics can do rather than just stating what they cannot or do not do (as for instance the H_1 and H_2 definitions provide). This speaks to the suggestion that heuristics are *useful* devices for inference and deliberation. But if a procedure fails to approximate entirely—if it produces random results, or results that are completely off the mark—it is utterly useless, and thus should not be considered heuristic.

Nevertheless, "approximates" is a two-place predicate. If heuristics provide approximations, then they must approximate *something*. This suggests that there must be some *norm* to which heuristics approximate. I have already discussed the norms of guaranteeing correct outcomes and optimization, and rejected the characterization of "heuristic" as contradistinguished from these norms. There are heated debates over what norms should be or are approximated by heuristics, which subsequently spur the "rationality wars" (Samuels et al., 2002). But perhaps it is inaccurate to say that heuristics provide approximations. For as I have argued, although there are problems for which there are, or in principle can be, norms to which heuristics can provide approximations, there exists an entire class of problems for which it is difficult to identify what is or should be approximated. As outlined above, the problems that readily admit norms are the well-defined problems exemplified by probability and logic, while the problems for which norms are hard to identify are the ill-defined problems typified by most real-life situations, such as choosing whether to buy a Mac or PC. To what would heuristics approximate with respect to ill-defined problems?

Rather than further belabour the issue, a natural way around it is to invoke Simon's notion of satisficing. Simon originally conceived of satisficing to refer to computational search processes—a program's search is stopped once a certain, flexible goal is met, i.e., satisfied. The term has since been co-opted to apply to more general (computational) cognitive procedures (e.g., Carruthers, 2006c; Samuels, 2005). Understood in contemporary terms, a satisficing procedure is one that sets an aspiration level and attempts to meet this goal. Rather than rigidly attempting to meet some theoretical ideal, the aspiration level is in some sense "good enough" relative to the desires, interests, and goals of the agent. It is possible for a satisficing procedure to set its aspiration level commensurate with an ideal, if one exists. However, the point of satisficing is that the bar is flexibly positioned according to the goals and desires of the satisficer.

Understanding heuristics as procedures that do not aspire to meet some ideal implies that heuristics are essentially satisficing procedures. Consider two examples of inferential heuristics. The first is Gigerenzer's Take the Last,

(16) Use the cue that discriminated on the most recent problem to choose an object; if the cue does not discriminate use the cue that discriminated the time before last; and so on.

This may not appear at first sight to be a satisficing procedure. However, if we take determining a choice method to be an initial goal in a problem of choice, a certain aspiration level might be set with respect to this goal, and using the method that was used to solve the previous problem would satisfice. Or consider Kahneman and Tversky's Availability heuristic,

(12) The frequency of a class or the probability of an event is assessed according to the ease with which instances or associations can be brought to mind.

Again, this may not appear to be a satisficing procedure, but upon closer examination we see that this heuristic opts not to determine the actual frequency or probability of an event via complex calculations, but instead sets an aspiration level relative to the ease with which instances come to mind; thus, availability of instances is a "good enough" indicator of frequency.

Methodological heuristics do not appear on the surface to be readily amenable to being characterized as satisficing procedures. For example, it is difficult to see how (2)-(6) are in fact satisficing strategies:

(2) Draw a diagram when trying to solve a problem.
(3) If you can't solve a problem right away, try an indirect proof.
(4) Try to restate the problem in different terms.
(5) Assume you have a solution and work backwards.
(6) Find a related problem that has been solved before, and try to use its result or method.

However, methodological heuristics are strategies that aim to learn or understand things in the world, not by reverting to some standard operations such as formal proofs or deductive reasoning, but by using flexible, defeasible procedures. Once this is acknowledged, methodological heuristics begin to look like strategies arising from a need to satisfice. Pólya makes such a point:

> We shall attain complete certainty when we shall have obtained the complete solution, but before obtaining certainty we must often be satisfied with a more or less plausible guess. We may need the provisional before we attain the final. We need heuristic reasoning when we construct a strict proof as we need scaffolding when we erect a building. (Pólya, 1957, p. 113)

Hence, methodological heuristics appear to be satisficing procedures that are often indispensable to attaining complete and certain solutions.

Notice that satisficing does not require or imply established norms. Satisficing therefore makes perfect sense with respect to ill-defined problems, whose states and goals are often not adequately known or understood, and therefore whose norms are generally unspecified. At the same time, satisficing procedures can be applied to well-defined problems, and they can be normatively assessed according to the established norms in those situations. Understanding heuristics as satisficing procedures also recovers the notions of computational and perceptual heuristics. For a procedure that approximates is essentially one that satisfices, though a procedure that satisfices does not necessarily approximate. In addition, by appealing to satisficing we are able to account for special cases of heuristics. Some influential accounts claim that heuristics may not produce an outcome at all (e.g., Gigerenzer, 2001; Richardson, 1998; cf. Simon, 1957). An example of the latter may be a procedure that instructs an agent to give up on a problem if it is not solved within a certain time-frame. This kind of special case cannot be sensibly said to approximate anything. Yet a satisficing strategy may naturally have built-in instructions to capitulate if the set aspiration level is not met in conjunction with some independent parameter.

Thus, understanding heuristics to satisfice rather than to provide approximations gives us a more robust notion. And, in fact, satisficing appears to be absolutely essential to heuristic processing. Let us therefore amend our characterization:

> **$\mathbf{H}_5$** Heuristics are cognitive procedures that satisfice.

There are further essential features of heuristics that the rule-of-thumb-definition illustrates. For what a rule of thumb lacks in precision is made up for by its convenience and reliability. A thumb is not as accurate for measurement as markings on a rigid stick, but it is *readily accessible*, *easily deployed*, and *dependable*. The same can be said of heuristics. A heuristic is dependable because it can be counted on to produce a satisficing solution. This is in part why heuristics are powerful inference devices, and may in part explain why heuristics are so ubiquitous in human reasoning and inference (Gigerenzer, 2000). Yet it is not by chance or happenstance that a heuristic succeeds or fails. Rather, heuristics process specific kinds of information in predictable ways. This accounts for why they work so well in certain domains, but fail miserably in others. According to some researchers, such as Gigerenzer, heuristics are attuned to specific domains that exhibit stable informational structures, and this is what enables a heuristic to be successful in that domain. However, if the heuristic is applied in another domain in which the same informational structures do not obtain, then the heuristic will fail, and biases will ensue (Gigerenzer, 2000, 2006).[36]

[36]Evolutionary stories can be told here about how heuristics get attuned to one domain or another. Gigerenzer (2000) provides such an evolutionary account. One can also interpolate Sperber's (1994, 1996) evolutionary

Heuristics are readily accessible because little cognitive resources are required to engage them. We rarely, if ever, have to invest much time and processing in thinking over whether we will employ one inferential heuristic or another (or some other procedure). On the contrary, the empirical evidence shows that inferential heuristics are typically available for use without the need for reflection. In fact, many psychologists agree that some inferential heuristics are hardwired mechanisms that are essentially automatic.[37] (This corroborates my position that heuristics are cognitive procedures that can be expressed as rules that one reasons in accordance with, as given by H_3.) In a similar vein, heuristics are also easily deployed. This means that comparatively little cognitive resources are needed in their processing. This is done by processing only readily available or easily accessible information (in the environment or within the mind). Many researchers—Gigerenzer in particular—take this to mean that heuristics only process very little information. But as we shall see in later chapters, many heuristics are informationally demanding processes, though no less easily deployed. For now, however, note that the importance of heuristics being readily accessible and easily deployed resides in allowing other cognitive resources—including time—to be spared, not only with respect to processing, but also with respect to search (among information as well as procedure alternatives). This enables swift, on-the-fly inferences, and frees up precious cognitive resources for other tasks—virtues upon which survival generally depends. Thus, we have it that:

> **H_6** Heuristics are cognitive procedures that satisfice, and that require little cognitive resources for their recruitment and execution.

Notice, however, that H_6 appears to preclude methodological heuristics. This is the point at which the present characterization of heuristic processes departs with methodological heuristics. Indeed, we often take our time in deciding which, if any, methodological heuristic to use in a given case of problem-solving. Moreover, methodological heuristics can require much cognitive resources to be recruited and executed, and as such their deployment may not be so

account of the etymology of representations—his distinction between *actual domains* and *proper domains* in particular—to apply to the origins and functions of heuristics (see e.g., Samuels et al., 2002).

[37] This appears to be why Wason and Evans (1975) contrast heuristic reasoning with analytic reasoning.

easy. The reason for this, it seems, is that methodological heuristics demand that much more conceptual information be integrated than what is demanded of inferential heuristics. As I suggested above, inferential heuristics do not involve deep conceptual processing, at least to the same extent as methodological heuristics. And it may be the case that the extent to which a heuristic engages conceptual information has to do with its respective function in cognition, and subsequently the kinds of information operated over. This is not to say that methodological heuristics always require excessive amounts of cognitive resources. As one develops certain cognitive skills, such as problem-solving skills, one would surely hone the ability to call upon and apply methodological heuristics with relative ease, and perhaps even automatically and unconsciously.[38] Nor is this to say that inferential heuristics will always require little cognitive resources. Some inferential heuristic strategies might consume quite a bit of cognitive resources and time. Nevertheless, it stands to reason that, in general, methodological heuristics will require more cognitive resources in their deployment and processing than inferential heuristics, since the former generally involve processing and integrating conceptual content. But again, these are empirical matters. I am content with focusing on inferential heuristics that require little cognitive resources for their deployment, leaving these further issues to be explored on some other occasion.[39]

[38]Perhaps this would be owed to repetition and practice creating and developing new representations, or representing old representations in new and different ways; and inferential heuristics operating over these new representations take over what methodological heuristics had once achieved. But this is just a thought in passing.

[39]A natural inclination might be to assume that inferential heuristics are the proper operations of mental modules, whereas methodological heuristics operate within central cognition. But I think this would be wrong. Although it is likely that modules do not (cannot?) employ methodological heuristics (perhaps for reasons having to do with the modules' automaticity, shallow inputs-outputs, and speed; Fodor, 1983), it seems perfectly reasonable that central cognition can and does employ inferential heuristics for given problems. In fact, the cognitive architecture that I develop in coming chapters corroborate this point. Notice, however, that whether and to what extent modules employ methodological heuristics and central cognition employs inferential heuristics depends on what we mean by "module". Those who endorse the massive modularity thesis (i.e., the thesis that the mind is entirely, or nearly entirely, composed of individuated modules) do not subscribe to Fodor's characterization of modules. And if those such as Sperber (1994, 2005) and Carruthers (2006b, 2006a) are right about the architecture of the mind, then modules in some sense can and do employ methodological heuristics, even if it takes collections of modules to pull it off. I will be making some remarks about modularity in the next chapter. However, a full assessment of modularity of mind and massive modularity is beyond the scope of this dissertation.

2.4.2.2 Beyond rules of thumb: exploiting information

Before being satisfied with our working definition of heuristics developed thus far, let us revisit and emphasize a key issue that has been pervasive in much of the discussion. Despite the illumination that the rule-of-thumb definition offers, heuristics are more than just rules of thumb. The nature of heuristics *qua* cognitive processes involves the utilization of conceptual information. Granted I have suggested that inferential heuristics do not process conceptual content; nonetheless, what they do process has everything to do with concepts. Inferential heuristics may not process conceptual content, but they do utilize conceptual information. I already mentioned that I will argue in chapter 4 that inferential heuristics operate over relations between concepts. Utilizing conceptual information is precisely what makes cognitive heuristics cognitive. This is provided by condition C, as presented above. A rule of thumb *per se*, on the other hand, is merely a static, rough unit of measure, and no conceptual information is implicated in its employment. In this vein, rules of thumb do not involve the type of information processing that heuristics do.

To put the point another way, when one employs a heuristic, as I am characterizing it here, it tells us something about one's concepts—particularly about the content of one's concepts and the structure within which one's concepts reside. When one employs (15) Take the Best, for instance, certain beliefs are implied about the cue upon which the choice in question was made. At the very least, we can infer that the said cue was believed to be the "best" upon which to make the choice in question; but that the cue is believed to be the "best" implies certain things about the conceptual content of the cue (as possessed by the chooser), as well as certain things about how the cue fits within the chooser's conceptual structure. Similarly, when one employs (11) the Representativeness heuristic, certain beliefs are implied about what one believes about and how one conceptualizes the objects or events under evaluation, and how the concepts involved in the evaluation fit within one's conceptual structure. We will see more detail on this matter as we continue through this dissertation. The point, however, is that one's conceptual wherewithal has a significant role in employing heuristics.

When one employs a rule of thumb, on the other hand, one's conceptual wherewithal may not play much of a role at all, and therefore the use of a rule of thumb may tell us nothing about one's concepts (their content or conceptual structure). "When constructing a truth tree, decompose those formulas that use stacking rules before those that branch"; (1) "To choose a ripe cantaloupe, press the spot on the candidate cantaloupe where it was attached to the plant and smell it; if the spot smells like the inside of a cantaloupe, it's probably ripe"; "Start in the centre square when beginning a game of tic-tac-toe"; "Measure twice, cut once". These are all rules that can be applied willy-nilly with very little requisite conceptual content or conceptual connections between the information processed. For example, we can imagine someone with little experience with sentential logic being presented with the truth tree rule, say in a lecture, and told to follow it (more on following presently). However, that this person is able to follow the rule tells us nothing interesting about her beliefs or concepts—she is just doing what she is instructed to do. It is in this sense that the person would be employing the rule as a rule of thumb rather than a heuristic. It is important to understand that what distinguishes rules of thumb from heuristics, as I am claiming here, is the manner in which the rule engages one's conceptual wherewithal, and not the rule itself. In other words, the distinction has to do not with the procedure but with the information structure over which the procedure operates.

A distinction between rules of thumb and heuristics can also be found with respect to the basic understanding of heuristics provided by H_3, namely that heuristics can be expressed as rules that one reasons in accordance with. Let us recall here the distinction I made earlier between following a rule in reason and reasoning in accordance with a rule. Following a rule in reason implies a causal link between the rule and some behaviour, and that the rule is an intentional object in cognition; reasoning in accordance with a rule likewise implies a causal link between the rule and some behaviour, but the rule does not likewise have to be an intentional object in cognition. Upon reflection, we can now see that, whereas heuristics do not have to be represented, rules of thumb usually are; that is, whereas heuristics are rules that one reasons in accordance with, rules of thumb are typically rules that are followed.[40]

I therefore resist simply equating heuristics with rules of thumb. The basis for this is ultimately that heuristics require richer representational content and conceptual information over which to operate; that is, heuristics *exploit informational structures*, and they are successful because they do so. This is similar to the idea present in Gigerenzer's conception of heuristics, as presented above, although as we shall see in the next chapter, my understanding is richer and goes beyond Gigerenzer's.[41] In any event, let us now draw the characterization of heuristics that the discussion has led us to:

> $\mathbf{H}_7$ Heuristics are cognitive procedures that satisfice, and that require little cognitive resources for their recruitment and execution; they operate by exploiting informational structures.

Following my comments above with respect to H_6, the present characterization of "heuristic" should be understood to refer only to inferential heuristics. It captures what seems to be important features of heuristics as they are discussed in the philosophical and cognitive science literatures, while at the same time being precise enough to pick out a specific kind of cognitive process. The characterization may therefore be used to help us understand a broad but distinct range of phenomena, and thereby make sense of certain psychological theories of reasoning, inference, and decision-making. We might thus stand to gain a clearer picture of how the mind works. Heuristics are supposed to be ubiquitous in human cognition, and so understanding the nature of heuristic processes significantly contributes to understanding the nature of the mind in general.

The present characterization also enables us to see the respects in which the disparate conceptions of heuristics given by different researchers and authors overlap, as well as come apart.

[40] On this account, it is possible that a rule starts out as a represented rule of thumb, and with enough experience the rule turns into a heuristic as one's conceptual wherewithal plays more of a role when using the rule. This I believe to be plausible especially with respect to methodological heuristics. Cf. footnote 38 above.

[41] Although more detail on the differences between my view and Gigerenzer's will come out in the next chapter, let us here simply note that Gigerenzer does not make the distinctions I do between kinds of heuristics, as evinced by his confounding what I had called stimulus-response and heuristic-produced behaviour (see above). Further and importantly, Gigerenzer does not develop an understanding of heuristics as exploiting conceptual structures as I am presently doing.

For example, as was indicated above, Gigerenzer's use of the term "heuristic" to refer to (what I am calling) stimulus-response "rules of thumb" of animals is precluded by the present characterization. Moreover, Gigerenzer often cites an algorithmic decision-tree as constituting a heuristic (Gigerenzer & Todd, 1999; Gigerenzer, 2001, 2007, 2008a, 2008b). But according to the proposed characterization, algorithmic decision-trees are not really heuristic since their deployment does not exploit informational structures embodied by representations or concepts in an interesting way. Rather, such decision-trees are simply rote algorithms (albeit perhaps simplified compared to full decision trees). Other supposed heuristics will also fall by the wayside accordingly. In addition, the proposed characterization helps us to understand more clearly whether and to what extent the examples of heuristics given above from Simon, Kahneman and Tversky, Gigerenzer, Pearl, and others refer to the same concept. Some heuristics, such as those proposed by Pólya, will be seen to be essentially methodological devices to provide insight, and are of a different species than the inferential heuristics Kahneman and Tversky researched.

2.5 Concluding remarks

Let us briefly review what has been done so far. We started with a vague understanding of what "heuristic" means and what heuristics are. With the motivation of clarifying the notion of "heuristic" in order to advance research and investigations in philosophy and cognitive science, I proceeded to consider two perfunctory accounts: (H_1) that heuristics are procedures that do not guarantee correct outcomes, and (H_2) that heuristics are operations that do not maximize/optimize. I argued that these perfunctory accounts are unsatisfactory, since they either fail to sensibly apply to a range of important cases of practical cognition, or they trivialize interesting features of cognition. In this way, these negative definitions are not useful for the kinds of phenomena that we are interested in. I then considered four influential, positive accounts of heuristics, and by drawing a number of distinctions between different kinds of heuristics, I narrowed in on a cluster of notions to do the work needed of them; the rule-of-thumb analogy proved most helpful for this purpose. The end result was a working characterization of inferential heuristics: (H_7) Heuristics are cognitive procedures that satisfice, and that require little cognitive resources for their recruitment and execution; they operate by exploiting informational structures. I believe this account is useful and interesting for investigating how the mind works, and moreover that it clarifies some of the uses of the term "heuristic" in the relevant literatures.

This chapter also serves to motivate further philosophical investigation into the nature of cognitive heuristics, and what it is about the mind that enables heuristics to be "fast and frugal". One matter in particular demands further attention, namely the nature of the informational structures that heuristics exploit. As indicated above, Gigerenzer (2000, 2004, 2006) shares a similar view; he contends that heuristics exploit informational structures in the environment. However, Gigerenzer neglects to sufficiently describe such informational structures. I have not sufficiently described the structures that I believe heuristics exploit either. We might observe, nonetheless, that on the present account the informational structures do not reside in the environment, but are cognitive structures. An investigation into the cognitive structures exploited by heuristics can yield insight not only into how heuristics work and why they can be fast and frugal, but also into more fundamental issues concerning the architecture of cognition.

In the next chapter I will explore more fully the idea that heuristics exploit informational structures. More specifically, I will offer an account of the kinds of cognitive structures that heuristics exploit, and argue that they are in fact rich structures that are highly organized in specific ways. I will develop this idea within the purview of a special class of problems related to heuristics, namely problems of relevance.

Chapter 3

Beginning to Understand How Heuristics Work

Now that we have a working definition of heuristics,[1] we are in a position to illustrate the work it does for us. The present chapter will do this by first considering the sorts of problems that many philosophers believe heuristics solve, namely what can be broadly described as "problems of relevance" (Samuels, 2005, forthcoming). As we shall see, however, Fodor (e.g., 2000) argues that heuristics do not in fact solve problems of relevance. In order to appreciate this controversy and what is at stake, we must get a firm grasp of the role heuristics are purported to play (or fail to play) in cognition. And in order to do this, we need to gain a better understanding of how heuristics work and the structures they operate over.

I will begin this chapter, in section 1, by discussing the problems of relevance I am referring to. As I explain below, problems of relevance are more commonly collectively referred to as *the frame problem*. The frame problem is a special problem for cognitive systems of circumscribing (read: framing) the information to be considered in their processing. But the frame problem actually constitutes a set of closely related problems for computational cognition. I will expound various aspects of the frame problem, labeling them as I go, and show how they generalize in terms of relevance. Philosophically significant implications follow from computational and epistemological dimensions of the frame problem. The computational aspect will be what guides the present chapter; the epistemological aspect will be a topic to be discussed in the final chapter.

Expounding the problems of relevance will set the stage for me to advance a general architecture of cognition. The cognitive architecture that I put forth in section 2 will be of so-called central systems, paradigmatically responsible for general reasoning. A widely agreed upon feature of central systems is that they allow for the free exchange of information (this is actually partly how relevance problems arise in the first place). This is usually taken to imply that central systems are not dedicated to computing specific information belonging to specific domains—that is, they are not believed to be domain-specific, but domain-general. However, some theorists—mainly evolutionary psychologists—argue that central cognition is best con-

[1] Henceforth, I will simply use "heuristics" to refer to inferential cognitive heuristics, unless otherwise specified.

ceived as a collection of a great many domain-specific mechanisms, or "modules", since (so the argument goes) a domain-general device cannot solve the specialized problems that humans are naturally very good at solving (e.g., predicting the intentions of others). Nevertheless, borrowing from Richard Samuels (1998, 2000), I will argue that central cognition is indeed domain-general, but that central systems are able to draw on specialized bodies of knowledge.

In section 3, I will explain more fully what I take these specialized bodies of knowledge to be. Samuels (2000) recognizes that it is possible for humans to possess innate, domain-specific, *non-computational* modules over which a domain-general, computational, central system operates.[2] My view comes close to Samuels'. However, I will argue that such non-computational modules are better conceived as what I call "k-systems". K-systems, as I will explain, are highly organized, informationally rich bodies of knowledge; the information within a k-system exhibits specific relations, and the k-systems themselves bear specific relations to one another. This view is not intended to be understood as wholly original, but we shall see that my proposal is novel with respect to how I envision heuristics to interact with such systems of knowledge.

Section 4 will be devoted to expounding the properties of k-systems in more detail. It will begin to be evident how heuristics work by exploiting k-systems. I will also show how my theory is similar to the views advanced by Gigerenzer (2000) and Kim Sterelny (2004, 2006), but I will also identify important respects in which my view is different from each of these. The idea of heuristics exploiting k-systems will pave the way for the next chapter in which I provide a more specific account of what k-systems are in order to advance a more specific account of how heuristics work. If it turns out that the assumptions about cognitive architecture which underlie my account of heuristics help in solving, or otherwise circumventing, the frame problem (or problems of relevance), that is reason to suppose that those assumptions are plausible.

[2]In anticipation, non-computational modules are not the types of modules that Fodor (1983) envisioned, although he more recently acknowledged that one can conceive of non-computational modules (Fodor, 2000).

3.1 Heuristics and relevance problems

A longstanding philosophical issue related to the purported use of heuristics in human cognition has to do with determining the relevant information to consider for a given task. Determining what information is relevant to a task is not as simple as it may appear. One problem is that no one has yet given a satisfactory account of what relevance is. This is a problem that plagues any account of relevance-determination, and fully addressing it is beyond the scope of this dissertation. I will therefore rely on an undefended intuitive account of relevance, as most authors do, but I will have some more to say on the topic in the concluding chapter of this dissertation. Nevertheless, even if we possessed a refined account of relevance, determinations of relevance are generally not *computationally* easy. This is because many of the problems faced by humans can (in principle) require that generous amounts of information be surveyed. This is where heuristics come into play—heuristics appear to be prime candidates for alleviating the computational burden of relevance determinations. Questions that demand answers, then, are: Do heuristics actually solve the problem of determining relevance? And if so, how do they do it?

The problem of determining what information is relevant to a given cognitive task has its roots in the frame problem. However, specifying what the frame problem is is a difficult task (cf. Dennett, 1984). As it is generally understood nowadays, the frame problem actually constitutes a set of closely related problems for computational cognition. Over the years, different authors have emphasized different aspects of the frame problem. Fodor's (1983, 1987, 2000) interpretation is perhaps the most popular in philosophical circles; and it has generated quite a bit of controversy, as he declares that the frame problem is a central and insurmountable problem for computational cognitive science. Yet Fodor's account of the frame problem is not always entirely clear, since he tends to gloss over some of the subtle dimensions of the problem. To get a better idea of what the frame problem is, how it gives rise to more general problems of relevance, how deep these problems run, and the potential role for heuristics, I will expound various aspects of the frame problem and explain their philosophical significance to begin the chapter. A further aspect of the frame problem will be revealed toward the end of this chapter.

3.1.1 The many guises of the frame problem

The frame problem was first introduced by John McCarthy and Pat Hayes (1969) as a problem faced by early logic-based AI. This was a problem of how to represent, in a logically succinct way, the ways in which properties of objects do not change when acted upon (Shanahan, 2009). For any given action, there will be many properties of many objects that remain constant throughout and after the execution of the action. When moving a ball from one location to another, for instance, the ball's colour does not change, nor does its shape or its size; other objects in the room do not change; the fact that $2 + 2 = 4$ does not change; and so on. If a cognitive system is to maintain a veridical belief-set[3] about the ball (and the world in general), determinations of which properties change with which actions will have to be coded into its program. In logic-based AI, this is done by coding what are called "frame axioms", which are sentence-like representations that specify a frame of reference for the action in question (Dennett, 1984; Viger, 2006a). But the number of frame axioms can become quite cumbersome very quickly since there will in general be (*number of properties*) × (*number of actions*) frame axioms that would have to be written into the program (Shanahan, 2009). Obviously, the more properties or actions to account for, the more cumbersome this will be. Thus, the problem for logic-based AI is how to most compactly represent the non-effects of actions. As such, it is often referred to as a problem of *representation*. Let us call this the *AI frame problem*:

> **AI frame problem** The problem of how to appropriately and succinctly represent in a system's program the fact that most properties are unaffected by given actions.

The main obstacle to overcoming the AI frame problem is the monotonicity of classical logic. As a monotonic logic, classical logic does not allow for different conclusions to be drawn with the addition of new premises. AI researchers have since developed a number of

[3] It is obviously too demanding of any system to maintain a completely veridical belief-set. What is required, instead, is a belief-set that is suitably veridical, or reasonably accurate (cf. Samuels, forthcoming).

nonmonotonic logics to remedy the situation. According to Shanahan (2009), the introduction of nonmonotonic logics adequately solves the AI frame problem, notwithstanding certain technical problems.[4] Whether Shanahan is correct is perhaps debatable (see e.g., Fodor, 1987, 2000), but I will not adjudicate his claim here. For the original AI frame problem and its purported solutions are not of concern to this dissertation or to philosophy more generally. Rather, what is of philosophical interest are the philosophical and computational concerns born out of the AI frame problem, to which I now turn.

Philosophers have interpreted the AI frame problem as a general problem of how a cognitive system updates its beliefs if it is to maintain (suitably) veridical beliefs about the world. As Fodor observes, "Despite its provenance in speculative robotology, the frame problem doesn't really have anything in particular to do with action. After all, one's standing cognitive commitments must rationally accommodate each new state of affairs, whether or not it is a state of affairs that is consequent upon one's own behavior" (1987, p. 27). This interpretation of the frame problem has an obvious connection to the AI frame problem. Upon the arrival of new information, a cognitive system will have to update its beliefs. In moving a ball from one location to another, for instance, a cognitive system will have to update its belief regarding the position of the ball, as well as the ball's new spatial relations with respect to other objects in the room. But there are arbitrarily many other beliefs that it *will not* have to update, such as the colour of the ball, the sizes of the other objects in the room, the relative spatial relations of the unmoved objects, the ambient temperature in the room, that $2+2=4$, and so on. The problem, then, is determining which of any of the system's beliefs is to be affected in belief-update. Such a determination cannot be made *a priori*. For instance, are the beliefs about the relative spatial relations of the other objects in the room affected? That depends: was something knocked over in the process of moving the ball? Unlike the original representational problem of logic-based

[4]One technical problem that is often cited is the "Yale Shooting Problem". In this scenario, a gun is loaded at time t, and shooting the gun at Fred at $t+1$ is expected to kill Fred. However, the formalization of nonmonotonic logics cannot uniquely determine whether Fred is dead. This is because the logic proves that Fred indeed dies, but the logic also proves that the gun becomes (mysteriously) unloaded and Fred lives. Yet the formalism has no way to determine which proof to keep.

AI, this problem is *computational* in nature (McDermott, 1987)—it concerns the *processes* a cognitive system employs as it updates its beliefs. Let us call this problem the *update problem* (cf. Samuels, forthcoming):

> **Update problem** The problem of determining which of any of the system's beliefs is to be affected in belief-update.

This interpretation of the frame problem can be generalized as a problem of relevance—that is, a problem of determining what is relevant to the task at hand. This is a deep problem especially for human reasoning since central systems—the cognitive systems that are paradigmatically responsible for general reasoning and decision-making—admittedly allow for free exchange of information, any of which can bear on their operations. Thus, central systems are what is often referred to as *holistic* (or, in Fodor's (1983) terminology, *isotropic*). This is to say that, given an appropriate set of background beliefs, any representation *in principle* can be relevant to any other. Who won tonight's football game is *prima facie* irrelevant to whether there is beer in your friend's fridge. But if you believe that your friend's favourite football team played tonight, and that your friend usually overindulges in beer consumption whenever his favourite team wins, then your belief about who won tonight's game actually is relevant to your belief about beer in your friend's fridge. Indeed, an itch on your left hand can be relevant to your belief about whether you will get a job promotion, if you carry an appropriate set of (superstitious) background beliefs. Again, there is seemingly no way *a priori* to circumscribe the subset of representations that are relevant in a given occasion of reasoning or inference. I will refer to this construal of the frame problem as the *generalized relevance problem*:

> **Generalized relevance problem** The problem of determining, from all the information that can in principle bear, what is relevant to the cognitive task at hand.

It is perplexing for cognitive science that humans have an astonishing ability for knowing (with reasonable levels of success) what is relevant in much of their reasoning and fixating thereon without having to spend much time deciding what is and is not relevant, or wasting time cogitating on irrelevant details. In terms of the generalized-relevance guise of the frame problem, we generally know quite easily what pieces of information bear on—or lie within the frame of—our reasoning tasks. The problem is understanding how we do it.

It is important to be clear, in light of these observations about human performance, what the generalized relevance problem is and what it is not. The problem is not how a cognitive system can possibly determine, from all the information that can in principle bear, *everything* that is relevant to the task at hand. Humans typically do not consider everything that is relevant, and it is normally too cognitively demanding (of human or machine) to do so. Not only is this demonstrably the case (witness the frequency of our errors, or of surprise), but it is often impossible to consider everything that is relevant to the task at hand. Suppose for instance that the number of hairs on Caesar's head on March 15, 44BCE is relevant to the extent to which the current state of the economy will improve over the next five years. Although the former piece of information can in principle be relevant to the latter, it is impossible to know the number of hairs on Caesar's head on March 15, 44BCE; and so it would be impossible to consider everything that is relevant to estimating the extent to which the current state of the economy will improve over the next five years.

The generalized relevance problem is, rather, a problem of how a cognitive system can make determinations of what is relevant to a given task *with reasonable levels of success*. The problem is for real-world cognitive systems—cognitive systems that are less-than-perfect, and that suffer from real constraints in terms of time and cognitive resources. Therefore, the generalized relevance problem suggests a methodological problem for a cognitive system of making reasonably accurate determinations of relevance efficiently and in a timely fashion; and not determining everything that is relevant to the task at hand, but enough to get the job done. This means, among other things, that there is minimal waste of time and cognitive resources considering what is irrelevant. The latter observation might appear to be a mundane point, but as we shall see presently it reveals an additional deep epistemological problem.

The methodological problem has to do with the computational burden put on a system in determining the set of representations to be considered for a given cognitive task. Since relevance cannot be determined *a priori*, and since any representation held by a cognitive system can in principle be relevant to any other (provided an appropriate set of representations or background beliefs), determining which representations to actually consider can be computationally intractable. If a system has a sufficiently small set of beliefs to begin with, relevance determinations can be performed without running into major computational problems: the system may simply consider all representations. But once we consider a cognitive system with sufficiently many and/or complex beliefs—as is the case with humans, who have a wealth of beliefs and stores of other complex representations—assessing every representation quickly becomes infeasible. For such a strategy would entail a vast amount of computations, occupying scarce cognitive resources and taking an unreasonable amount of time. Indeed, assessments of relevance of individual representations are computationally taxing enough, but assessments of relevance *between* representations exponentially adds to the complexity (Samuels, 2005). What is needed are computationally tractable methods to pick out, in a timely fashion, which representations are to be brought to bear on a given cognitive task (cf. Shanahan, 2009). Let us call this the *computational relevance problem*.

The computational relevance problem The problem of how a cognitive system tractably delimits (i.e., frames) what gets considered in a given cognitive task.

It is important to notice that the computational relevance problem would not be solved or circumvented if a cognitive system were merely to employ a method to differentiate between a relevant representation and an irrelevant one, and then subsequently ignore the representations that fall into the latter category. This is a divide-and-conquer strategy in which the dividing part is too computationally taxing. For not only can such a strategy possibly entail considering each representation held by the system, but seeing whether a given representation is relevant means considering the representation and drawing implications. Even if this strategy does not entail considering each of our held representations, we do not have the time to consider many representations and their implications with respect to the task at hand. In other words, the system would be performing quite a lot of computations to see whether a given representation is relevant and therefore should be brought to bear. Of course, most beliefs and other representations a system holds will be irrelevant to the cognitive task at hand, and this would then translate to wasting quite a lot of resources on considering what is irrelevant, and performing computations by drawing conclusions that do not need to be drawn. This is what Fodor (1987) aptly referred to as "Hamlet's problem: How to tell when to stop thinking" (p. 26).

The computational relevance problem is therefore a problem of how a cognitive system can *ignore* most of what it knows (Dennett, 1984). More precisely, the problem *qua* relevance problem is how a system ignores *what is irrelevant*. This brings us to the other dimension of the generalized relevance problem hinted at above. The problem here has to do with how a cognitive system considers *only* what is (mostly) relevant. This is what I take to be the "new, deep epistemological problem" that Dennett (1984, p. 130) spoke of. Again, since relevance cannot be determined *a priori*, and since any representation held by an agent can in principle be relevant to its task (provided an appropriate set of background beliefs), how does a cognitive system *know* what is relevant? I will refer to this problem as the *epistemological relevance problem*.

The epistemological relevance problem The problem of how a cognitive system considers only what is (mostly) relevant, or equivalently, how a cognitive system knows what is relevant.

It is important to note that this epistemological problem[5] does not have to do so much with the computational costs of delimiting what gets considered in a given cognitive task (although this remains an important aspect) as it has to do with considering the *right* things.[6]

[5]What I am calling here the epistemological relevance problem does not coincide with what Shanahan (2009) calls the epistemological frame problem. According to Shanahan, "The epistemological problem is this: How is it possible for holistic, open-ended, context-sensitive relevance to be captured by a set of propositional, language-like representations of the sort used in classical AI?". It appears that what Shanahan has in mind for the epistemological problem is something closer to the AI frame problem, but for humans in determining relevance.

[6]We might understand the distinction between the computational and epistemological relevance problems, as I presented them here, as in some respects corresponding to the distinction I made in chapter 2 between computational and cognitive heuristics (see section 2.4.1).

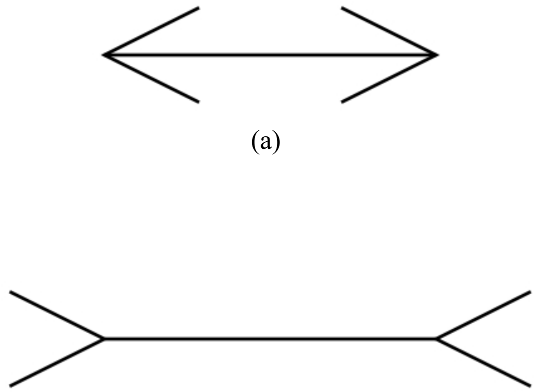
3.1.2 The heuristic solution

Many philosophers and cognitive scientists have posed the question: How do we solve the frame problem? However, this question is rarely, if ever, posed within the context of demarcating the various aspects of the frame problem outlined here. The question, "How do we solve the frame problem?", might not even be meaningful without specifying which guise of the frame problem is being referred to. It does appear as if humans somehow manage to solve all aspects of the frame problem, and as we shall see some philosophers believe that heuristics are the operative solution in human cognition. But we might want to ask whether heuristics really offer a tenable solution to the frame problem, and indeed which, if any, guises of the frame problem humans actually solve despite appearances.

By way of addressing these concerns, I will now introduce the heuristic solution to the frame problem as a response to worries emphatically advanced by Fodor concerning a plausible theory of central cognition. My intention will then be to show that the heuristic solution, as advanced as a response to Fodor, applies only to the computational relevance problem, ignoring the epistemological problem. In chapter 6, I take up the question of whether heuristics do in fact circumvent the epistemological problem. For ease of exposition, I will use "relevance problems" to refer to all guises of the frame problem here expounded, save for what I have called the AI frame problem; as mentioned, the AI frame problem will not be of concern to this dissertation.

Fodor (1983, 1987, 2000) argues that only a suitably *informationally encapsulated* system can avoid relevance problems. Roughly stated, informational encapsulation refers to a property of a cognitive system that can draw on only a highly restricted database of information in the course of its processing. There is some debate on how to interpret informational encapsulation and what is entailed by it.[7] However, for our purposes let us understand it thus: A cognitive system is encapsulated to the extent that, of all the information held by other systems that is relevant to its processing, only a strictly delimited portion is available to it. The peripheral

[7]For further discussion see Carruthers (2006a, 2006c); Samuels (2005, 2006); Sperber and Wilson (1996).



(a)

(b)

Figure 3.1: The Müller-Lyer Illusion. (a) appears shorter than (b), although the lines are the same length; and (a) continues to appear shorter than (b) despite *knowing* that the lines are of equal length. The visual system is therefore said to be encapsulated with respect to the belief that the lines are of equal length.

(as opposed to central) cognitive systems dedicated to linguistic and perceptual processing are paradigm cases of encapsulated systems. A common example is the Müller-Lyer illusion (see Figure 3.1): the belief that the two lines are actually equivalent is relevant to determining their relative lengths, but the visual system is encapsulated with respect to this information; it has access to only a highly restricted or proprietary database (of visual information) in the course of its processing.

Relevance problems do not arise for sufficiently informationally encapsulated systems, since there would be a sufficiently small amount of information over which to compute. To be more precise, encapsulated systems avoid relevance problems in two subtly distinct ways: Not only does the small amount of information contained in the system's database constitute *all* the information that the system *can* consider, thus considerably reducing the number of computations needed for information search, but that small amount of information *ipso facto* constitutes the one and only set of background information against which relevance is determined. The more encapsulated a system is, the more tractable its computations will be, and the less that relevance problems will be problems.[8]

Nevertheless, Fodor stresses that central cognition cannot be encapsulated given its holistic character and the wealth of complex beliefs and other representations it has access to. Central processing thus appears squarely faced with relevance problems. In addition, Fodor asserts that central cognition is, what he calls, "global". There has been some misconceptions in the literature about what Fodor is referring to, but the basic idea is this: Computations are, by definition, operations over local, context-insensitive properties, i.e., the syntax of the items being computed. In other words, it is only local properties that are causal determinants in computations. Fodor points out, however, that there are properties that central cognition assesses, such as relevance, simplicity, and conservativism. Unlike the local properties of syntax, these are extra-syntactic, global properties of representations in relation to (what can be very large) contexts, requiring in the limit that entire belief systems be assessed.[9] The result is intractable cognition. According to Fodor, there can be no computational account of how central cognition is sensitive to global properties, since only syntax is supposed to be a causal determinant in computations. He therefore concludes that CTM is not a viable theory for cental cognition. And since he claims that CTM is the only plausible theory of mind, Fodor is therefore pessimistic about the prospects of a cognitive science for central cognition. Hence:

> 'Fodor's First Law of the Nonexistence of Cognitive Science' ...: the more global (e.g., the more isotropic [holisitc]) a cognitive process is, the less anybody understands it. *Very* global processes, like analogical reasoning, aren't understood at all. (Fodor, 1983, p. 107)

As Fodor (2000) later makes clear, we can add assessments of relevance to the "*very* global

[8]Of course, this does not mean that the system's database will necessarily contain everything that is relevant to its tasks. And so it is possible that an encapsulated system can still have a problem with respect to considering what is relevant. But this poses a relevance problem of a different sort, which I will not be discussing here. In addition, even if an encapsulated system had access to all relevant information (but not only what is relevant), the system might still face the epistemological relevance problem if its database was not organized in such a way that would facilitate *determinations* of relevance, or that would enable the the system to consider only what is relevant. See the discussion toward the end of this chapter on what I call the representational frame problem. Nevertheless, at the very least, a suitably encapsulated system will avoid the computational problem of tractably delimiting what gets considered in its tasks.

[9]Fodor (2000) argues that this implies that entire theories are the units of cognition. See Samuels (forthcoming) for arguments against this view.

processes" he has in mind.

In response, however, certain philosophers argue that Fodor overstates what tractable computation requires. Samuels (2005, forthcoming) and Carruthers (2006a, 2006c), for instance, both claim that, although encapsulation will ensure computational tractability, tractability does not necessarily require encapsulation.[10] Rather, they argue that all that tractable computation requires is that computations be suitably *frugal*, in the sense of simply putting limits on the amount of information used or processed. A frugal process may therefore have every bit of information at its disposal, and hence be unencapsulated, but considers only a tractable amount according to certain parameters (e.g., the number of computations needed to gather the information).

Frugal computational processes can, of course, be realized by "techniques that often, though not invariably, identify a substantial subset of those representations that are relevant to the task at hand" (Samuels, forthcoming), namely *heuristics*. The idea, then, is that by employing heuristics a system can avoid computationally taxing assessments of relevance, and they do this by instantiating suitable search and stopping rules that effectively limit the amount of information surveyed in its computations, but that still bring to bear an appropriate subset of information on the system's tasks. This escapes the need to invoke encapsulation in explaining computational tractability. For a system can very well be unencapsulated, where *any* information *can* be brought to bear on a system's computations, but the use of heuristics will ensure that only a small (and hopefully relevant) amount of this information *will in fact* be brought to bear in its processing.[11] Hence, for these theorists heuristics ensure that central cognition is computational and unencapsulated, but yet operates in a tractable way.[12]

[10] See also Sperber and Wilson (1996).

[11] Carruthers actually believes that we may still speak of systems as informationally encapsulated if we construe encapsulation as a property that simply restricts informational access in processing. In this way, Carruthers believes that encapsulation can mean just that most information simply will not be considered in a system's computations—a notion he calls "wide encapsulation". However, Samuels (2006) criticizes Carruthers' view, claiming that "not only is it [wide encapsulation] different from what most theorists mean by 'encapsulation,' but it's simply what you get by denying exhaustive search; and since virtually no one thinks exhaustive search is characteristic of human cognition, the present kind of 'encapsulation' is neither distinctive nor interesting" (p. 45).

[12] We might think of it like this: To determine everything that is relevant to the task at hand, the powerset of the

Recently, Fodor (2008) exclaims that resorting to heuristic procedures to circumvent relevance problems "is among the most characteristic ways that cognitive scientists have of not recognizing how serious the issues are that the isotropy [holism] of relevance raises for a theory of the cognitive mind" (p. 116). By referring to what has come to be known as the "sleeping dog" strategy (McDermott, 1987),[13] he goes on to argue that the appeal to heuristics begs the question it is supposed to answer. Specifically, Fodor claims that the sleeping dog strategy takes for granted some principle that individuates situations; but the task of individuating situations presupposes that determinations of relevance have already been made (i.e., regarding what is relevant to circumscribing a situation). However, Fodor not only ignores possible heuristics that may not beg the relevance problem, but he fails to consider non-question-begging possibilities to individuate situations. It is possible, for example, that a heuristic can identify a situation as a certain type given a small set of cues, and this identification procedure can be completely based on local considerations. There is of course an accompanying risk of the heuristic being wrong in identifying a given situation, but (as was made evident in the previous chapter) heuristics provide no guarantees.

In an earlier work, Fodor (2000) gives a different, though equally bad, argument against heuristic solutions to relevance problems. In short, his argument is this: In deciding which heuristic to use for a given cognitive task, there is a meta-problem of deciding how to decide which heuristic to use. According to Fodor, not only is this the beginning of an infinite regress, but the meta-problem of deciding how to decide is faced with the relevance problems associated with holism and global inference just as the first problem is. However, here again Fodor fails to consider the possibility of heuristics being cued by a small set of abstract features of the problem (cf. Gigerenzer & Goldstein, 1999). This is not a failsafe plan and it will sometimes

available information would have to be computed for relevance. (If S is the set under consideration, the powerset is the set of all subsets of S including S itself (and the empty set, though this does not matter much for present purposes).) For an unencapsulated system with lots of information, this is an intractable task. Frugal processes, on the other hand, would consider only a small number of different subsets of the available information, hence remaining tractable.

[13]Roughly, the sleeping dog strategy is a rule to the effect of: *Do not alter (update) a representation unless there is explicit indication of change*, according to the dictum "let sleeping dogs lie".

misfire, but it would work nonetheless without invoking any regress. Or indeed, if we use my characterization of heuristics as cognitive procedures that can be expressed as rules that one reasons in accordance with (as given by H_3 in the previous chapter), heuristics need not be represented in cognition, but rather can be primitive processes, perhaps built into the architecture of cognition. And if this is the case, then there may not be any deciding which heuristic to use, and *ipso facto* no deciding how to decide which heuristic to use.

Thus, despite Fodor's worries, heuristics may very well play a crucial role in achieving frugality in cognition, as is certainly the case for our AI counterparts. At the very least, heuristics appear to circumvent the problem of delimiting the amount of information a system considers in its processing—heuristics can simply pick out an appropriately limited set of representations from the total set possessed by an agent according to certain parameters,[14] and the system can make use of them accordingly. There will be no guarantees that all and only relevant representations will be picked out (Samuels, 2005, forthcoming), but again, such follows from the nature of heuristics. Presumably, heuristics will somehow be attuned, via the specified parameters, to picking out mostly relevant representations, but there will be times when what is considered will happen to be irrelevant. Humans are fallible, so we should not expect perfection in their cognition. At any rate, it appears as if heuristics do indeed avoid relevance problems.

However, appearances can deceive, and we must therefore look at the details of these claims. To see if heuristics do in fact circumvent relevance problems, let us take a closer look at what is being proposed in response to Fodor's worries about a computational theory of central cognition.

In specifying the heuristics that make cognition tractable, Carruthers (2006a, 2006c) draws on the work of Gigerenzer and his collaborators (e.g., Gigerenzer et al., 1999). Carruthers appears to take Gigerenzer's analysis of fast and frugal heuristics as suggestive of the general processes that we might expect to find in human cognition to enable suitably frugal computations.[15] However, the domains of inference for which most of Gigerenzer's heuristics perform

[14] We will see some of these parameters below when I discuss in more detail how heuristics work.

[15] To be sure, Gigerenzer likewise believes that the majority of human inference and decision-making involves

best are not general enough to support the claim that the said heuristics are commonly employed for most human cognitive tasks. The tasks for which Gigerenzer's heuristics are best suited are decision tasks of choice between two alternatives, or of some value to be estimated among alternatives, and wherein the cognizer has limited familiarity of the alternatives. These tasks constitute only a small subset of the cognitive tasks humans commonly face (cf. Sterelny, 2003, p. 208). Even for those instances when we have to make choices or estimate values, we are not always forced to choose among pairs of alternatives—we are often faced with many options. Moreover, when we are faced with tasks of choice or value-estimation, we are usually confronted with things we not only recognize but that we are very familiar with. Indeed, many of our choices demand an intimate acquaintance with the alternatives (cf. Sterelny, 2003, 2006). It therefore seems as if Gigerenzer's proposed heuristics have only a limited role in our reasoning. This is in stark contrast to Gigerenzer and Todd's (1999) assertion that their research on heuristics is the study of "the way that real people make the *majority* of their inferences and decisions" (p. 15, my emphasis). This also undermines Carruthers' suggestion that Gigerenzer's fast and frugal heuristics can provide tractable processes for cognition *tout court*.

Samuels (2005, forthcoming), on the other hand, does not specify any heuristics that he thinks can ensure computational frugality in determining what representations get considered in a given cognitive task. He rests his arguments instead on an intuitive, though vague, appeal to web-search-engine-like heuristic or approximation techniques (Samuels, forthcoming; cf. Shanahan, 2009). The prospect of web-search-engine-like heuristics initially looks promising to deliver the desired frugality since actual web search engines are able to search through billions of websites in fractions of a second. Yet it is not entirely clear from Samuels' account exactly what role such heuristics would play in *relevance determinations* in human cognition. To be sure, web-search-engine-like heuristics can certainly search and retrieve information quickly, efficiently, and frugally, but these heuristics *per se* do not determine the relevance of

heuristics (Gigerenzer & Todd, 1999; see also Gigerenzer, 2007). But Gigerenzer's analysis extends to reasoning and inference *qua* central cognition, whereas Carruthers, in arguing for the massive modularity thesis (Sperber, 1994), suggests that Gigerenzer-style heuristics can be deployed by the various *modules* that purportedly compose central cognition. More on modularity below.

the information it retrieves. This is made apparent by Samuels' (correct) assertion that the representations that a heuristic will bring to bear on a given task will not be invariably relevant. To put the point more directly, web-search-engine-like heuristics may very well circumvent the computational relevance problem, but this solution does not address the epistemological relevance problem (cf. Shanahan, 2009).[16]

Certainly, one can point to the advances in AI, computer programming, and other real-world computational applications, and proclaim that there are real instances where a program successfully deployed heuristics without having to go through computationally taxing processing to determine relevance beforehand (much like how web search engines are real-world examples of success). However, this observation fails to acknowledge the fact that the AI researchers and computer programmers have already determined the relevance of certain information to certain tasks, and that they design heuristics to operate according to *those* determinations. This is what I take Fodor's (1987) point to be: In describing the frame problem as a problem of formalizing computationally relevant facts from irrelevant ones—and hence delimiting the candidates for belief-updating when certain events occur—Fodor asserts,

> The programmer decides, case by case, which properties get specified in the data base; but the decision is unsystematic and unprincipled. . . . [T]here appears to be no disciplined way to justify the exclusion [of properties that cause a system to update indefinitely many facts about the world,] and no way to implement it that doesn't involve excluding indefinitely many computationally relevant concepts as well. (Fodor, 1987, p. 33)

The question is not: How do AI and computer programs determine what information is relevant? We already know the answer to this, namely such information is *programmed into* them. The question is rather: How did the AI *researchers* and computer programm*ers* determine what information is relevant to what tasks in the first place? In the absence of an answer to this question, the epistemological problem remains. (And this is not to mention Fodor's worry about excluding indefinitely many computationally relevant concepts when implementing a program,

[16]We shall see in the next chapter, however, that there is an interesting theory of web search engine strategies expounded by Andy Clark (2002) that can play a role in addressing the epistemological relevance problem. This role will be made clear once we understand the cognitive architecture I develop in this chapter and the next.

which is a problem that also remains unaddressed.)

Both Samuels (2005) and Carruthers (2006a) also appeal to the notion of *satisficing* to explain how human cognition can be computationally tractable. However, such an appeal is empty as it stands. For satisficing is really a substitution for a stopping rule for search among alternatives—a rule that instructs search to stop once a certain goal is reached (see e.g., Simon, 1979). But relevance determinations involve more than just search. Even if we generalize the notion of satisficing (which seems to be the case in many discussions in cognitive science) to a procedure that arrives at solutions that are not optimal but are in some sense "good enough"—processing (e.g., evaluation) continues until some specified aspiration level is reached—the very process *that* satisfices remains unaccounted for. That is, asserting that a system satisfices says nothing of how the goals or aspiration levels are specified, nor of how it is determined when these goals or levels are reached; and these are the very problems of relevance under consideration. Certainly, heuristics can satisfice, but it is a further question how *they* make relevance determinations in their processing. On this Samuels and Carruthers are silent.

Fodor rightly points out that (what I am calling here) the epistemological relevance problem "goes as deep as the analysis of rationality" (1987, p. 27). We can see why—namely, because the demands to be satisfied in determining what is relevant just are the demands of rationality. To repeat, humans appear to be largely rational—in Fodor's words, "witness our not all being dead" (2008, p. 116)—and determining what is relevant is part and parcel of rationality. Samuels (2005, forthcoming) demurs at Fodor's suggestion that the totality of one's beliefs must be consulted in determining relevance, since this is the "only guaranteed way" (Fodor, 2000, p. 36) of classically computing a global property like relevance. As Samuels remarks,

> guarantees are *beside the point*. Why suppose that we always successfully compute the global properties on which abduction depends? Presumably we do not. And one very plausible suggestion is that we fail to do so when the cognitive demands required are just too great. In particular, for all that is known, we may well fail under precisely those circumstances that the classical view would predict—namely, when too much of a belief system needs to be consulted in order to compute the simplicity or conservativism of a given belief. (Samuels, 2005, p. 119)

Samuels is right that guarantees are beside the point (cf. Shanahan, 2009; Sperber & Wilson, 1996), and that we often fail in determining relevance. Nevertheless, we are still left to explain our successes in determining what is relevant to our cognitive tasks. And, again, our successes are not few. To paraphrase Fodor,[17] the moral is not that the heuristic solution to the frame problem—or to relevance problems more generally—is wrong; it is that the heuristic solution is empty unless we have, together with the heuristic strategy (or strategies), some idea of what is to count as relevant. Thus, despite the fact that heuristics can probably avoid the computational relevance problem, we are still faced with the epistemological relevance problem.

Let us briefly pause to see where we have gotten to. We have been discussing what heuristics are supposed to do for us in cognition. The part of cognition that is of concern is central cognition, since this is where problems of relevance arise. We saw that philosophers such as Samuels and Carruthers believe that heuristics offer a solution to problems of relevance.[18] I argued, however, that heuristics do not offer the kind of solution that Samuels and Carruthers think; specifically, given the distinctions of the kinds of relevance problems I described above, heuristics do not offer a complete solution to relevance problems, as they fail to solve or circumvent the epistemological problem.

We also saw that Fodor believes that heuristics fail as a solution to relevance problems in central cognition. Yet Fodor's rejection of the heuristic solution is not motivated by the concerns I gave. Rather, Fodor believes that typical tasks in central cognition require global assessments of information, requiring in the limit consideration of one's entire set of beliefs, and this is something that heuristics cannot do. If Fodor is right, then it is hopeless to look to heuristics as a solution to relevance problems in central cognition. Nevertheless, not only do I think that the demands of global assessment that Fodor believes is placed on central cognition are too lofty (I am not sure if anyone except for Fodor believes that central cognition is global in his sense; cf. Samuels, forthcoming), but, despite my arguments against Samuels' and

[17]"The moral is not that the sleeping-dog strategy is wrong; it is that the sleeping-dog strategy is empty unless we have, together with the strategy, some idea of what is to count as a fact for the purposes at hand" (Fodor, 1987, p. 31).

[18]Notwithstanding that Carruthers believes that central cognition is massively modular (see below).

Carruthers' suppositions about the work heuristics do for us, I believe that heuristics actually help in circumventing the epistemological worries associated with relevance problems. To see how this is so, we will have to gain a better understanding of the role heuristics play in cognition, and this means investigating more thoroughly how heuristics work and the structures they operate over.

In the next section, I will discuss the architecture of cognition in order to develop a view of central cognition which facilitates heuristic processes. The section that follows will be devoted to elaborating the kinds of cognitive structures that heuristics operate over (or exploit, as suggested in the previous chapter). In the final chapter of this dissertation I will attempt to address the epistemological relevance problem. I will argue that humans do not actually solve the epistemological relevance problem, and therefore the heuristic solution is empty. However, I will also show that the theory of cognition and how heuristics work that I advance in this chapter and the next suggests its own means of alleviating us of the epistemological worries.

3.2 Toward a general architecture

Heuristics are specified reasoning patterns, or, as explicated in the previous chapter, cognitive procedures. Some believe that heuristics are executed by specialized cognitive mechanisms, conceived as "modules" of the mind (e.g., Gigerenzer, 2000; Carruthers, 2006a). There are many ways that one can characterize a module, but there are two core features that are generally believed to facilitate fast and frugal reasoning (Samuels, 2005): *domain-specificity* and *informational encapsulation*.[19] These two properties were those (among others) that Fodor (1983) had originally used to define the input-output modules that appear to subserve our peripheral systems (paradigmatically, those responsible for visual and language processing). We have already discussed the role that informational encapsulation would play in facilitating fast and frugal reasoning, and in avoiding the computational and epistemological worries in cognition

[19]There is substantial debate over what constitutes a module and how to best characterize such properties. I want to avoid this debate, however. So I will rely on a relaxed understanding that entails at least that a module is a relatively autonomous component of the mind.

generally. Encapsulation is supposed to confine the amount of information that can be surveyed to a highly restricted proprietary database. In consequence, there is little computational burden associated with information search, and the information possessed by the proprietary database constitutes the one and only background against which relevance is determined.

Nevertheless, a defining characteristic of central systems is that they do not exhibit modular encapsulation. For it also may be recalled that a widely recognized feature of central systems is that they allow for the free exchange of information.[20] For example, when he discovered the chemical structure of benzene, Kekulé credited a dream he had of a snake biting its own tail (Hempel, 1966, pp. 15-16). Thus, for Kekulé—and for anyone, really—his dream of a snake bore in novel and important ways on his theorizing about chemical structures. All of us can understand this; in fact, we all can envision a snake biting its own tail and acknowledge its resemblance to a ring that is benzene's structure. The purported feature of our general reasoning that allows us to understand how images of snakes bear on chemical structures—two disparate representations, and *a priori* not obviously relevant to each other—is that any conceptual piece of information we hold in our head can bear on any other. But if general reasoning is subserved by modular encapsulated mechanisms, then it would seem, *prima facie* at least, and contrary to fact, that some conceptual information cannot be brought to bear on others.

Of course, this is not a decisive case against the possibility of encapsulated modular central systems. But it shows that there is room to doubt the plausibility of the thesis. This in turn casts doubt on the hypothesis that heuristics are executed by encapsulated modules in central cognition—if central cognition is not modular, then *ipso facto* modules would not be the structures that are executing heuristics in central cognition.[21] Yet, if heuristics are not executed by mental modules, then there is no indication that heuristics are informationally encapsulated

[20]This is a feature acknowledged by those who deny central systems modularity (e.g., Fodor, 1983, 2000; Samuels, 2000, 2006) as well as those who advocate for it (e.g., Carruthers, 2006a; Sperber, 1994).

[21]The present discussion is about the role of heuristics in central cognition. Though it is an interesting question what role(s) heuristics have in peripheral systems, this is not a matter that will be addressed in this dissertation. Therefore, whenever I speak of heuristics, I should be understood as referring to them as operations or procedures in central cognition only.

processes. In turn, if heuristic processes are not encapsulated, one might wonder how they can perform rapidly and frugally, and avoid the computational relevance problem. An appeal to heuristics would be circular; so without encapsulation it appears that we are left with the task of explaining how they work the way they do.

The solution might come by way of the other aforementioned feature of modules that is supposed to enable quick and frugal cognition, namely domain-specificity. Before we continue, it is important to note that there are two ways to understand domain-specificity, one in terms of *the function a mechanism computes* and the other in terms of *the range of representations a mechanism takes as input* (Samuels, 2005; Carruthers, 2006a). On the former understanding, a cognitive mechanism's domain is the *task* it is dedicated to performing, and on the latter the domain is the type of *content* or *class of representations* that it takes as input. Hence, these two senses of cognitive domain are often called *function*- or *task-domain*, and *content*- or *input-domain*, respectively. However, as we continue on, it will become apparent that my argument does not hang on which interpretation is adopted, although it will be more amenable to the content/input reading.[22] Nevertheless, it is important at this point that the notion of domain-specificity does not get confused with informational encapsulation or with frugality. Domain-specificity is about restrictions on what a system computes or what representations can serve as input to a system, depending on the interpretation. In other words, domain-specificity is about the subject matter computed or the kinds of representations the system can compute. In contrast, informational encapsulation has to do with *access*—what information a system has access to. A system might be informationally encapsulated, but the information that it has access to might span many domains (which would mean that the system can compute either different domain functions or representations that belong to different domains). On the other hand, a system might be domain-specific, but have access to quite a lot of information (regardless of domain), and thus be unencapsulated. Finally, let us remind our-

[22] Carruthers (2006a, p. 6, n4) suggests that the content/input reading may be tied to a function/task reading, insofar as an evolutionary story can be told about how the representations that a cognitive mechanism (or module) computes (constituting its content-/input-domain) are the ones that it was designed to process (constituting its function-/task-domain) (cf. Sperber, 1994).

selves that frugality is just a general notion of providing limits on the information processed in a system's computations. Frugality can be achieved by informational encapsulation or domain-specificity (or indeed, as we saw, by heuristics, which may not be encapsulated (at least for the case of central cognition) but may be domain-specific; more on this presently).

Returning to the matter at hand, within the framework of CTM, the potential benefits of domain-specificity and its role in cognition become especially apparent:

> if a mechanism is sufficiently domain specific, then it becomes possible to utilize a potent strategy for reducing computational load, namely, to build into the mechanism substantial amounts of information about the domain in which it operates. This might be done in a variety of ways. . . . But however this information gets encoded, the key point is that a domain-specific mechanism can be *informationally rich* and, as a result, capable of rapidly and efficiently deploying those strategies and options most relevant to the domain in which it operates. (Samuels, 2005, p. 111)

A domain-specific mechanism can thus be understood as an "expert system" (with respect to its domain), and can thereby be frugal in its computations.

Some researchers believe that humans possess a wide variety of domain-specific mechanisms, such as a "theory of mind" module which is dedicated to reasoning about the mental states and behaviour of people (e.g., Baron-Cohen, 1995). This module is understood to be domain-specific because the class of representations that it can process is restricted to a distinct kind—namely, intentional mental states. The purported theory of mind module can draw rich inferences about the beliefs, desires, and intentions of people without being computationally expensive by having a lot of information about intentional mental states built into the mechanism. The suggestion, then, is that heuristics can likewise be fast and frugal, not by being encapsulated, but by operating according to built-in information about the respective domains in which they operate.[23]

[23]Many evolutionary psychologists, such as Cosmides and Tooby (1992), Sperber (1994, 1996), and Baron-Cohen (1995), do not follow Fodor's characterization of modules. They reject the supposition that informational encapsulation is a core feature, and insist only that domain-specificity is essential to modularity. Thus, what is being suggested here with respect to how heuristics operate is consistent with such a conception of modularity. In other words, on the evolutionary psychology conception of modules, heuristics may very well be executed by modules. I will add, however, that I think that a conception of module that does not include informational encapsulation as a central feature is uninteresting. But I will not argue the point here (though see the following footnote).

However, we are faced with an ambiguity with respect to the *kind of system* that is domain-specific. This is because domain-specificity can be a property of either a *computational* system or a system of *representations*. The distinction between these two types of cognitive systems is made by Fodor (1983, 2000), and is acutely emphasized by Samuels (1998, 2000). A computational (cognitive) system is a (classical) device which takes symbols (or representations) as input and manipulates them according to formally specifiable rules to generate outputs, which may either be symbolic (or representational) or action-producing. These computational systems—or "computational modules"—may be presumed to be domain-specific insofar as they are capable of carrying out only a restricted class of computations or computing only a specific range of representations (depending on how one reads domain-specificity). On the other hand, a system of mental representations is simply a stored body of knowledge or, more generally, information which may be accessed by other cognitive systems. These systems of mental representations are individuated according to the specific (content) domains to which each is dedicated. They are typically thought to be innate, but need not be. Nevertheless, such systems encode various kinds of information about their respective domains. Both Fodor and Samuels refer to such systems as "Chomskian modules", so named because of Chomsky's hypothesis of innate Universal Grammar; however, other proposed innate domain-specific knowledge besides that of language, such as theory of mind, or so-called "folk physics" or "folk biology", qualify as Chomskian modules as well.

What is important to notice is that

> computational modules are a very different kind of mental structure from Chomskian modules. . . . In a sense, [Chomskian modules] are 'inert'. They only eventuate in behaviour when manipulated by various cognitive mechanisms. By contrast, computational modules are . . . mechanisms that *manipulate* representations. Computational modules and Chomskian modules are thus importantly different with respect to the functional roles that they play in our cognitive economy. (Samuels, 2000, p. 19)

It should be clear that, although Chomskian modules and computational modules alike can be characterized as domain-specific, Chomskian modules cannot be characterized as informationally encapsulated. For it is only a module *qua* mechanism that can properly be encapsulated; it makes no sense to claim that a system of mental representations has constraints on the amount of information it can access, since such a system does not access, or indeed *do*, anything. (We might think of Chomskian modules instead as *encapsulated information*.)

Samuels makes a further observation worth mentioning. He reports that computational modules can coexist with Chomskian modules. And indeed, a natural way of conceiving the operations of the mind is that the proprietary databases to which access to information is restricted for computational modules are just Chomskian modules. Nevertheless, this does not have to be the case. As Samuels emphasizes, Chomskian modules do not entail computational modules. He goes on to argue instead that we may very well possess a variety of Chomskian modules without possessing any domain-specific computational mechanisms at all. Indeed, he contends that it is possible that the "domain-specific knowledge [of Chomskian modules] is only utilised by *domain-general*, and hence non-modular, computational mechanisms" (Samuels, 2000, p. 19).

Let us return now to heuristics. Heuristics are certainly computational insofar as CTM holds. But if heuristics indeed operate according to built-in information about the respective domains in which they operate, the question is whether such information is built into an instantiating computational module or an accessed Chomskian module. It seems that there is nothing particular about the nature of heuristics that entails computational *modules*. For, mirroring Samuels' argument, it is possible that heuristics are instantiated by a domain-general computational mechanism which draws on domain-specific bodies of knowledge or information (i.e., Chomskian modules). This would permit heuristic procedures to be applied across many domains (i.e., they would be domain-general), and yet still operate quickly and frugally.

Carruthers (2006a, 2006c), in advancing a massive modularity hypothesis (i.e., the thesis that the mind is entirely, or nearly entirely, composed of individuated modules), argues for the possibility of multiple copies of the same *type* of heuristic instantiated by different modules. If this is right, then there can exist a number of different domain-specific computational modules, each dedicated to different tasks and carrying out a restricted class of computations or operating over a specific class of representations, but each employing its own heuristic of the same type. However, this possibility rests on an independent assumption about the architecture of central cognition—that it is massively modular. As explained above, however, the thesis that central cognition is subserved by a number of distinct computational modules is implausible.[24] Thus, in the absence of evidence to the contrary, there is no reason to reject the domain-general view of central systems and the execution of heuristics given here.

This is not to say that heuristics are not domain-specific independent of the mechanism that instantiates them. However, the heuristics that are of interest here—i.e., cognitive (inferential) heuristics—appear to be domain-general. For instance, Gigerenzer's Take the Best or Recognition heuristic, or Kahneman and Tversky's Representativeness or Availability heuristic, can operate over many different ranges of representations. That is, such heuristics themselves do not specify any content. It is not the case that Take the Best says

> (i) Search among alternative *cities*; (ii) stop search when one *city* scores on *whether it has a university*; (iii) choose the *city* identified in step ii;

or that Availability says

> The probability of *a plane crashing* is assessed according to the ease with which instances or occurrences can be brought to mind.

Rather, cognitive (inferential) heuristics are general procedures of inference that can be applied across many different domains—they can be employed in contexts of choosing which car to buy inasmuch as in contexts of deciding which city is larger; or in contexts of estimating the probability of plane crashes inasmuch as estimating the frequency of spousal abuse. In fact, the domain-generality of heuristics is what leads Gigerenzer to assert that the heuristics he

[24]The notion of "module" adopted by evolutionary psychologists suggests that the possibility that central cognition is subserved by a number of *un*encapsulated, domain-specific mechanisms (see previous footnote); in which case, Carruthers' argument would go through. However, I agree with Fodor that "it's informational encapsulation, however achieved, that's at the heart of modularity" (Fodor, 2000, p. 63). I therefore view any modularity thesis that foregoes informational encapsulation as a central property as no modularity thesis at all. In addition, there are a number of authors who argue that empirical evidence suggests the implausibility of modular central cognition (Prinz, 2006; Samuels, 1998, 2000, 2006; Sterelny, 2003; for a conceptual argument against central systems being subserved by a number of domain-specific mechanisms, see Aspeitia, Eraña, & Stainton, 2010).

studies are *robust*: they can be effectively applied to a variety of different (sometimes new and novel) situations and environments (Gigerenzer & Todd, 1999; Gigerenzer, 2004, 2006). At the same time, however, on the function/task reading of domain-specificity, these heuristics may be construed as domain-specific.

It is important to be clear about the scope of the claim. The claim being made is about the architecture of central cognition. This says nothing about the extent to which the mind in general is modular. Considerable empirical evidence supports the claim that there are a number of modular mental systems, which are both encapsulated and domain-specific, but these are most likely restricted to the peripheral systems (Fodor, 1983; Samuels, 2000, 2006). Furthermore, this says nothing about whether computational modules use certain heuristics in their processing. By all lights, it appears that peripheral modules do indeed utilize heuristics, such as perceptual heuristics as explicated in the previous chapter.

The claim being made here is that the use of heuristics in general reasoning does not entail a modular view of central cognition, and thus there is no apparent need to invoke domain-specific computational modules in explaining the use of heuristics and how they generally work. Instead, it is possible to characterize heuristics as cognitive procedures that are instantiated by a domain-general computational system, but operate with specialized bodies of domain-specific information (i.e., Chomskian modules). Once again, this is not a decisive argument against the possibility that heuristics are executed by computational modules in central cognition. But unless there is empirical evidence to the contrary, rejecting the domain-general view of central systems and heuristics is unmotivated.

3.3 Informationally rich structures: k-systems

If the above picture is correct, and heuristics are deployed by a domain-general computational mechanism in central cognition, two further questions arise: What is the nature of the representational modules heuristics exploit? and How does the suggested architecture facilitate heuristic reasoning? It is not obvious from the surface how heuristics work, since they are essentially simple algorithms; there is nothing in the heuristics themselves that specifies how they make use of informationally rich structures of representations. I will address these issues in turn in the present section and the next.

Let us begin by scrutinizing what Chomskian modules are and their role in the proposed architecture. Chomskian modules are essentially systems of mental representations. But Samuels (1998) and Fodor (1983, 2000) specify certain other features. One is that Chomskian modules are supposed to be bodies of mentally represented *knowledge*. Of course, there are different sorts of knowledge, declarative (or propositional) and procedural being most commonly distinguished. But the kind of knowledge that Samuels and Fodor are referring to is *truth-evaluable* knowledge. This implies, at least, that the representations of interest have propositional content. The importance of this is evident vis-à-vis CTM: Computations are transformations of representations which respect semantical relations, such as implication and logical consequence; and, in Fodor's words, "It is . . . a point of definition that such semantic relations hold only among the sorts of things to which propositional content can be ascribed" (Fodor, 1983, p. 5).

Another feature of Chomskian modules, according to Samuels, is that they encode various kinds of information about their domain. This, as we saw, is how Chomskian modules are informationally rich. But I think we can add to this that the information encoded in a Chomskian module is encoded in a highly organized fashion. A domain-general mechanism can take advantage of generous amounts of domain-specific information encoded in a Chomskian module, and thereby enable quick and frugal cognition. But if the information is structured or organized in specific ways, it would enable quicker and more frugal cognition. Contrariwise, if such information is not structured or organized in specific ways, it would retard the speed and efficiency of search and processing. Characteristic human performance on many cognitive tasks suggests that the structures heuristics exploit are distinctly organized to produce robust reasoning and inference patterns.

Moreover, it would seem that Chomskian modules must not only have a specific internal structure, but also bear specific intra-system relations, which would further facilitate fast and frugal reasoning. Reasoning within one domain, such as “folk psychology”, often bears in a number of ways on other domains, such as “folk biology”. For example, in deciding what intentions to ascribe an organism, one must make inferences about the kind of organism to which one is ascribing intentions. Without such connections between bodies of knowledge, we would not be able to make the rich inferences we characteristically do. I will return to these issues concerning informational organization below.

Samuels introduces a further feature in response to the possibility of counting any domain-specific collection of truth-evaluable representations as Chomskian modules: “We do not want, for example, to treat a child’s beliefs about toy dinosaurs as a module” (Samuels, 2000, p. 18). Thus, Samuels asserts that Chomskian modules are innate. Along with Fodor, Samuels understands Chomskian modules to be the very kind of system that Chomsky’s Universal Grammar is supposed to be (hence the namesake), and Chomsky’s Universal Grammar is supposed to be *inter alia* innate. However, I do not see why heuristics, or any domain-general operation for that matter, must be conceived to exploit only innate domain-specific systems of knowledge. It is quite conceivable that there is at least some acquired knowledge contained within many of our domain-specific systems of knowledge.[25] Furthermore, there is no principled reason why there cannot be entire systems of acquired information that is stored and organized in a fashion similar to innate domain-specific knowledge, and exploited by domain-general systems just as innate systems are.

Certainly, the way something like Universal Grammar is supposed to get unpacked during childhood development and early language learning cannot be emulated by acquired bodies of knowledge. However, the architecture suggested above and its operations are not about cog-

[25] “Acquired” should be read as “acquired in consequence of experience”, thus contrasting “innate”. Learning is sufficient for acquiring, but it is not necessary. One can be hit on the head and thereby acquire beliefs, though these are not learned beliefs. We might hesitate calling non-learned acquired beliefs *knowledge*, since such acquired beliefs may not meet a criterion of justification. However, I want to avoid speaking of learned knowledge to avoid issues concerning what constitutes learning. Fodor (2008), for instance, believes that learning consists in “a rational process like inductive inference” (p. 145). Yet I do not want to suggest here that the systems of knowledge to which I am referring in the text are learned in this way. In fact, I do not want to commit to any particular account of learning.

nitive development, but about cognitive *deployment*. The discussion that took place there, as well as the discussion provided by Samuels (1998, 2000), concerns how the mind can do what it does given a domain-general architecture of central cognition. Nothing was said, nor need be said, about the poverty of stimulus with respect to how we know the grammar of language, or conceptions of (folk) psychology, or (folk) biology, or whatever. Thus, it seems perfectly acceptable within the foregoing cognitive architecture that a domain-general computational device can utilize informationally rich, acquired domain-specific systems of knowledge. If this is right, then I believe it is necessary to forego the Chomskian namesake to refer to such informationally rich representational structures.

In addition to dropping the innateness requirement, I propose to jettison the domain-specificity requirement. This is not as bold of a move as it may appear. It seems that what is really doing the work in facilitating fast and frugal cognition is not the domain-specificity *per se* of a system of representations, but the kinds of information encoded in these systems and the manner in which it is encoded. For, to partially echo what was said above, it is possible that one can have an unorganized system of lots of domain-specific information, but it will be doubtful that a domain-general computational mechanism would exhibit the same speed and efficiency operating over this unorganized body of knowledge as it would operating over a highly structured system. Moreover, it is possible that one can have a domain-specific body of knowledge that is quite impoverished; and although a domain-general device would likely be able to operate quickly and frugally over such a system, owing primarily to the little information that ever gets considered, the inferences made would not be as robust and accurate as those made by operations over richer systems.[26] That a system of information is dedicated to a specific domain no doubt contributes to the extent to which the system is organized, since there are natural relations among items of information within a domain. However, domain-specificity *per se* does not seem to be necessary for a body of knowledge to be structured and organized in ways conducive to fast and frugal exploitation by a domain-general computational mecha-

[26]Notwithstanding that accuracy is not a general feature of heuristics, as illustrated by the work of Kahneman and Tversky.

nism. To be sure, it is conceivable that someone has an exceeding amount of knowledge about the French Revolution, and thereby has built a non-domain-specific system of representations on this topic, which is highly organized and informationally rich.[27] When drawing inferences about or relating to the French Revolution (via her domain-general reasoning mechanism), this person would almost certainly carry them out quickly and frugally (and probably robustly and accurately too).

To summarize, a domain-general computational mechanism can deploy suitable strategies (such as heuristics) to enable quick and frugal cognition by exploiting informationally rich systems of representations, or truth-evaluable knowledge. These latter systems encode various kinds of information in a highly organized fashion; not only is there internal organization, but there is also intra-system organization that exhibits specific relations between systems. But unlike Chomskian modules, these systems have to be neither innate nor domain-specific. Instead, they can be acquired and/or non-domain-specific, so long as they possess the requisite structure and organization. This should not be taken to mean that the systems of representations under consideration can be domain-general. A domain-general collection of representations would not bear the kinds of organization and relations that I take to be characteristic of such informationally rich systems. Working memory may be a domain-general system capable of storing any kind of representation, but it may store, at the same time, completely unrelated representations. For a system of representations to be non-domain-specific, on the other hand, simply means that it fails to pick out a proper domain as cognitive scientists usually understand the term (which is closely related to natural kinds).

I already suggested dropping the Chomskian namesake to refer to such non-domain-specific systems. But since domain-specificity is considered by most, if not all, theorists to be essential to modularity, it appears that "module" must also be omitted. Hence, for lack of a better name, I will call these informationally rich systems of representations presently under consideration

[27] Few cognitive scientists, if any, would consider the French Revolution as a proper domain, since the class of representations that belong to such a topic do not pick out anything resembling a natural kind. I do not want to enter into any debate over what constitutes a domain. The reader who is dissatisfied with this can substitute any non-domain-specific example. See Hirschfeld and Gelman (1994) for debates on what constitutes a domain.

k-systems (where "k" stands for knowledge). We shall see in the next chapter that the knowledge that k-systems are supposed to embody does not necessarily have to be truth-evaluable. Indeed, as I will argue, perceptual information can be, and often is, included within the systems of representations over which a domain-general computational device can operate. For now, however, it will do no harm to assume that such knowledge is truth-evaluable, but it should be kept in mind that this is only part of the story.

K-systems are what heuristics partially owe their speed and efficiency to, but k-systems are also what heuristics owe their robustness and potency to. To be sure, heuristics are robust and potent strategies insofar as they can be applied in a vast number of contexts while producing satisficing inferences. The way I suggest that heuristics can perform like this is by being sensitive to specific parameters that are built into k-systems. For example, a k-system for chess might contain numerous parameters for openings, endgames, the pieces and their positions on the board, etc. Depending on what the task at hand is, certain representations will be thereby more readily available and more easily brought to bear than others. And indeed, the parameters may affect how the present game is *represented*. Other k-systems that exhibit similar parameters will be amenable to the same heuristics and strategies deployed by the central system for chess—for example, k-systems for other chess-like games, or even for activities such as battle or sports that bear certain relations to chess. The parameters possessed by k-systems thus control the flow of information, but more importantly, they partly constrain the operations performed by the domain-general computational device. More specifically, I claim that k-systems have control parameters that partly constrain how the information possessed by the system can be manipulated; in terms of what is of interest here, the control parameters partly constrain the way heuristics operate, or equivalently, what heuristics are brought to bear. These parameters embedded in k-systems are either acquired or innate, depending on the nature of the k-system in question.[28]

[28]We might note a similarity to Marvin Minsky's (1975) notion of *frames* (not to be confused with the frames of the frame problem, though there is a relation). In Minsky's words, "A frame is a data-structure for representing a stereotyped situation, like being in a certain kind of living room, or going to a child's birthday party. Attached to each frame are several kinds of information. Some of this information is about how to use the frame. Some

Cognitive tasks will also have certain structures of their own. For instance, a task of deciding between two alternatives will have a specific structure that enables comparisons between the two alternatives on a number of dimensions, and the consideration of relevant representations and other information. Such a decision structure will be different from the structure of, say, estimating the probability of an event, which might enable the recall of past occurrences and perhaps other knowledge about probability and frequencies. Moreover, the content under consideration will influence the structure of the task. For a decision of choosing which car to buy will not share the exact structure with a decision of which move to make in chess, or with a decision of whether to accept a job offer. My suggestion is that the structures of cognitive tasks will help in cuing the activation of particular k-systems. Thus, it is in conjunction with the informational structure of the task that the control parameters of the activated k-systems constrain what heuristics are deployed; in the right conditions, the informational structure of the task and the activated k-systems' control parameters *determine* what heuristics are deployed.[29] In this way, a domain-general device can deploy suitable task-specific heuristics without being task- or domain-specific itself. This is a lot like how a Universal Turing Machine can simulate any specific Turing machine given the appropriate input. Thus, the claim here is that heuristics are powerful inference strategies in the same way that Universal Turing Machines are powerful computing devices.

There will be instances, however, when k-systems will lack the appropriate control parameters. This can occur especially with impoverished k-systems. Yet this does not mean that the k-systems in question will lack control parameters altogether, although this can be the case. When a k-system lacks the appropriate control parameters, the speed, efficiency, and/or accu-

is about what one can expect to happen next. Some is about what to do if these expectations are not confirmed" (p. 211). The view I am advancing, however, is different from Minsky's frames (or the related notion of *scripts*). K-systems are not to be understood in terms of stereotypes or paradigms, and not in terms of "situations" either. As we shall see in the next chapter, k-systems are to be conceived as *concepts*. Moreover, the function of frames (or scripts) in cognition is supposed to be quite extensive, having a role in the processes of vision and imagery, linguistic and other kinds of understanding, and memory storage and recall. On the other hand, I am not making such lofty suggestions here, as I am confining my account of k-systems to facilitating heuristic processes (though future work might extend this to include other aspects of cognition).

[29]This is another possible way to avoid Fodor's (2000) worry about an ensuing infinite regress when employing heuristics, as outlined above.

racy of inferences are negatively affected. In such instances, heuristics may be deployed based on the structure of the problem alone, or default heuristics may be deployed, or no heuristics will be deployed at all. In some of these instances, I believe, we witness the mistakes and biases that Kahneman and Tversky (and their followers) emphasized and researched. In fact, I believe that such mistakes and biases result from impoverished k-systems generally, in terms of the kinds, amount, accuracy, and organization of information. But I will save discussion on this issue for chapter 5.

In sum, k-systems influence the kinds of computations that operate over them in central cognition via the information contained in the k-system along with the relations within or between them. What is entailed is that the operations of heuristics exploit ordered and structured systems of knowledge.

Let us now switch gears and consider a more general view of how heuristics exploit informationally rich structures and of the role of k-systems in cognition. This will be my task in the next section. In the chapters to follow I will return to the matters of what kind of entity k-systems are, and *how* heuristics exploit them.

3.4 Exploiting informational richness

So far I have expounded a general cognitive architecture for central systems, and discussed how such an architecture can facilitate heuristics or heuristic reasoning. What I have been describing is consistent with the characterization of "heuristic" presented at the end of the previous chapter, viz.

> **H**$_7$ Heuristics are cognitive procedures that satisfice, and that require little cognitive resources for their recruitment and execution; they operate by exploiting informational structures.

I will now discuss in more detail the role of k-systems in cognition, and more specifically how heuristics can harness the informational richness of k-systems. I will argue that a lot of information does indeed bear on many of our cognitive tasks, but by the end of this section it will become clear how heuristics remain frugal in their processing. This sounds paradoxical, but as we shall see, frugality in information-processing can be realized in the appropriate management of the information that bears on cognitive tasks.

3.4.1 High information load cognition

Let us consider a common problem that someone may face. Alice is deciding what to do on Saturday night. There are a number of viable options: go to a movie, stay in to watch a movie, go out to a pub, go out dancing, go out for dinner, dinner and dancing, dinner and a movie, stay in for dinner, stay in for dinner and go out for a movie, etc. Alice's choice will be guided by her multiple beliefs, desires, and preferences, as well as by a number of external factors: What is playing at the local movie theatre? Is there any particular movie she has been wanting to see? Is there any movie that she can rent that she has been wanting to see? Does she want to stay in? What did she do last Saturday night? What are her friends doing? Does she want to be with her friends? Which friends? If Alice does in fact want to spend Saturday night with her friends, she must participate in coordination problems: Would they all be happy to go dancing? Would some rather go to a movie while others would rather stay in for dinner? Are there any friends with which Alice prefers to spend time over others? Will she want to go out to a movie with Bob, even though he has an annoying habit of talking throughout a film? Will she want to invite Carol, even though she suspects that Carol will invite Ted, whom Alice dislikes? Alice might have to consider her financial situation: Can she, and does she want to, spend the money to go out for dinner and a movie? Do her friends? Alice will also have to consider her other obligations: Does she have to be somewhere important early Sunday morning? This will bear on how late she would want to stay out; whether she would rather have a quiet night in; etc. Much more can certainly be added here for Alice to take into consideration, but the point should be clear that a lot of information must be used in making typical, everyday decisions such as these. Indeed, they are problems of the ill-defined variety discussed in the previous chapter, so it is no wonder why navigating through them can be a complex thing to do.

Gigerenzer would have us believe that such a decision can be made on one reason. For instance, he suggests that Darwin may have decided to marry based on one reason (namely, having a constant companion), despite the fact that Darwin had written out a list of reasons for why he should or should not marry, so as to help him in his decision process (Gigerenzer & Todd, 1999). Gigerenzer observes that Darwin's pro/con list contained incommensurable items. How does one, for instance, compare "having a constant companion" to "if many children forced to gain one's bread"? Such incommensurability is why Gigerenzer claims that a one-reason decision-strategy is needed—a strategy that would forego the effort of attempting to convert all the reasons into some common commensurable scale. Likewise, Gigerenzer would likely claim that the multiple issues that Alice must consider are incommensurable, and thus so too would her situation call for a one-reason decision-strategy. One-reason decision-making ensures that reasoning is frugal. (Cf. the discussion regarding problems of comparing incommensurability in the previous chapter (section 2.2.2).)

Nevertheless, I believe that Gigerenzer is overlooking the potential of human cognition when dealing with issues or problems that are cognitively demanding, such as those faced by Alice and Darwin alike. And in so doing, Gigerenzer is making the converse error that he accuses standard statistical models and theories of rationality of committing, namely of making too many simplifying assumptions about the problems to which they are applied. Gigerenzer and Todd (1999) assert that "the way information is structured in real-world environments often does not follow convenient simplifying assumptions" (p. 19). The converse error that I accuse Gigerenzer of making (and to some extent his colleagues as well) is making simplifying assumptions about human cognition. In a way that paraphrases Gigerenzer and Todd, human cognition often does not follow convenient simplifying assumptions, and it would be folly to presume so. For as Alice's situation demonstrates, generous amounts of information are in fact brought to bear on many of our decisions and inferences. Therefore, we should not make simplifying assumptions about the way humans reason, but instead look to model human cognition in a way that accounts for how humans manage such informational complexity.

To see what I mean, let us consider an argument recently put forward by Kim Sterelny (2004, 2006). Sterelny claims that the majority of real-world problems have a *high information load*—they are informationally demanding, requiring significant amounts of information. Sterelny believes that problems that bear a high information load, as complex as they are, are typical rather than extraordinary, especially in a social world:

> Human decision-making often has a high informational load, for we depend on knowledge-intensive methods of extracting resources from our worlds. . . . [H]uman social worlds are complex, demanding, and only partly cooperative. They are complexly structured: divided by gender, status, occupation, generation. They are operationally complex: much human action requires coordination with others. And they are complex in their resource demands: successful human life requires access to a large range of goods, not just a few. (Sterelny, 2006, p. 221)

Yet humans are impeccable at handling these high-cognitive-load problems in a quick and frugal way. The question then is: How do we manage to do it? This is, of course, an incarnation of the frame problem discussed above.

Sterelny argues that our uncanny ability to respond successfully to problems with high information loads can reside in modules—i.e., informationally encapsulated, domain-specific computational mechanisms. But such a solution is restricted to problems that are predictable and occur in a stable environment over evolutionary time. Thus, while Sterelny concedes that we likely possess innate modules dedicated to certain capacities, such as visual perception, language, and perhaps even folk physics, he goes on to claim that "human environments are heterogeneous in space and time, and as a consequence there are many high-cognitive-load problems that we face whose informational requirements are not stable over evolutionary time" (p. 218), as described in the passage above. This means that it would be difficult for natural selection to have equipped humans with specialized mechanisms to determine relevant information for the unstable, heterogeneous problems typical of human life.

Nevertheless, Sterelny believes that humans deal with high-cognitive-load problems by relying on a distinctively human evolved strategy: *epistemic niche construction*. Epistemic niche construction is evident in humans' propensity to develop *epistemic technology*.[30] Humans epis-

[30]Sterelny is actually picking up and responding to the ideas of Clark (1999, 2002b; Clark & Chalmers, 1998),

temically engineer their environments "in ways that circumvent the cognitive limits of individuals on their own" (p. 229). This is a special form of niche construction: it is *informational*, which consists in either the transformation of the informational structures of the problems that are to be solved, or the creation of "cognitive tools" to "enhance the capabilities of our naked brains. . . . To take the simplest of examples, the practice of marking a trail while walking in the bush converts a difficult memory problem into a simple perceptual problem" (Sterelny, 2006, p. 225; see also Sterelny, 2003, 2004).[31] Sterelny further points out that epistemic niche construction can consist in the use of epistemic technology to store and organize information in our environment. External representations, such as pictures or words, are obvious examples. But, perhaps more interestingly, artifacts can encode other kinds of information, such as their function; or they can act as exemplars or templates, and in this way carry information about how to make others like it (cf. Mithen, 2000; Dennett, 1991). In short, "we alter our environment to ease memory burdens. We store information in the environment; we recode it; and we exploit our social organization through a division of intellectual labour" (Sterelny, 2006, p. 230).

Sterelny concludes that "Good decisions require access to, and use of, generous amounts of information" (p. 233), and therefore much of human decision-making really carries high information loads. Even the use of epistemic technology is informationally demanding, according to Sterelny. (I will return to this point presently.) However, instead of speculating about innate cognitive mechanisms to assist us in our high-cognitive-load problem-solving and decision-making, he claims that most of our cognitive abilities are the products of learning how to exploit the right kinds of information in the right ways. This requires structured learning, or "scaffolding", to mitigate informational burdens, and to allow us to harness what is realized by epistemic technology. Thus, according to Sterelny, human competencies that many evolutionary psychologists (such as Gigerenzer and Carruthers) believe to be the products of innate

Dennett (2000, 1993), and Mithen (2000). I am choosing to focus on Sterelny's position in particular because he attends to some of the central points relevant to this dissertation in a more direct fashion than Clark, Dennett, and Mithen.

[31]Similar ideas are found in Dennett (2000).

mechanisms can be the result of perceptual preadaptation to relevant features of the natural environment (inherited from innate perceptual tuning to certain objective features of the world), plus structured training to be sensitive to such features by a community or culture (i.e., social learning); social learning will direct a learning-individual's attention to the relevant salient features in the environment (through, for example, cultural representations such as pictures or words), thereby further entrenching their respective roles in cognition.

It is interesting to note a certain parallel between Sterelny's position and Gigerenzer's understanding of how heuristics work, as outlined in the previous chapter. Gigerenzer appears to recognize that humans are informational niche constructors, though he does not use such terms. For example, he advocates for changing the way information is externally represented (such as probabilistic information) in order to facilitate correct or better inferences and judgments (e.g., Gigerenzer, 1991, 1996). Further, with respect to his notion of ecological rationality, Gigerenzer asserts, "From an ecological view, environmental structures … directly influence cognition and can be designed to improve it" (Gigerenzer, 2006, p. 128). In another, more recent work, he speaks of "design[ing] human environments" (Gigerenzer, 2008c, p. 27) to support better decisions (cf. Gigerenzer, 2010). Nevertheless, not only are Gigerenzer's discussions on the actual phenomenon of informational niche construction too brief and underdeveloped for what the topic deserves, he fails to appreciate the resulting informational richness of such designed environmental structures. Yet his fundamental features of heuristics—that they (i) exploit evolved capacities, and (ii) exploit informational structures in the environment—appear to correspond to ideas advanced by Sterelny. In particular, the exploitation of evolved capacities appears to correspond to what Sterelny believes are perceptual preadaptations to relevant features of the environment, which are requisites for many human competencies. Furthermore, according to Sterelny, humans use these perceptual preadaptations to exploit informational niches and epistemic technology, which are really just special kinds of informational structures in the environment. This is to say that, on Sterelny's account, some of the environmental structures that Gigerenzer talks about are human-engineered materials and niches for the purposes of exploitation.

It should be noted, however, that Gigerenzer refers to *naturally occurring* informational structures in the environment, and how heuristics are mechanisms adapted to specific environments. Although epistemic technology is, by definition, human-made and therefore not naturally occurring, Sterelny's broader story about the evolution of cognition is compatible with Gigerenzer's, so long as the environmental structures in question are stable enough to invite specialized adaptations. In sum, Gigerenzer believes that heuristics work very well for us—they are ecologically rational—just because they exploit environmental structures, and Sterelny gives an account of precisely how this is done for a certain range of environmental structures, namely through epistemic technology along with structured learning which finely tunes our awareness to these structures and our skills to exploit them.

This is not to say that Sterelny's position is continuous with Gigerenzer's. For Gigerenzer believes that the ubiquity of heuristics makes much human decision-making informationally cheap, whereas Sterelny's point is that this is not so. As we shall see, my view agrees with Sterelny's. If Sterelny is correct in that much of human cognition bears a high information load, then contrary to the common conception about the nature of heuristics, they actually draw on and use a great deal of information. However, I will argue that heuristics are still computationally frugal despite the high information loads. This is one respect in which my view is importantly different from both Sterelny's and Gigerenzer's—whereas Gigerenzer is too optimistic with respect to the amount of information heuristics process, Sterelny does not see heuristics as providing computational frugality. Moreover, both Sterelny and Gigerenzer emphasize the information in the environment that is exploited by humans—information manifested in epistemic technology for Sterelny, and information embodied by "environmental structures" for Gigerenzer. I claim, on the other hand, that the informational structures that humans exploit are *in the head*—specifically, information embodied by k-systems.[32] We

[32]Sterelny (2004) emphasizes the elaboration of mental representation as a result of the development and use of epistemic technology. In the next chapter we will see that my view is along the same lines. However, we will also see that my view goes several steps further than Sterelny's in that I develop a thesis about the cognitive structures that embody our mental representations (what I am here calling k-systems), and I argue that the informational

shall see more of this presently. But this does not mean that my view is incompatible with Gigerenzer's or Sterelny's. Rather, I understand my view to complement theirs. To see this, first note that Sterelny's account is not so much about how humans *exploit* informationally rich structures as it is about how humans *create* ways to store and organize information. Sterelny certainly advocates for the role of special cognitive preadaptations (i.e., modules) and structured learning in honing our abilities to exploit epistemic technology. However, he is silent on the underlying mechanisms by which we harness such informationally rich structures. On the other hand, as recently pointed out, Gigerenzer believes that simple heuristics altogether avoid the need to utilize significant amounts of information in decision-making. In these respects, then, my account offers additions to Sterelny's and Gigerenzer's views, and thereby suggests a more comprehensive understanding of human cognition.

Let us return to the central topic of this section. If we are to have any hope in exploiting the informationally rich structures manifested in epistemic technology or environmental structures in the ways that we in fact do, we had better have the appropriate cognitive wherewithal to do so in the first place. What this partly entails is that we had better have the cognitive wherewithal to conceptualize epistemic technology as things that we can exploit in very specific ways. A fish trap is an epistemic technology that *inter alia* encodes information about its function (Sterelny, 2006; cf. Mithen, 2000). But if one cannot recognize the fish trap *as* a fish trap, then one cannot exploit its function (or at least, one could only do so by accident). Instead, one may only be able to exploit it *qua* some decorative artefact, rather than as a specific tool for a specific purpose. To recognize and conceptualize a fish trap as a fish trap, one needs to have specific knowledge about fishes, traps, hunting, bodies of water, and so on. In other words, one needs to have a specific k-system (or a set of k-systems) from which to draw in cognition. In yet more general terms, how we exploit epistemic technology is *via the k-systems* we possess.

This view is Dretskean in an important sense. In developing his information-theoretic epistemology, Fred Dretske commented that there is "a *relativization* of the information contained

organization of these structures is what enables heuristic processes. Also see below.

in a signal because *how much* information a signal contains, and hence *what* information it carries, depends on what the potential receiver already knows about the various possibilities that exist at the source" (Dretske, 1981, p. 79). Although we might not know (or even cannot know) absolute measures associated with the amount of information generated by an event or carried by a signal, Dretske believes that *comparisons* can be made, "in particular comparisons between the amount of information generated by the occurrence of an event and the amount of information a signal carries about that event. . . . For example, if I tell you that Denny lives on Adams Street in Madison, Wisconsin, I give you *more information* than if I tell you, simply, that he lives in Madison, Wisconsin" (p. 54). With respect to the epistemological account I am suggesting, we might consider epistemic technology as sources of information, and, of course, the k-systems a person possesses determine what one knows about the source. Thus, what a person can learn from epistemic technology depends on what one already knows about it. If I give you a fish trap, and you know it as a fish trap (i.e., you can conceptualize it as a fish trap), then you can get more information from it (to manipulate, for heuristics to exploit) than if you do not know it as a fish trap.

What I am suggesting, then, is that, contra Sterelny and Gigerenzer, heuristics do not exploit environmental structures directly. Rather, heuristics exploit k-systems within which substantial amounts of information about the environmental structures in question are built, and this is how we manage high-information-load cognition. That is, the information embodied by k-systems, along with the extant relations within and between k-systems, are what enable heuristics to operate as they do; without the information and relations within and among k-systems, heuristics certainly would not be as powerful and robust as they in fact are. Epistemic technology and other environmental structures certainly can store and organize information, and thus exist as informationally rich structures, but cognitive exploitation occurs via what is known about the objects or events in concert with the problem(s) in question. We will get a better idea of how this works in the next chapter.

3.4.2 Exploiting k-systems, but still frugal

Before concluding this section, I will return to the issue of frugality with respect to heuristics. Notwithstanding that little cognitive resources are required to deploy heuristics, it appears that the foregoing account suggests that heuristics can entail the processing of quite a lot of information, since heuristics operate over informationally rich k-systems, and since heuristics operate in high-information-load cognition. All sorts of information and assumptions can bear in a variety of ways in much of our reasoning. We already saw this with respect to central cognition, and we discussed its implications with respect to incurring relevance problems. But on the foregoing account of central cognitive systems, it seems not only that a lot of information *can* bear on a given cognitive task, but a lot of information *does* bear. If so, this would run against what most researchers (most notably Gigerenzer and his colleagues) believe of the nature of heuristics—i.e., that heuristics are frugal processes, requiring little information in their processing—as well as against what is generally believed to be necessary to avoid computational relevance problems—i.e., that the computational burdens associated with relevance assessments can be relieved only if little information is processed, as outlined above.

It is instructive to consider here Dennett's (1984) account of the frame problem. He offers as an example a mundane task of making a midnight snack (complete with beverage):

> Now of course I couldn't do this without knowing a good deal—about bread, spreading mayonnaise, opening the fridge, the friction and inertia that will keep the turkey between the bread slices and the bread on the plate as I carry the plate over to the table beside my easy chair. I also need to know about how to get the beer out of the bottle into the glass. Thanks to my previous accumulation of experience in the world, fortunately, I am equipped with all this worldly knowledge. . . . Such utterly banal facts escape our notice as we act and plan. (Dennett, 1984, p. 134)

Dennett continues:

> my midnight-snack-making behaviour is multifariously sensitive to current and background information about the situation. The only way it could be so sensitive . . . is for it to examine, or test for, the information in question. The information manipulation may be unconscious and swift, and it need not (it better not) consist

> of hundreds or thousands of seriatim testing procedures, but it must occur somehow, and its benefits must appear in time to help me as I commit myself to action. (Dennett, 1984, p. 138)

> [The frame] problem concerns how to represent (so it can be used) all that hard-won empirical information .... Even if you have excellent knowledge (and not mere belief) about the changing world, how can this knowledge be represented so that it can be efficaciously brought to bear? ... A walking encyclopedia will walk over a cliff, for all its knowledge of cliffs and the effects of gravity, unless it is designed in such a fashion that it can find the right bits of knowledge at the right times, so it can plan its engagements with the real world. (Dennett, 1984, pp. 140-1)

In this light, the frame problem can be understood as the problem of how cognition achieves the informational organization, and enables access to the relevant information, that seems to be required for human-like cognitive performance. Interestingly, this idea was intimated by Fodor (1975) in *The Language of Thought*, though his purposes for broaching the issue were slightly different (though related). He there commented that "a fundamental and pervasive feature of higher cognitive processes [is] *the intelligent management of internal representations*" (p. 164, emphasis in original). In light of these considerations, we have yet a further aspect of the frame problem—what I will call the *representational frame problem*:

> **The representational frame problem** The problem of how a cognitive system embodies the informational organization, and enables access to the relevant information, that seems to be required for human-like cognitive performance.

I am claiming that k-systems are precisely the cognitive structures that exhibit the requisite informational organization—that appropriately represent knowledge—to ensure access to the right information for given cognitive tasks. And heuristics exploit such informational organization in their operations, effecting fast, on-the-fly cognition with reasonable levels of success.

At the same time, despite the exploitation of generous amounts of information, nothing of what I have said implies that heuristics are not frugal. Frugality is (in a certain sense) the essence of computational tractability, without which the computational relevance problem threatens. Moreover, on the characterization given in the previous chapter, heuristics by their very nature satisfice. And this entails that most information will not get considered in their processing; that is, satisficing and frugality go hand-in-hand. How, then, can heuristics' frugality be reconciled with their being informationally taxing? The answer, I believe, lies in the way information is organized and encoded in informational structures, such as k-systems and epistemic technologies. It is these epistemic structures that shoulder the informational burden, not the heuristics themselves. On my account of the architecture of central cognition, highly organized epistemic structures (or systems) bear dense relations among its constituents, and in addition, there are specific relations between epistemic structures. This is to say that the architecture of k-systems themselves affect information flow. When one part of a structure is utilized by a heuristic, there will be implications for other constituent parts (or other systems), depending on the organization of the structure and the nature of the relations among its parts, or to other systems. This might mean that other items of knowledge are activated (to a greater or lesser extent) when a heuristic exploits one (or more) items in a given k-system.

For example, Gigerenzer's Take the Best heuristic is a simple algorithm that produces a decision solely based on a single discriminating cue.

(15) (i) Search among alternatives; (ii) stop search when one alternative scores higher on some predetermined criterion; (iii) choose the object identified in step ii.

The specifics of Take the Best was presented at the end of the previous chapter, where I argued that heuristics are more than mere rules of thumb. But let us here remind ourselves of the details of the heuristic as I draw implications for its frugality. Take the Best assumes that an individual has a subjectively ranked order of cues stored in her memory which are believed to discriminate between objects, and which are sequentially searched to make predictions.[33] The highest ranked cue which discriminates is believed to be "the best", and the object which has a positive value with respect to this best cue is chosen or assumed to score higher on some criterion.[34] The idea behind this heuristic is something like the following: Suppose you had to

[33] I suppose that one can produce cue rankings on the fly, as the situation arises. However, according to Gigerenzer, we must have predetermined cue rankings stored in memory (Gigerenzer & Goldstein, 1996, 1999). For further discussion see chapter 5 (section 5.2.2).

[34] Take the Best is supposed to apply only when there are two alternatives, but I believe that the heuristic can be generalized to apply to situations in which there are more than two alternatives.

choose which of a pair of cities is larger; you believe that having a university is the best cue which discriminates on the criterion of relative city size; you know that one of the pair of cities has a university and that the other does not. Using Take the Best, you would infer that the city with the university is the larger of the two. But note that significant amounts of information are probably built into the very belief that a cue discriminates. That having a university is a good predictor of relative city size may presume, for instance, that one understands what a university is; its function in society; how it houses a number of faculty, staff, and students; perhaps the relative social classes of members of a community typical of such faculty, staff, and students; maybe the ways in which having a university relates to the economy of a city; and probably much more. Indeed, it is the presumption of this information that can *make one believe* that having a university is a good predictor of relative city size, and enables one to recognize and understand the cue *as* a (discriminating) cue. On my account, all this information can very well constitute a dense web, but it would be embedded and organized within the k-systems of an individual. The important point is this implies that when one uses a heuristic like Take the Best, which operates according to a single cue to discriminate, one is actually (implicitly) relying on much more information than is revealed by the heuristic's surface simplicity.

Here are two more examples in a little less detail. Simon's chess heuristic (9) and Kahneman and Tversky's Recognition heuristic (11) each uses a small number of factors in its operation.

(9) In chess, consider first those moves that remove or protect pieces under attack.

(11) Probabilities are evaluated by the degree to which one thing or event is representative of (resembles) another; the higher the representativeness (resemblance) the higher the probability estimation.

Similar to Take the Best, the Representativeness heuristic operates according to resemblance, which is just one cue. Yet judging resemblance is not such a simple process. Rather, it is an informationally demanding task, involving the extrapolation of relevant features from items in perception and/or memory. Whether Dick's description is judged to resemble an engineer rather than a lawyer requires that one understand what each of these occupations are, what members of each occupations are typically like, why members of each occupation are typically like their stereotype, and so on.[35] On the other hand, (9) considers whether any of one's own chess pieces is threatened, and the feasibility of the moves that would remove the threats, if any. At any given point in a typical chess match, the number of pieces under threat of capture is small (usually no more than three). And the potential moves to remove or protect the piece are also typically relatively few. However, the heuristic must rely on other information if it is to operate at all, such as what moves each piece can legally make (and really, all, or most, of the rules of the game). Since chess is a finite, well-defined game, the amount of information that (9) invokes and relies upon, despite how simple it appears on the surface, may not be as much as the Representativeness heuristic or Take the Best, but the point holds just the same.

It is in this way that I believe heuristics exploit informationally rich k-systems, but remain frugal in their computations. In fact, we might say that it is because heuristics exploit k-systems that they can be frugal, since it is the informational richness that is bearing the cognitive load. If this is right, then I believe we have the beginnings of a plausible model of how heuristics operate in central cognition.

3.5 Concluding remarks

To recap, we began by considering the frame problem and saw that there are many aspects to it. This chapter was guided by the computational relevance problem, and more specifically with the appeal to heuristics to circumvent the problem. I showed that such an appeal to heuristics has implications for cognitive architecture.

Philosophers such as Samuels (2005, forthcoming) and Carruthers (2006a, 2006c) believe that heuristics solve the frame problem, whereas Fodor argues that they do not. I argued that heuristics do not offer the kind of solution that Samuels and Carruthers think; specifically, heuristics fail to circumvent the epistemological relevance problem. However, despite these

[35] In chapter 5, I will be giving a fuller description of the Dick example, and an analysis in light of the thesis of this dissertation.

arguments, I want to maintain that heuristics actually have a significant role in circumventing such epistemological worries (though, of course, not in the way that Samuels and Carruthers suggest). So I went on to gain a better understanding of the role heuristics have in cognition, which meant investigating more thoroughly how heuristics work and the structures they operate over.

Some authors believe that tractable cognition requires domain-specific computational mechanisms, and thus heuristics must be domain-specific. A domain-specific mechanism is supposed to have built into it a significant amount of information about the domain in which it operates so as to render it *informationally rich*, capable of quickly deploying strategies to (re)solve the problems with which it is faced. Nevertheless, following Samuels (2000, 2005), I argued that central cognition can be a domain-general computational system that draws on a variety of specialized bodies of knowledge. Samuels and Fodor (2000) construe such domain-specific bodies of knowledge as Chomskian modules, which are innate and domain-specific. But I argued that such specialized bodies of knowledge need be neither innate nor domain-specific. These k-systems, as I call them, are simply highly organized and structured systems of knowledge.

We saw that Sterelny (2003, 2006) believes that humans tend to increase their cognitive powers by informational niche construction, and creating and utilizing epistemic technology. Thus, Sterelny offers a plausible account of how informationally rich bodies of knowledge can be constructed, developed, and possessed by humans, as well as how such informational richness can be instilled and stored in the environment for humans to, perhaps heuristically, draw on and exploit. In contrast to Sterelny (and Gigerenzer), however, I claimed that heuristics do not exploit information in the environment directly, but the information about the environment embodied by the appropriate k-systems *in our heads*.

Finally, I argued that, although heuristics carry a high informational load, they remain frugal in their operations or computations by exploiting informationally rich k-systems. In short, exploiting k-systems is ultimately what allows heuristics to be fast and frugal, as k-systems bear much of the information load in cognition. This view is consistent with the characterization of “heuristic” I developed in the previous chapter.

In the next chapter I will provide a more specific picture of k-systems and of how heuristics exploit their informational richness. I will claim that what I have been referring to as k-systems are in fact concepts. And after developing a specific theory of concepts, I will explain in more detail how heuristics work.

Chapter 4

Concepts, Heuristics, and Relevance

Now that I have described a general architecture of cognition, it is time to provide a more detailed account of the cognitive structures I have been calling k-systems. The account I will give will show that *concepts* are the cognitive structures that fulfill the role of k-systems explained in the previous chapter. The basic idea is that our concepts embody the kind of informational structure that k-systems are supposed to have; and as such, concepts guide our reasoning and inference, and they have a special role to play with respect to facilitating the operations of heuristics in particular.

There are many theories of concepts. The theory that I will adapt is Lawrence Barsalou's (1999, 2003, 2005, 2008b, 2009; Barsalou, Simmons, Barbey, & Wilson, 2003). Barsalou's account of concepts is unlike many of the leading philosophical accounts insofar as it comes from a psychologist, and more importantly it is grounded in perception rather than nonperceptual cognition or a language of thought hypothesis.[1] More specifically, Barsalou claims that concepts are partly realized in collections of neural patterns in the perceptual centres of the brain that were originally activated upon perceiving instances of entities in a concept's extension. Barsalou's theory of concepts is situated in his more general view of conceptual cognition, which he calls the *perceptual symbol systems* (PSS) view of cognition.

Although I will be drawing heavily from it in giving my own account of cognition and how heuristics work, Barsalou's theory will go largely undefended. Providing a full defence of a theory of concepts is beyond the scope of this dissertation. However, my reliance on PSS is justified to the extent that the theory has psychological and neural (biological) plausibility. Moreover, since my proposed cognitive architecture enables us to model heuristic cognition, my reliance of the PSS model is justified insofar as we are able to explain a wide range of phenomena (which, it may be recalled, was a goal set out in chapter 2).

As I shall show, the PSS theory of concepts is in some respects related to an account of cognition that many philosophers, psychologists, and cognitive scientists have converged on, namely the *file model* of cognition. Such widespread convergence indicates that there likely is

[1] Although see Prinz (2002).

something right—or at least interesting—about the theory. According to the file model, concepts come with "files" that contain information (e.g., representations and beliefs) about the concepts, or more precisely about the entities in the concept's extension. In section 1 of this chapter, I will expound Barsalou's PSS theory as well as the file model of cognition. It will be seen that these two accounts are not immediately compatible with each other. However, in section 2 I will offer a critical assessment of each account, which will entail certain modifications and qualifications. The result will be a reconciled theory of concepts and cognition that is more robust and, in my view, more fruitful.

In section 3 I argue that concepts, understood in terms of my reconciled theory, fulfill the role of k-systems described in the previous chapter. I will there pick up the discussion from the previous chapter regarding the exploitation of information in the environment, such as epistemic technology, and argue that it is in fact one's conceptual wherewithal to which one owes the ability to exploit such information in the right ways. We will then be in a position to see that concepts—or more precisely, the information embodied by concepts and the relations among such information—facilitate heuristic inference. Specifically, I will contend that concepts *constrain* heuristic inference, and furthermore that the characterization of "heuristic" given at the end of chapter 2 (viz. that they are cognitive procedures that satisfice, and require little resources for their recruitment and deployment) is met by virtue of such constraints.

Section 4 will be devoted to providing a more detailed understanding of how heuristics work. I contend that heuristics do not operate over conceptual content, but over higher-order relations between active and primed concepts and their content (conceptualizations and representations). I describe two sorts of informational structure that heuristics exploit. These two kinds of structure are not meant to be exhaustive, but in the next chapter I will illustrate that they have a role in facilitating some of the heuristics studied by Kahneman and Tversky, and Gigerenzer.

As we shall see, the present chapter will be one largely of synthesis, but it will provide substance to many of the ideas central to this dissertation.

4.1 Concepts and representation

4.1.1 Perceptual symbol systems

As alluded to, Barsalou's (1999) theory of mental representation is perceptually grounded. His view is in a similar vein as the British Empiricism of Locke, Berkeley, and Hume. According to the Empiricist tradition, sense experience is the only source of ideas (or mental representations), and moreover, ideas are imagistic in nature. The British Empiricists also believed that thought proceeded via associations among ideas. Though Barsalou's theory does not align precisely with British Empiricism, his theory steps away from language of thought approaches to mental representation by grounding representation in the perceptual system of the human brain.

As Barsalou explains, the language of thought hypothesis originated from developments in logic and computer science, from which formalisms such as the predicate calculus and programming languages inspired new ways to conceive of mental representation.[2] The new conception of representation departed from experience-based empiricism, and this had considerable consequences, not least of which was a drawn principled distinction between *perception* and *cognition*. One important difference believed to exist between perception and cognition is that perceptual representations are understood to be *modal* (i.e., they have specific properties that are bound to specific mental/neural systems), whereas cognitive representations are conceived to be inherently modality-neutral, or *amodal*. This immediately leads to the assumptions that there must be separate mental/neural systems dedicated to perception on the one hand, and cognition on the other, and that each system must use different means of representation and operate on different principles. Thus, for instance, a perceptual representation of a chair will have specific properties and be produced by specific systems that are able to represent its shape, colour, tactile features, and so on. On the other hand, a cognitive representation of the same chair will have none of these features, but will be represented—via a transduction of information from perceptual systems to cognitive systems—as a nonperceptual, amodal symbol

[2]Cf. the discussion in chapter 2 of this dissertation regarding the rise of the interest in heuristics.

chair,[3] that behaves a lot like and shares similar properties with words or sentences of natural languages. It is believed that, as amodal symbols are transduced from perceptual systems, they enter into larger representational structures that operate as a fully functional symbolic system which supports all of the higher cognitive functions. Assuming CTM, mental conceptual processing can thereby be understood to be computational operations defined over the word- or sentence-like syntax of the representations of the language of thought. Thus, we get in thought and thinking the productivity, systematicity, compositionality, and everything else that we get with natural languages (Fodor, 1975, 1983, 2008; Fodor & Pylyshyn, 1988).

Nevertheless, Barsalou believes that mental representations are not amodal language-like structures. Rather, he argues that representations are modal "perceptual symbols". Perceptual symbols "are records of the neural states that underlie perception" (Barsalou, 1999, p. 582), and therefore perceptual symbols "are modal because they are represented in the same systems as the perceptual states that produced them" (p. 578).[4] Thus, on Barsalou's account, perception and cognition share representational systems. This does not mean that identical systems subserve perception and conceptual thought; for indeed, mechanisms additional to perceptual systems are needed for conceptual thought. Instead, perceptual symbols are essentially based in or arise from perceptual systems in virtue of capturing and maintaining information from perceiving entities and events in the environment and the body. Hence, for Barsalou, "cognition is inherently perceptual" (p. 577).

We should note that in important ways perceptual symbols are unlike the ideas or mental images envisioned by the British Empiricists. Perceptual symbols are not conscious mental images, but are patterns of neuron activations. Therefore, although perceptual symbols can function in consciousness or unconsciousness, they are essentially unconscious mental representations. Moreover, whereas Berkeley and Hume believed that representation cannot occur without some specific particular being represented, Barsalou claims that perceptual symbols do

[3]The convention I adopt here is to represent concepts in small caps, contents in italics, and uses in quotes.

[4]It is not entirely clear what makes perceptual symbols, symbols in the philosophical sense (i.e., in terms of meaning, referring, and being grounded). I want to avoid these issues, however, so I will uncritically follow Barsalou's terminology.

not necessarily represent particulars. This is because perceptual symbols are not entire neural states that are activated in perception. Rather, they are only schematic aspects of the neural states of active perception. According to Barsalou, *selective attention* allows the cognitive system to isolate and focus on only certain features of a perceptual state, depending on the nature of the cognitive/perceptual task. As he observes, psychological work on attention has shown that cognition can attend to the shape of an object while ignoring its colour, texture, and other surrounding objects. Those features of a perceptual state (i.e., the subset of neural patterns active in perceptual systems) that are attended to in the right sort of way[5] are consequently stored in long-term memory (although some information not attended to may get encoded as well to a lesser degree). It is these attended-to features that constitute mental representations.

As a result of the selective attention process of symbol storage, perceptual symbols are componential, not holistic. This is to say that perceptual symbols represent certain features determined by the stored neural patterns, not entire "images" as understood in the Empiricist tradition in which Locke's primary properties (such as shape or size) are necessarily present. As Barsalou explains, "It is well known that high-level neurons in perceptual systems can code information qualitatively. For example, a neuron can code the presence of a line without coding its specific length, position, or orientation" (p. 585). One can therefore represent a line with few or no other specific qualities.[6] This may seem counterintuitive, but it is important to keep in mind here that perceptual symbols are essentially *unconscious* representations (that can function in consciousness). Thus, despite the fact that it is difficult, or maybe even impossible, to consciously construct a nonholistic representation (e.g., an image), it is not unreasonable to assume that this can be done unconsciously.

[5]Barsalou does not offer a theory of attention.

[6]One might conjecture here that it is the *use* of one's concept that enables someone to represent a line without representing a specific length, position, or orientation. That is, once one grasps the concept LINE, one need not represent any specific features of a particular line when mentally invoking the concept. This may very well be the case, but it misses the point being made here, which is that Barsalou's conception of perceptual symbols is not to be equated with static images (Barsalou, 1999, 2008b) such as those envisioned by the British Empiricists, and moreover, that Barsalou cites empirical evidence for the possibility of human neurology to realize representations of an object without representing many of its specific features. (Pylyshyn (2003) gives evidence that properties of objects can be tracked without being represented, and this lends support to the possibility of representing an object without representing specific properties of it.)

Perceptual symbols are not only supposed to be modal, according to Barsalou, but *multi*modal in the sense that they encode information from any of the sensory modalities. The idea is that different neural/sensory systems will become active upon perceiving an object, and information captured by these systems will serve to represent various perceptual features of the object. Thus, during visual processing of a chair, edge and surface detector neurons fire, while other neurons fire for colour, orientation, and motion. The overall pattern of activation across this distributed system represents the entity in vision. Analogous patterns of neural activation in other sensory modality systems represent how the chair feels, smells, sounds, etc. (Barsalou, 2009). Attended-to aspects of these activation patterns are conjoined by "conjunctive neurons" that bind features within a given modality (Barsalou, 2009; Barsalou et al., 2003; Simmons & Barsalou, 2003; cf. Damasio's (1989) notion of *convergence zones*), and these bound patterns are stored in memory as perceptual symbols. "Later, in the absence of [perceptual] input, these conjunctive neurons partially reactivate the original set of feature detectors to represent the [chair]" (Barsalou et al., 2003, p. 85), and higher association areas of the brain (i.e., the temporal, parietal, and frontal lobes) integrate activations across modalities (Barsalou, 2009). These integrated activations thus constitute multimodal perceptual symbol representations.

A very important feature of PSS is that perceptual symbols do not exist independently of one another. Rather, "related [perceptual] symbols become organized into a *simulator* that allows the cognitive system to construct specific simulations" (Barsalou, 1999, p. 586) of the perceptual component it represents.[7] The simulations Barsalou has in mind are analogous to

[7]In his original paper, Barsalou (1999) was not explicit on what makes perceptual symbols "related", though he hinted toward the roles played by spatial and temporal relations of external events. More recently, Simmons and Barsalou (2003) propose that the spatial proximity between conjunctive neurons (i.e., the neurons that bind the neural patterns that represent perceptual features) in the brain's association areas is indicative of similarity relations between representations. In more specific terms, Simmons and Barsalou propose what they call the *similarity-in-topography principle*, which states that the closer that conjunctive neurons are in the spatial topography of an association area, the more similar those features will be; in their own words: "In general, the topographic proximity of two conjunctive neurons reflects the similarity of the features they link" (p. 458). As an example, Simmons and Barsalou claim that the conjunctive neurons that serve to represent a human face will lie close to, and even partly overlap with, the conjunctive neurons that serve to represent a monkey face; the conjunctive neurons for the face of an elephant, on the other hand, will not reside as close to those for the human face, but will be much closer than conjunctive neurons that serve to represent a completely different type of object, such as a chair. Simmons and Barsalou do not have an answer as to what makes conjunctive neurons bind certain features rather than others. However, they suggest that conjunctive neurons may be "tuned" to certain features that are of

the simulations that are conducted in mental imagery. When multimodal perceptual information is extracted and integrated into an organized system, one can, at any point thereafter, mentally simulate one's experiences. Barsalou goes on to assert that, according to his theory, "the primary goal of human learning is to establish simulators. … Once individuals can simulate a kind of thing to a culturally acceptable degree, they have an adequate understanding of it" (p. 587). In other words, once one is able to construct a range of multimodal simulations of an object or event (in a manner that is congruent within one's community), one *possesses the concept* for that object or event. Hence, Barsalou claims that "a *concept* is equivalent to a simulator" insofar as a concept consists of "the knowledge and accompanying processes that allow an individual to represent some kind of entity or event adequately" (p. 587).[8] It will become much clearer as we work through this chapter how important the knowledge to represent an entity or event adequately is; specifically with respect to enabling certain neural activations (e.g., those activated for words) to activate the appropriate representations for a given concept.

Moreover, a concept is a general kind, but since simulators produce simulations, a concept can undergo any number of *conceptualizations*—different representations of a concept according to the many different possible simulations of it. Again, for example, a chair can be conceptualized (represented) in different ways concerning its size, shape, colour, material

evolutionary significance; and they leave the door open to the possibility that (at least some) "tuning" is learned. These matters are tangential to the present discussion, so I will not assess them here.

[8]Since the publication of his original paper, Barsalou has wavered on his stance on equating simulators with concepts. For instance, in a paper published some four years later, he and his collaborators discuss some empirical data on grounding conceptual knowledge in modality-specific systems. They there state that "nothing is explicitly called a concept", and that the simulator is "one central mechanism" of conceptual processing (Barsalou et al., 2003, p. 84, Box 1). Yet, in a publication that same year, Barsalou asserts, "a concept is a simulator that constructs an infinite set of specific simulations" (Barsalou, 2003, p. 522; more on this below). More recently, Barsalou contends that simulators "implement" concepts and conceptual processing, while at the same time stating that "Theoretically, a simulator functions as a concept or type in more traditional theories by integrating the multi-modal content of a category across instances, and by providing the ability to interpret individuals as tokens of the type" (Barsalou, 2009, p. 1282). In another work, he states, "simulators are roughly equivalent to concepts in traditional theories of knowledge" (Barsalou, Santos, Simmons, & Wilson, 2008, p. 251). Barsalou's noncommittal stance toward strictly identifying concepts with simulators may be a reflection of his effort to distance his view from language of thought theories of mental representation and concepts (Barsalou, 2003, 2005b). It is also possible that Barsalou is using the term "concept" to refer to categories, as psychologists are wont to do, as opposed to what philosophers mean by concepts. However, on some occasions it appears as if Barsalou is in fact using "concept" the way philosophers use the term. In any event, to avoid the vagueness of Barsalou's position, I will assume here that simulators are to be identified with concepts. As we shall see presently, the theory of concepts I develop adopts this position.

composition, feel, etc.[9] Different conceptualizations will be activated depending on the context and goals of the agent.

Concepts are standardly conceived to represent categories or classes of things in the world, i.e., the objects in their extension. CHAIR represents the class of chairs; CAT represents the class of cats; etc. As such, Barsalou claims that once a simulator becomes established in memory, and as knowledge is accumulated and added to the simulator, one is in a position to identify members of its class, adequately interact with its members, and make categorical inferences about them. Categorical inferences proceed from knowledge associated with the category, which "provides predictions about [an] entity's structure, history, and behavior, and also suggests ways of interacting with it" (p. 587). Furthermore, since a simulator contains an enormous amount of (multimodal) knowledge about a category, one can simulate aspects that *go beyond* what is perceived, such as anticipating an entity's structure or behaviour. For example, Barsalou claims that the simulator for JET "might suggest that the jet contains pilots, passengers, luggage, and fuel ... that the jet is likely to fly horizontally, not vertically, and is eventually likely to decrease altitude, put down its wheels, and land at an airport" (p. 596). Thus, simulators, according to Barsalou, afford a wealth of top-down inferences.

It would be instructive to pause here to note that, as Barsalou describes the affordances of simulators, simulators appear to be as potent as propositional knowledge. We should be hesitant to accept this, at least not without some critical reflection. Simulators consist of perceptual symbols which are in some ways similar to images (though, recall, they are distinct from images). There is a problem, however, with images expressing propositions—a problem that goes as far back as Fodor (1975) in *The Language of Thought* (chapter 4). Fodor argued that

[9] Barsalou (1999) claims that, despite the fact that an individual can produce different simulations or conceptualizations of a given concept, concepts are stable for an individual insofar as the same simulator (i.e., the same areas of neural activation) produces the different simulations. Moreover, Barsalou claims that, although different individuals conceptualize concepts differently, concepts are stable between individuals insofar as (i) they share similar cognitive systems and experiences, and (ii) they have the ability to simulate each others' conceptualizations. I mention these issues to anticipate some philosophical worries that some may have with Barsalou's theory, such as how we are to individuate concepts, how to account for concept stability (i.e., how a concept can remain the same concept under different simulations or conceptualizations), how the same person can maintain the same concepts over time, and how different people share the same concepts (this is what Fodor (2008) calls "publicity").

we cannot think in images because images cannot express propositions, and this is because images "are insufficiently abstract to be the vehicles of truth" (p. 180)—a problem that does not exist for a representational system like language. Consider for example a picture of Bob, who happens to be a short and fat man. Can we say that the picture truly express "This is Bob"? Or "This is a man"? Or "Bob is a short, fat man"? Or "Bob is pregnant"? (For the last, think about what picture would express that proposition.) In short, images do not do the kind of work that language does, since images are essentially analogue representations while linguistic representations are digital. In light of this, then, we might be skeptical about Barsalou's claims regarding the work that simulators *qua* collections of perceptual symbols do. As indicated, perceptual symbols are not to be identified with images.[10] Nevertheless, the fact that simulators are based in the perceptual centres of the brain should cause us to question whether simulators can "suggest" the propositional knowledge that Barsalou claims (e.g., can a simulator for JET be a vehicle for the truth that the jet will eventually land at an airport?). This problem may be circumvented if we include a role for natural language in our conceptualizations. I will be returning to this suggestion when I adapt Barsalou's theory to achieve an understanding of richer concepts.

In any event, simulators, as Barsalou envisions them, also impose constraints on conceptual processing, which arise from associative connections between perceptual symbols that represent individuals within the simulator. Thus, activating a subset of perceptual symbols having to do with the back of a chair can activate associated subsets of perceptual symbols having to do with the seat of the chair. Constraints weaken or strengthen over time, and default perceptual symbols can be constructed. Hence, what one simulates (or at least the ease with which one simulates) is constrained by the perceptual symbols one possesses and the relations between them. Furthermore, simulators provide background structures that support understanding concepts. For example, the meaning of "lock" is ambiguous when left unspecified relative to some background—it can mean (among other things) *a tuft of hair* or *something that fastens a door*.

[10]Maybe they are better likened to film strips, though with the possibility of only certain properties being represented.

But different perceptual symbols are accessed for “lock” depending on the context. An active simulator for HAIR will determine a background of organized knowledge against which “lock” is understood and conceptualized as a tuft of hair; whereas “lock” will be conceptualized as something that fastens a door when entertained against the background simulator for DOOR. Such constraints on conceptualization and representation will have a significant role in the account I develop below of how heuristics work.

In summary, multimodal perceptual symbols are derived from previous perceptual experiences with certain events or objects. Perceptual symbols get bound up and integrated into a simulator that produces a range of simulations or conceptualizations of the events or objects in question. In this way, simulators are concepts, and the simulations produced by a simulator represent multimodal features of entities that belong to the concept’s extension (i.e., simulations are conceptualizations).

Barsalou’s theory of concepts is certainly unlike the standard accounts found in philosophy, especially those that rely on the language of thought hypothesis. Although it is beyond the scope of this dissertation to fully defend Barsalou’s theory, I believe that it is a psychologically and neurologically (biologically) plausible account of mental representation. As we shall later see, I will adapt PSS theory to advance my account of how heuristics work. Nevertheless, some justification for adopting Barsalou’s account can be found vis-à-vis a model of cognition that is becoming increasingly popular in philosophy and cognitive science, namely the file model of cognition. I will now briefly expound the file model, and afterwards I will critically assess it in concert with Barsalou’s theory of concepts in an attempt to reconcile the two accounts.

4.1.2 The file metaphor

A number of cognitive scientists and philosophers have converged on an idea about the nature and function of mental representations (e.g., Bach, 1987; G. Evans, 1982, 1985; Fodor, 2008; Lawlor, 2001; Perry, 2001). The idea is a *file model* of cognition (Kahneman, Treisman, & Gibbs, 1992; Pylyshyn, 2003; Treisman, 1982; Treisman & Schmidt, 1982). There is no standard doctrine accepted by everyone who endorses the file model, but there is some common ground that can be identified. I will here roughly follow Fodor's (2008) account since it brings to bear a number of issues that are of concern to this chapter.

According to the file model of cognition, each object that one has beliefs about (i.e., knows of, is acquainted with, erroneously believes exists, etc.) has its own mental "file" in one's head. Each mental file has a "label" that both picks out the file, and names the object that is associated with the file. Further, each file holds a number of "notes" or "memos" that contain various kinds of information or beliefs about the given object. Recently, the file model has been adapted as a way to understand the nature and function of concepts.

According to Fodor, mental file-labels are expressions in Mentalese, and the file-notes are written also in Mentalese. For example, when you think *office*, you token an *office*-label, which calls up the corresponding file. You can then open the *office*-file and find in it notes containing representations about desks, office doors, books, shelves, papers, computers, and so on, depending on what you believe about offices. Conversely, the *office*-file is what gets brought to mind when you, for instance, recall where you left your copy of *Modularity of Mind*.

Of course, files get created and altered subsequent to new experiences and changes of beliefs. When you meet Alice for the first time, a new mental file gets created, labeled *Alice*,[11] and in it a number of notes are created according to the beliefs you acquire about Alice. Over time new notes are added and some notes get revised: You had initially thought that Alice was single, but now you have come to believe that she is married, and so you make the appropriate corrections to the notes in your *Alice*-file. The number of files and the number of notes in each file can be very large. Fodor believes that they can be "arbitrarily large" and speculates that this might mean "potentially infinite" (p. 94). But surely, what files and notes one has are bound by limitations on memory. This does not mean, however, that one cannot have an extremely large number of files and notes.

There is also the possibility, as acknowledged by Fodor, that beliefs not contained within a

[11]If one knows more than one Alice, the mental file would be labeled $Alice_n$, where n differentiates. (This is how Fodor (2008) claims to have solved Frege-style problems of sense and reference.)

given file may be inferred "online". Thus, the belief that your cat is not in the top drawer of the desk in your office is probably not in your *office*-file, and is generally not something that you store anywhere in memory, but can be inferred when and if there is an occasion to do so. This certainly mitigates the demands that the file model places on memory. Moreover, there will likely be constraints on these sorts of inferences. An inference that your pen is in your desk drawer is easily made, since that is the place where you usually keep your pen—you therefore have a store of memories and representations of your pen in your desk drawer. Inferring that your cat is not in your desk drawer may not be as readily made, given that your cat is never in your office and may not even fit in the drawer. However, you will probably never infer that your car is not in your desk drawer (except when doing philosophy), since your car is simply not the kind of thing that is or can be in a desk drawer (not to mention in an office).[12] We should acknowledge, however, that this is an empirical matter, as is what is actually stored in one's files and gets inferred online.

The file model of cognition is useful in explaining the flow of thought. For the notes of a file can themselves contain the names of labels for other files. Thus, when one searches through the contents of a given file, other files may be brought to mind. When you recall that your copy of *Modularity of Mind* is on top of a stack of *papers* on your *desk* by opening your *office*-file, the files for *papers* and *desk* may be made available for opening and searching among their contents. Alternatively, one may make inferences and thereby make available further files. Thus, even so your *office*-file does not contain a note conveying the belief that your cat is not in the top drawer of your desk, you can infer this and thereby bring to mind the file for *Felix*, your cat.

It is important for the present purposes to understand how the file model can be understood

[12]We might think of these constraints in terms of pattern matching of information stored in memory. In our example, you have memories and representations of your pen in your desk drawer, whereas you have no memories or representations of your cat (or your car) in your desk drawer. And, by assumption, it would be much easier to match and retrieve associated representations in memory (your pen and your desk drawer) than unassociated representations (your cat (or your car) and your desk drawer); and there may even be resistance in associating certain representations if there are conceptual problems in bringing them together (such as your car and your desk drawer, as cars are not the kinds of things that can fit in a desk). See below for further remarks.

more particularly to be a theory of the structure of concepts. On this reading, a concept names a file that contains beliefs about the things in the concept's extension. When thinking about cats, for instance, one calls upon one's file for CAT, which may contain such information as *is a living thing*, *is a furry creature*, *is a quadruped*, *is grandma's favourite animal*, etc. Notice now that what is contained in the file—what notes there are in the file—name other concepts: LIVING THING, FURRY CREATURE, QUADRUPED, GRANDMA'S FAVOURITE ANIMAL, etc. This is not classical decomposition. Rather, concept files contain notes that convey information about the concept in question. Hence, unpacking the notes of one's CAT concept file is unpacking one's beliefs and other stored information about cats, the representations of which invoke further concepts. Moreover, similar to what was explained above, whatever is inferred from the beliefs one holds about a given concept may also excite further concepts, thereby bringing to mind files of their own. The upshot of this picture for Fodor is that, since concepts name files, one thinks in file names; and one can subsequently entertain the notes of a file once the appropriate file name (i.e., concept) is tokened. In short, "file names are Janus-faced: one face turned towards thinking and the other face turned towards what is thought about" (Fodor, 2008, p. 100).

What I have presented here in brief is a basic file model of cognition. Some authors add a number of details beyond this basic model, but considering these extra details will take us too far afield. We will gain a better understanding of how Fodor envisions the file model to contribute to the nature and function of concepts when I critically assess his account below. For now, however, let us observe that Barsalou's theory of concepts bears a certain relation to the file model, as both theories characterize cognitive representations and concepts in relation to ensembles of information that exhibit specific structures and relations. More specifically, the information constituting the perceptual symbols involved in a concept's simulations can be understood to be what is contained in the concept's file. Nevertheless, there are some details that need to get ironed out if Barsalou's theory of concepts is to be reconciled with the file model of cognition, and, importantly for the purposes of this dissertation, if the resulting reconciliation can fulfill the role of k-systems as conceived in the previous chapter. For instance, Barsalou's perceptual symbols are multimodal and grounded in perception whereas the labels and notes of a file are supposed to be conceived in terms of an amodal language of thought. In the next section I will argue for some amendments to both Barsalou's theory and the file model, the result of which will be a reconciled theory of concepts and cognition. I will subsequently show that this reconciled theory indeed supports the role of k-systems; I will then be in a position to propose that they underlie the operations of heuristics.

4.2 A theory of richer concepts

This section will proceed in two steps. First, I will discuss a problem faced by the file model. This problem is regarding whether it is possible to token a concept without invoking any representational content. I will argue that this is not possible on Fodor's version of the file model, since it runs contrary to some assumptions of the language of thought hypothesis. I will suggest more generally that conceptual content and associations between representations are invariably invoked when mentally entertaining a concept. This view will play a crucial role for my account of how heuristics work. In suggesting this view, moreover, I will propose that Barsalou's PSS theory be interpreted in light of the file model, thereby initiating a reconciliation between the file model and Barsalou's theory of concepts.

The second step entails a critique of the PSS theory of concepts as envisioned by Barsalou. In this subsection, I will argue for a role for natural language that augments Barsalou's understanding of simulators and simulations. The result will be an enriched theory of concepts and cognition which will underwrite the remainder of the discussion in this dissertation.

4.2.1 Conceptual content and associations

The degree to which notes are extracted from files will depend on the cognitive task and task demands. In general, the whole of what one believes about an object or concept will not be brought to bear in one's thinking. Instead, one will tend to (or at least hope to) bring to bear only what is relevant. In recalling where you left your copy of *Modularity of Mind*, for example, you will not recall everything you know or believe about offices—that is, you will not consider the *office*-file in its entirety. You would likely extract some information to the effect of, say, *Modularity of Mind* being on your office desk; and other representations may also concomitantly be brought to mind, such as those having to do with the stack of *papers* you have sitting on your *desk*, on top of which lies *Modularity of Mind*. (I italicize *papers* and *desk* to signify, as explained above, that these representations can also be the labels of files themselves (i.e., concepts), and thus any of a host of representations related to these objects may be activated.)

On the other hand, Fodor contends that label-representations can serve in one's thoughts without inviting any other representations contained in their corresponding files to be brought to mind. This means, for instance, that you can token an *office*-representation *qua* file label without attending to any other beliefs about offices—that is, token the label without opening the mental file. Whether one does or does not open files when labels serve as constituents of one's thoughts will, again, depend on the cognitive task and the task demands. Although this is an empirical claim, it is a matter of principle according to Fodor. As he asserts, there is a principled distinction between a file and the notes it contains; and this allows for a correspondingly principled distinction between entertaining a representation and entertaining other beliefs about that representation (Fodor, 2008, p. 97).

Whether one can in fact entertain a representation without entertaining other beliefs about that representation, or whether and to what extent this actually does happen, are issues that deserve serious attention. I suppose, however, that one cannot *think* a representation (have a representation *in thought*) without entertaining other (related) representations. Indeed, it is hard to understand how one can token a file label without invoking any content on Fodor's view, since file labels are supposed to be the constituents of thought, namely concepts. But tokening a concept supposes content. The crux of this issue has to do with the way in which labels are connected to their referents. Since, according to Fodor, labels are concepts, tokening a label in thought must bring the referent to mind, and this implies that some content must be invoked. Thus, the tokening of a concept in thought supposes entertaining other (related) representations.

It is important to keep in mind that the principled distinction between a file and the file's contents, and the correspondingly principled distinction between entertaining a concept and entertaining other beliefs about that concept, are grounded in the language of thought hypothesis. It is the *labels* of files that are supposed to be the representations we think in, i.e., our concepts. In other words, file labels are the nonperceptual, amodal, language-like symbolic representations that are the constituents of the language of thought. Notes that are in attending files are also written in the language of thought, though they would be complex constructions of the basic constituents, i.e., complex constructions of concepts. According to Fodor's account of the file model, it is because we think in file labels that we are able to pull files, as it were, without necessarily opening them; and hence, so Fodor claims, one can think OFFICE without entertaining any beliefs about offices. Fodor claims that this is possible when *the name* of the *office*-file is the constituent of one's thoughts. But it is difficult to understand what he means by this when he is committed to the position that the proper constituents of thought are supposed to be concepts and not the names of concepts. As we shall see below, there is a sense in which Fodor's claim is tenable, namely if we understand file labels to be natural language forms (e.g., words). Labels might thereby be connected to the referents expressed by concepts, but tokening a linguistic form in thought will not necessarily bring the referent to mind—linguistic forms *per se* are contentless. That is, if we understand file labels to be natural language forms, the label would be connected to the file, but the content is the stuff in the file, and whether and the extent to which the file is opened and explored will depend on the task, cognitive demands, etc. Nevertheless, this possibility is not open to Fodor since, again, he is committed to the assumption that file labels just are concepts. Yet concepts are unlike natural language words in that *inter alia* concepts cannot be divorced from their representational content. Natural language words, on the other hand, are not essentially mental representations; rather, they are *linguistic* representations which can be mentally represented. This is an ontological distinction that should not be ignored. I will return to this distinction between natural language and Fodor's file labels below.

It therefore does not appear that Fodor can consistently maintain that file labels can be entertained without invoking (at least some of) the file's representational content.[13] In general, I assume that concept-files are brought to mind partly opened with some of its contents spilling out—some of its contents are activated, while other contents are primed. This may be due to unconscious[14] entertaining of representations before they are "brought to mind" in conscious thought. But whatever the case may be, if this is correct, the principled distinction between entertaining a representation and entertaining other beliefs about that representation does not hold.[15]

We already saw that Barsalou does not subscribe to the language of thought hypothesis, and that concepts, according to his theory, are multimodal and grounded in perception. Furthermore, concepts are richly structured on Barsalou's account, consisting of a simulator and a potentially infinite set of simulations. Indeed, Barsalou argues that such structure can provide for the productivity and compositionality we get with natural language. I propose, however, that his theory of concepts be interpreted in light of the file model. Accordingly, the PSS view would entail that concepts are not mere file labels, but are the files in their entirety, notes and all. The integrated perceptual symbols in a simulator and its simulations can be understood as the notes in the file, while the simulator can be understood as the file itself. Barsalou's theory thus interpreted is consistent with the assumption that files are brought to mind partly opened with some of its contents spilling out. For on the foregoing account, entertaining a concept by conducting simulations will necessarily activate and prime a variety of information about the

[13]Save for the possibility of contentless concepts. I am unsure, however, whether there can be contentless concepts. At the very least, we can safely assume that the concepts we usually entertain in thought will have content.

[14]To be succinct, I will use "unconscious" to refer to anything that is not conscious, including preconscious thought or representation.

[15]We might also note that if it is possible to token a concept without tokening any other representations, then it would be possible for a mind to contain only a single concept which it grasps and of which it has thoughts (cf. Fodor, 1990)—in Fodor's words, thinking about the concept "as such". However, this is seen to be absurd, since thinking in concepts supposes content, and content supposes more than a single representation.

concept.

Fodor maintains that it is wrong to characterize concepts as files themselves, rather than labels, since not only does this dissolve the distinction between entertaining a file and considering its contents, but it also threatens associationist commitments that, being a computationalist, Fodor is loathe to endorse. For instance, if we accept that concepts are files themselves, then when one entertains a file one must entertain all thoughts implicated by the notes that happen to "spill out". Fodor might be right that not everyone thinks mouse when she thinks cat, even though these are highly associated concepts. However, the co-tokening of representations need not occur in conscious thought. Indeed, it may be recalled that Barsalou believes that perceptual symbols are mostly utilized unconsciously. Thus, co-activations or primings of concepts and representations can very well occur unconsciously, even if they do not reach consciousness. It can be that I consciously token cat, but I unconsciously token mouse representations. Or, in terms of Barsalou's theory, I consciously token cat, and perceptual symbols having to do with mice (say, a cat chasing a mouse) are activated unconsciously; in fact, on Barsalou's theory, unconscious perceptual symbols may cause me to consciously token cat in the first place.

This picture, however, should not be confused with the associationists' doctrine that Fodor argues against; Barsalou's theory is not an associative-dispositionalist theory. Whereas associationists believe that *fixed* associations are what guide thought, the connections between representations on Barsalou's account are dynamic and vary in degree, depending on the context and the cognitive task, as well as on the strength of connections between perceptual symbols. If I am simply thinking about how lazy cats can be, I may entertain no representations of mice at all, consciously or unconsciously. Again, in terms of Barsalou's theory, my cat simulator would not be simulating anything having to do with mice.

Moreover, Barsalou need not be committed to the claim that thought generally proceeds via associations among representations, despite the claim that *ex hypothesi* people have a general tendency to co-token representations associated with the concept one is entertaining. When tokening cat, one can very well entertain a number of beliefs about cats (from the notes in the cat file or the perceptual symbols involved in a simulation) and thereby activate a network of associations between various other concepts, without thought proceeding via those associations. Certainly, one can proceed from one thought to the next based on associations, but this is optional and depends on the cognitive task and effort being invested in the task. At the same time, it is perfectly acceptable on the foregoing account that thought generally proceeds via inference (consciously or unconsciously). One can certainly entertain representations and attending concepts associated with CAT without inferring anything from those associations. Yet inferences can be made by *utilizing* the associations or co-tokenings of representations. In this way, one's thoughts might be "guided by" associations in a loose sense. We shall see in more detail below the role of co-tokened representations when entertaining concepts in inferential cognition. The present point, however, is simply that the analysis of Barsalou's theory of concepts that I am giving here should not be understood as a brand of associative-dispositionalism.

To sum up the discussion so far, it is possible to reconcile Barsalou's theory of concepts with the file model, even if the language of thought hypothesis is rejected. Concepts are thus understood as files in their entirety, notes and all, and with at least some associated representations active and others primed (most of which occurs unconsciously) while the concept is being entertained. This is approaching the kind of structure that I envision serves the role played by k-systems in cognition developed in the previous chapter. Before explaining this in more detail, however, I will point out a problem that Barsalou's account faces, and offer a remedy. This will lead to an enrichment of conceptual structure better suited to the informational structure of k-systems.

4.2.2 A critique of the PSS theory of concepts

Barsalou has recently acknowledged the involvement of more than just perceptual systems in conceptual knowledge, including systems subserving language, affect or emotion, and statistical representation (Barsalou, 2009; Barsalou et al., 2008; Santos, Chaigneau, Simmons, & Barsalou, forthcoming). As we shall see below, he assigns a special role to the language faculty in particular for representing and processing concepts. Nevertheless, the involvement of these other cognitive systems in conceptual cognition is restricted to guiding conceptual processes, and they do not contribute to the nature of concepts themselves. That is, Barsalou continues to maintain that concepts are inherently perceptual—concepts are collections and activations of perceptual symbols, which are based in the perceptual systems and have been stored in long-term memory. (Although he is open to revising his position pending empirical evidence; Barsalou et al., 2008.)

One thing that Barsalou's theory fails to account for, however, is how we acquire concepts for which no perceptual events are associated, or for which no perceptual events are possible. Barsalou discusses how we might acquire and use abstract concepts, such as TRUTH or DISJUNCTION, via simulations of perceptual symbols of events (e.g., he claims that we can simulate TRUTH by simulating an event conveyed by a sentence, and then check to see if the simulated event—or something sufficiently close to it—is in the actual world). I am not convinced by Barsalou's account of how we might acquire and use abstract concepts, but I will not adjudicate the matter here. For in any event, these are not the kinds of concepts that I have in mind by those for which no perceptual events are associated or possible. Rather, the kinds of concepts that cause serious trouble for Barsalou's account are concepts like PROTON. A proton is a subatomic particle of which any type of perceptual experience is impossible. How are we to simulate PROTON if we cannot acquire any perceptual symbols for it? Certainly we can conceive a proton to be a tiny sphere spinning around, but no serious (contemporary) scientist believes this to be true. In fact, protons have a specific structure made up of three quarks. Protons also have a +1 elementary charge. What perceptual symbols would be used to simulate these properties? Analogies can be made with visual diagrams, but these do not adequately capture the conceptual information afforded by grasping the concept itself. Quarks are not circles, nuclear bonds are not squiggly lines, and a positive elementary charge is not a "+" symbol. This is not to say that these visual things cannot represent the properties of protons. But we need some requisite conceptual knowledge of the referents before we can understand the representation relation and what is being represented. A red octagon does not look like "Stop!" any more than a sphere looks like a proton, but we need knowledge of STOP if we are to understand that a red octagon represents a command to stop (e.g., via, perhaps, perceptual symbols having to do with ceasing motion). If someone does not already have some understanding of STOP, then she would not know what is being represented by the red octagon.

Imagine trying to teach the concept PROTON solely by perceptual means to a complete neophyte with respect to science. You might draw a few circles, some squiggly lines connecting them, and a "+" symbol, and then tell the neophyte that it is a diagram of a proton. You might even try to convey that a circle represents QUARK, a line represents NUCLEAR BOND, and the "+" represents POSITIVE ELEMENTARY CHARGE. You mention, however, that protons, quarks, and so on, do not *really* look like what is presented in the diagram. But you continue to enhance her perceptual experiences, maybe by presenting many such diagrams. The neophyte might then store her perceptual experiences as perceptual symbols in her long-term memory, integrate them in a simulator, and perform a number of simulations. It seems that according to Barsalou's theory, these perceptual experiences and her resulting perceptual symbols may be the only means to understanding the concepts PROTON, QUARK, NUCLEAR BOND, POSITIVE ELEMENTARY CHARGE, and really any concept for which it is impossible to have direct perceptual experiences (e.g., UNICORN). Yet most of us would doubt that the neophyte really does possess the concepts PROTON, QUARK, etc., based on her perceptual symbols and simulations of whatever perceptual information she received from the diagrams. At the same time, however, professors and teachers get their students to understand concepts such as PROTON, QUARK, etc., at considerable rates of success. Thus, our understanding of concepts in general cannot be owed simply to perception—something more than perception must ground our (human) concepts.[16] We might recall here Barsalou's claim that "a *concept* is equivalent to a simulator" insofar as a concept consists of "the knowledge and accompanying processes that allow an individual to represent some kind of entity or event adequately" (Barsalou, 1999, p. 587). It seems that it is "the

[16]This is to say nothing of what grounds the concepts possessed by other animals, or how other animals acquire concepts, or whether they possess or acquire concepts at all.

knowledge" part that requires further explication.

Natural language presents itself as a way around the problem I am posing for Barsalou's account of concepts—a way in which we might cash out "the knowledge", in addition to perceptual symbol simulations, in which a concept consists (or at least in which some concepts consist). In many ways, one can do more with language than what can be done by perception alone. Metaphor and analogy are just two examples.[17] It would be very difficult, if not impossible, to simulate with perceptual symbols: "Juliet is the sun". One might try to simulate this by somehow combining the perceptual symbols one has for Juliet and for the sun. But not only would this not convey the full intended meaning behind this statement, it would convey the wrong meaning. For Romeo, Juliet is the sun not in any literal sense, but in the sense that his love for Juliet overpowers his feelings for other women, and that (when one understands the play as a whole) Juliet is in a way necessary for his life. It is unclear how something like this can be simulated by perceptual symbols. And yet, all of us can very well understand the metaphor, at least after it is *explained*.

My suggestion, then, is that Barsalou's theory of concepts be augmented by natural language for the human case, since perceptual symbols do not have the resources needed to grasp many of the concepts we in fact grasp. This would entail two important related additions. First, simulators would be enriched by including linguistic forms; in terms of the file model, concept-files would contain as notes both perceptual and linguistic information. Second, natural language affords connections between simulators (or concept-files), thereby substantially increasing the operational power of our central cognitive processes. I will discuss these in turn.

4.2.2.1 Integrating linguistic information

Natural language increases the informational resources of our conceptual systems such that all our concepts are enriched. The idea is that linguistic information gets encoded and integrated into our concepts as we learn language(s), hear linguistic descriptions about the classes

[17] Lakoff and Johnson (1980) argue that our concepts are metaphorical in nature. I do not endorse their argument. However, there is an interesting connection, I think, between (our use of) natural language and *understanding* our concepts.

or referents of concepts, and form beliefs that are expressed in natural language. As a human infant learns a concept, she learns words and other linguistic forms accepted and used by her linguistic community to refer to the extension of the concept in question. I am not prepared to offer here a full theory of content or symbol grounding, but the process would likely occur by first learning the relation between the said linguistic forms and direct stimuli in the environment; and as her conceptual and linguistic skills develop, the infant would cognitively realize a similar relation between the linguistic forms and the perceptual symbols (stored in long-term memory) associated with the stimuli (cf. Viger, 2007). During this learning process, the infant would increasingly be able to recall and use the appropriate linguistic forms as she learns to adequately represent the concept in question. I am suggesting that the appropriate linguistic forms eventually get integrated in the simulator created for the concept just as multimodal perceptual symbols get integrated. More specifically, we may consider the neural activations involved in producing and processing linguistic forms (in the Broca's and Wernicke's areas) as "linguistic symbols" (to adopt Barsalou's terminology) on par with perceptual symbols. If this is accepted, then linguistic symbols can get coded and stored in long-term memory along with the other perceptual symbols associated with a simulator via the higher association areas of the brain that integrate activations across cognitive systems (i.e., the temporal and frontal lobes),[18] as Barsalou describes and outlined above. The only difference is the cognitive systems from which the symbols originate and are (re)activated upon tokening or simulating the relevant representations.[19] Hence, I contend that linguistic symbols come to be part of an individual's concept-simulator, or what amounts to the same on the foregoing account, linguistic symbols

[18]Recall that Barsalou includes the parietal lobe as an association area for perceptual symbols. I omit this lobe here since it does not seem to have much of a role with respect to processing language.

[19]My suggestion that linguistic symbols get integrated in a simulator inasmuch as perceptual symbols means that linguistic symbols would have to get bound up with perceptual symbols in specific ways. Barsalou borrows Antonio Damasio's (1989) idea of *convergence zones* to explain how the brain integrates different features of perceptual experience via conjunctive neurons to form perceptual symbols. According to Damasio, convergence zones are neural ensembles that store the relevant neural patterns active upon perception, and enact "binding codes" that specify the neural patterns to be (re)activated during mental representation (in the absence of stimuli). If this right, then my proposal would require that linguistic symbols are encoded in the binding codes along with perceptual symbols. At the very least, something in the neural operations underlying mental representation would have to (re)activate the neural patterns of linguistic symbols in addition to those of perceptual symbols.

serve as notes in a concept's file.[20]

If Barsalou (1999, p. 587) is correct that an individual is understood to grasp a concept to the extent that she possesses the knowledge that allows her to represent the concept to a culturally acceptable degree, then linguistic symbols ought to be considered part of a simulator inasmuch as perceptual symbols are, since the appropriate application and use of linguistic forms is certainly indicative of the knowledge one possesses with respect to the concept in question. For example, experiments by Karen Wynn (1990, 1992) reveal that most children begin counting when they are two or three years old. However, it is not until months or years after that they are able to understand the meanings of the number words, or the meaning of the counting routine generally. For instance, children between 2 and 2.5 years of age can count; but if they are allowed to count an array of objects and then are asked to give the experimenter, say, four objects, they grab a handful of objects at random, notwithstanding that "four" is a part of their counting routine. This is certainly evidence that such children do not grasp natural number concepts, even if they are able to employ the words designating such concepts in a limited range of situations. Thus, coming to understand a concept includes being able to *appropriately* employ the correct linguistic forms in the *correct* circumstances;[21] and according to the view I am suggesting here, this is facilitated by having specific linguistic symbols integrated in the concept's simulator or file.

Allowing for linguistic symbols to be integrated in simulators or files enriches the content of concepts in significant ways. For not only do linguistic symbols afford additional resources to token conceptual representations in different ways (i.e., via various linguistic forms, in addition to perceptual symbols), but linguistic forms may be used in sentences of natural language

[20]Throughout the remainder of this dissertation I shall employ talk of simulators and talk of files more or less interchangeably.

[21]Of course, this does not mean that we must get things right all of the time if we are to possess a given concept. I am sure that competent people sometimes (or even often) inappropriately employ words. The point I am making, however, is that, if one understands a concept, one would be able to correctly and appropriately employ its referring word(s) *on the whole* or *in general*; and one's competence can be assessed according to the standards of one's community. Let me point out also that this applies only to linguistically competent humans. We certainly would not hold a nonlinguistic, nonhuman animal to this criterion; and we probably would not hold humans suffering from aphasia to this criterion.

to produce novel descriptions of conceptual content without necessarily invoking perceptual symbols. Thus, when I entertain the concept CAT, I can engage in *cat*-simulations *or* I can activate the linguistic symbol for "cat". Similarly, I can represent various aspects of cats or a particular cat by using my perceptual symbols (e.g., the furriness of cats), *or* I can describe these aspects, mentally to myself or aloud to others, by linguistic means (e.g., expressing the belief *cats are furry* by saying "cats are furry").[22] The "or" in these cases should be understood as inclusive here, for we have no reason to believe that thinking in language cannot be accompanied by perceptual symbol simulation, or *vice versa*. Moreover, if Barsalou is right in that perceptual symbols can be entertained, and that simulations can occur, unconsciously, then we might expect that linguistically coded information, such as beliefs, can be entertained and processed unconsciously as well. In fact, that sentence parsing and other linguistic processes occur unconsciously suggests that this is not implausible.

Furthermore, on the present account it is entirely possible to learn a word in natural language without any conceptual content save for grammatical information, such as whether the word is a noun or verb. At this stage, a person—who can be an adult or child—would know the role of the word in language but nothing about the concept the word names. And so the person could very well use the word properly in some contexts, but would not be able to appropriately employ the word in all the correct circumstances. Over time and with experience, however, the person would learn what the word refers to and would thereby build up conceptual information (filling the concept-file with notes), and this could continue until the person has sufficient knowledge to grasp the concept in question. This example shows that a word token can initially create a concept-file which is then subsequently filled with conceptual content (notes). The result is a concept which one can represent with linguistic or perceptual information (together

[22]The question may arise here as to whether sentences are included among the linguistic forms that get integrated in concept-simulators or files. My answer would be the following. Whatever neural activity it is that constitutes a linguistic symbol and gets (re)activated upon tokening conceptualizations of a given concept is that which is a proper part of the concept's simulator. I see no reason thus far to preclude the possibility of whole sentences being (re)activated in certain conceptualizations. I suspect, however, that limitations on memory recall and other cognitive resources would restrict the length of sentences that get integrated in a concept's simulator, if sentences are the units available for integration at all.

or independently, as mentioned). Indeed, this seems to be how we in fact learn some of our concepts.[23]

The integration of linguistic symbols in concept simulators or files also explains at once how one can acquire and appropriately employ concepts for which there are no possible perceptual experiences (such as PROTON), as well as how one can learn abstract concepts more generally. A student can learn PROTON because the teacher explains in natural language that protons are subatomic particles composed of three quarks, etc., while explaining at the same time the concepts QUARK, NUCLEAR BOND, etc. (the teacher can also draw diagrams to supplement the learning process). The student's simulator for PROTON would thus consist mainly of linguistic symbols with which she would be able to use as and productively incorporate in locutions pertaining to protons and related matters. The student is subsequently prepared to draw inferences about protons and comes to have beliefs about such entities, and hence she grasps the concept PROTON. The same sort of learning process would apply to how one can acquire concepts for nonexistent entities, such as UNICORN, along with abstract concepts. The import of this picture is great, since many of our concepts are initially formed in the absence of direct perceptual experience. As Christopher Viger (2007) observes, "My acquisition of 'shark' was more like my acquisition of 'unicorn' than that of 'tree' " (p. 133). In a word, linguistic symbols are constitutive of most, if not all, of our concepts, and have a significant role in conceptual cognition.

The importance and utility of concepts containing linguistically coded information (as part of a simulator or as notes in concept-files) extends further. For there are certain events that language can represent generically. Consider, for instance, the thought *Alice stole the pen*. This thought need not be completely perceptually grounded.[24] The manner in which Alice stole the pen may be irrelevant; it might not matter whether Alice took the pen off Bob's desk and slyly placed it in her pocket, or whether Alice forcefully grabbed the pen out of Bob's

[23] I owe this point to Chris Viger. See below.

[24] I say "completely" because there may be certain aspects of this thought, such as *Alice* or *pen* that is perceptually grounded, notwithstanding that the entire thought *Alice stole the pen* is not.

hand as he was writing with it, or whether Alice borrowed the pen and then decided to keep it without permission. If the thought *Alice stole the pen* would be completely perceptually grounded, specific perceptual symbols would have to be activated for specific simulations, and thus such details (e.g., the manner in which the pen was stolen) would be filled in. However, if the thought were to be linguistically coded, "Alice stole the pen", all that matters (viz. that Alice stole the pen) can be coded, brought to mind, and conveyed, and irrelevant details need not be represented. In a way, linguistically coding representations saves cognitive resources in storing and entertaining certain thoughts. But more importantly for the present purposes, it is a means by which specific, relevant information gets represented without representing what is irrelevant.[25]

4.2.2.2 Affording connections between concepts

So far I have discussed the way in which augmenting Barsalou's theory of concepts with natural language would enrich the simulators or files of concepts. As indicated above, the other role for natural language is to provide a medium for establishing connections between simulators (or concept-files), thereby substantially increasing the operational power of our conceptual cognition. The idea here is that natural language is a powerful combinatorial system that affords a means by which to combine diverse representations from different systems—an ability that otherwise may not have been possible. This suggestion follows the accounts offered by Elizabeth Spelke (2003a, 2003b, 2009; Condry & Spelke, 2008; Hermer-Vazquez, Spelke, & Katsnelson, 1999) and Viger (2005, 2006c, 2007).

Spelke (2009) observes that many species are able to combine representations from differ-

[25]Barsalou asserts that perceptual symbols can be indeterminate and generic representations insofar as qualitatively specified neurons can encode specific qualities but not others (as explained above). His example is of a triangle: A given set of qualitatively specified neurons can encode the presence of lines without encoding length; length may be encoded by a different set of neurons. Thus, we can have a generic representation of a triangle in the sense that a triangle can be represented without having a specific size. Though I do not doubt that this is possible—and in fact, Barsalou cites empirical support for the existence of qualitatively specified neurons—this is not the kind of generic representation that I have in mind here. For it does not appear that representations like *Alice stole the pen* can be represented by qualitatively specified neurons in the same way as *triangle*, and there is no evidence that this can be done neurologically. With the triangle case only part of what is perceptual is specified (i.e., shape), whereas with the stolen pen case nothing perceptual is specified in the relevant sense.

ent cognitive systems, at least to a certain extent, in making inferences. It is possible that an organism's cognitive architecture enables such combinations by embodying certain pathways between representational systems; and some pathways may get established through the neural/cognitive development of an organism. It is also possible that an organism may learn to combine representations from different systems, though this would be a slow and piecemeal process. Humans, on the other hand, have available in addition to these two options a capacity that enables arbitrary combinations between representations, and as Spelke contends, this combinatorial capacity may be identical to the human capacity for natural language. Spelke acknowledges that the human brain might develop a rich combinatorial power *per se*, but she believes that this power is at least bolstered or guided by, and mutually developed in conjunction with, the development of natural language. Spelke cites empirical evidence to support her position. For the sake of brevity, I will only discuss one kind of evidence here.

Spelke and her collaborators (Gouteux & Spelke, 2001; Hermer & Spelke, 1994, 1996; Wang, Hermer-Vazquez, & Spelke, 1999) reveal interesting limits on navigational abilities in children. In their experiments, children up to 2 years of age were tested in reorientation tasks in which an object was hidden in a room, the child was disoriented, and then asked to find the object. It was evident that the children were able to represent that the target object was located, for example, *at a corner with a long wall on the left*, but they were unable to represent that the object was *at a corner with a blue wall on the left*. According to Spelke, the children were unable to combine geometrical representations (e.g., *left of a long wall*) with object-property representations (e.g., *blue wall*) in order to represent the locations of the target objects.

As may be expected, human adults exhibit the ability to combine such geometrical and object-property information under similar circumstances with ease. Spelke claims that the difference in performance between the young children and adults can be accounted for by the emergence of spatial language. The developmental research of Hermer-Vazquez, Moffett, and Munkholm (2001) reveal that the transition to flexible navigation in humans occurs around the time when children master spatial expressions in natural language, such as "left" and "right".

Furthermore, performance on productive language tasks with items involving the terms "left" and "right" turn out to be the best indicator of success on reorientation tasks for children at this transitional age. These findings suggest that natural language has a key role to play in navigation for humans. This suggestion is supported by further experiments on human adults. For instance, Hermer-Vazquez et al. (1999) presented human adults with navigation tasks accompanied by interference tasks. Participants exhibited a general impairment in navigation performance for all cases, but those who were given an interference task that specifically interfered with language production performed like the young children—they were unable to reorient with respect to nongeometric properties.

Spelke goes on to suggest that natural language assists in improved navigation performance due to the combinatorial properties of language. The property of compositionality of natural language allows speakers to represent the meanings of arbitrarily many combinations of words, once the meanings of a given set of words and the syntactic rules for their combination are known. Moreover, natural language crosscuts representational systems, and in this sense it acts as a domain-general system of representation. Thus, natural language allows linguistically-competent people to combine information in such a way that allows them to form representations that might not have been otherwise possible. With respect to the spatial and navigation tasks,

> the combinatorial machinery of natural language allows children to formulate and understand expressions such as *left of the blue wall* with no further learning. This expression cannot be formulated readily outside of language, because it crosscuts the child's encapsulated domains. Thanks to the language faculty, however, this expression serves to represent this conjunction of information quickly and flexibly. (Spelke, 2003b, p. 296)[26]

Though the evidence discussed here is only with respect to spatial representation, I believe that the role of natural language can be generalized. Indeed, Spelke (2003a, 2003b, 2009) argues that natural language enables the creation of new concepts and changes within concepts,

[26]Spelke (2000, 2003a, 2003b; Spelke, Lee, & Izard, 2010; Spelke, 2007) also provides convincing evidence that shows that distinct and more-or-less encapsulated systems in the human brain produce representations of the geometric properties of spatial layouts, on the one hand, and properties (e.g., colour and shape) of objects, on the other. The work of Susan Carey (2004, 2009) also suggests this.

and that natural language is significantly involved in our representational capacities with respect to artifacts, natural number, and social cognition. Extending this idea further, there is no principled reason why the productivity, compositionality, and combinatorial power of natural language cannot be harnessed in cognition generally. The latter is suggested by Viger (2005, 2006c, 2007). Viger argues against the language of thought hypothesis, claiming that concepts need not possess the logical or compositional relations exhibited in natural language, and also that the language of thought hypothesis does not possess the wherewithal to deal with the problem of how conceptual symbols are grounded. According to Viger, our conceptual apparatus is enhanced once we learn natural language. He therefore advances what he calls the *acquired* language of thought (ALOT) hypothesis: "some of our *ability* to think is acquired in learning . . . a natural language" (Viger, 2005, p. 322). Unlike the properties ascribed to concepts by the language of thought hypothesis, natural language terms are connected to each other in many different associative respects. And so tokening a natural language word, for instance, can co-token other associated words. Whether and to what extent co-tokening occurs will depend on the organization of connections among words in cognition, but if the foregoing view is correct, the result is an interconnected network of words that allows for any attending content to *in principle* activate any other content (which, as discussed in the previous chapter, is paradigmatic of central cognition; Viger, 2006c). "The interconnected network is like a roadway through the mind, a constructed central superstructure that makes for easier travel among the modular cognitive operations embedded within this superstructure" (Viger, 2007, p. 135). Thus, in a sense, natural language encodes information in our concepts in such a way that makes it easier to activate certain cognitive operations (Viger, 2006c). This picture is certainly consistent with how Barsalou envisions conceptual thought.

4.2.3 Putting it together

The application of Spelke's and Viger's work to my account of concepts and cognition is best understood in terms of the file model of cognition. It may be recalled that Fodor believes that one can think in file labels without invoking any content (i.e., any notes in the file), and that I argued against this view since Fodor takes file labels to be concepts. I had proposed earlier that concepts be understood as files in their entirety. I now further propose that we revise the file model such that natural language representations (linguistic forms) that name concepts are the labels attached to the files. On this revised model of concept-files, it may very well be possible for the language faculty to token file labels in the absence of conceptual content (cf. my comments above regarding learning a word and building up conceptual content afterward). That is, it is possible that we can entertain natural language words without invoking any content, such as when I mindlessly associate words like "salt" and "pepper". Fodor endorses such associations between words (see Fodor, 1983, pp. 78-83; Fodor, 2008, pp. 98-99). However, Fodor "suggests ... that associations are the means whereby stupid processing systems manage to behave as though they were smart ones" (Fodor, 1983, p. 81).[27] On my account, on the other hand, associations among words offer a means by which smart systems exercise their intelligence. Furthermore, unlike Fodor, I have two distinct ways that associations can occur, namely among items within a file *or* between natural language words *qua* file labels. This bimodal means of association thus provides resources for thought and reasoning within and between files, and this would certainly play a role not only in making inferences, but also in reinforcing and corroborating inferences.

Perhaps more importantly, if we understand labels as linguistic forms, then the language faculty can operate over file labels, and thereby enable the connections and combinatorial operations suggested by Spelke. For it is in virtue of being linguistic forms that file labels would possess the formal syntactic properties that lexical items have in a natural language. Thus, according to my view, lexical items *qua* linguistic symbols can be integrated in and serve as labels for concepts, and thereby serve as units over which different cognitive systems can operate (cf. Viger, 2006c). The language faculty may therefore independently operate on file labels

[27]To give this passage context, Fodor is arguing that informationally encapsulated systems (namely, modules) are stupid systems, and so they do not intelligently evaluate information; according to Fodor, one recourse is for them to operate according to associations.

to establish the connections initially needed to bring diverse concepts to bear on one another. Indeed, words are independent objects—independent from concepts—and therefore they can be independently manipulated. Thus, as the language faculty operates over file labels *qua* lexical items, all the linguistic machinery can be brought to bear when thinking in file labels. We have noted that the linguistic machinery includes the combinatorial power of language, which, as suggested, can enable connections between and combinations of concepts. Natural language may not be necessary in combining content across concepts, but it certainly seems as if it plays a significant role in facilitating the process.

In addition, as intimated above, the linguistic machinery includes the grammar of natural language, which can yield important and rich information about the concept the label names. For example, the label for one's CAT concept is a noun (i.e., "cat"), and "cat" does not act as, say, a verb in language. Information such as this can indicate something about the nature of CAT—at the very least, it indicates that it is a thing—and this in turn tells us something about the role the concept plays (or can play) in thought.

Construed as such, file labels really are Janus-faced, but not precisely in the way that Fodor (2008, p. 100) described (viz. "one face turned towards thinking and the other face turned towards what is thought about"). In light of the present considerations, it is more appropriate to claim that file labels are Janus-faced where one face is turned toward the concepts they label (what the thought is about) and the other face turned toward their role as a lexical item in natural language.

I have shifted here from talking about concepts in terms of simulators to talking about them in terms of files, but this was mainly for expository purposes. In terms of simulators, we can envision a natural language form *qua* linguistic symbol to be integrated into a simulator in such a way that it is flagged and understood to name (or label) the simulator. I suppose this happens by a neural ensemble that encodes in a specific way the relevant neural patterns in tokening the linguistic symbol (cf. Damasio, 1989; see footnote 19 above); this neural ensemble would thereby specify the neural patterns that point or refer to the concept, and in this way the linguistic symbol acts as a symbol insofar as it names or stands in for the concept in question. In this way, a linguistic form can function in cognition as a symbol that represents or expresses a concept.[28]

As explained above, I am assuming that linguistic forms are grounded in the linguistic systems of the brain and (re)activated as linguistic symbols. The language faculty may be an independent representational system, but the lexicon can be understood to exist separately from the linguistic system. So linguistic symbol activations can belong to a simulator, where the linguistic symbol continues to be operated on by the language faculty according to the formal syntactic properties it exhibits as a lexical item in natural language. Thus, simulators can encode natural language forms *qua* linguistic symbols, and produce simulations and conceptualizations with them (where the content of the conceptualization, it may be recalled, will depend on the context and cognitive task). As an independent system, the language faculty can process the linguistic symbols flagged as naming simulators and thereby make connections between simulators according to the syntactic properties of the linguistic forms as exhibited by natural language. Subsequent simulations then serve as deeper conceptual processing, providing deeper integrations of conceptual information between concepts, notwithstanding that language initially makes possible many connections between concepts, as Spelke contends (rightfully, I believe). Moreover, linguistic symbols activated by simulation will suggest further conceptual connections, since the linguistic symbols embedded within a simulator can very well pick out the names of other simulators.

In summary, my suggestion is that Barsalou's theory be augmented by natural language. Natural language enriches the content of simulators by allowing the integration of linguistic symbols; in terms of the file model, concept-files contain as notes both perceptual symbols and linguistic forms. Furthermore, if we understand file labels to be natural language words, then, following Spelke and Viger, natural language affords combinatorial resources needed to

[28]This approaches what Viger (2007) claims. He posits the existence of *interface structures* between cognitive systems that are activated during object recognition. Viger asserts that "a symbol is neurally encoded with a projection to the interface site of what the symbol stands for" (p. 132).

make connections between concept-files or simulators and their content. Natural language thus enables more powerful cognitive operations than would otherwise be available.

It should be acknowledged that Barsalou (1999) assigns a role for natural language in his symbol systems theory. However, the role he describes is not precisely the role I am advocating here. Rather, Barsalou envisions linguistic symbols as perceptual symbols. More specifically, he claims that spoken and written words are the perceived events from which selective attention focuses and extracts perceptual symbols from the modes of perception (typically of hearing spoken words or seeing them written), which then "become integrated into simulators that later produce simulations *of these words* in recognition, imagination, and production" (Barsalou, 1999, p. 592, my emphasis). The simulators for words are subsequently associated or "linked" to the entities or events they name, or more specifically, to the concepts the words label. Once this linking happens, according to Barsalou, words can control simulations: "On recognizing a word, the cognitive system activates the simulator for the associated concept to simulate a possible referent. On parsing the sentences in a text, surface syntax provides instructions for building perceptual simulations" (ibid.). Barsalou goes on to claim that "the productive nature of language, coupled with the links between linguistic and perceptual simulators, provides a powerful means of constructing simulations that go far beyond an individual's experience" (ibid.).

Nevertheless, I do not believe that these simulations will go far enough. For notice that Barsalou is not claiming that words or linguistic symbols are integrated into the same simulator of the concepts they label. Instead, words have their own simulators, the constituents of which are perceptual symbols for the perceived events of hearing spoken words or reading written words; these simulators then become linked to the simulators (or concepts) they express. On this account, then, Barsalou still faces the problem presented by nonperceptual concepts like PROTON. Certainly, one can develop a simulator from the perceptual symbols extracted from perceptual experiences of hearing or reading "proton" (call this the ⌜PROTON⌝ simulator, since the concept is not PROTON but of the word "proton"), but without any perceptual experiences from which to extract perceptual symbols for PROTON itself, there would be nothing to link ⌜PROTON⌝ to. This is similar to the problem I discussed earlier about needing requisite knowledge of STOP to understand that a red octagon means "Stop!".

In contrast, on the account I am suggesting here linguistic symbols are not merely linked to the concept(s) they express, but get integrated into the simulators themselves as, for example, words or beliefs expressed in natural language, and some word(s) would function as names for the simulators (concepts). Natural language words and sentences may not be amodal,[29] just as Barsalou claims, but as I argued above there is a way in which language enables its users to abstract away from modalities to express a proposition, make an exclamation, ask a question, perform a speech act, or whatever. This is what allows us to understand what protons are (along with everything else we cannot perceive), as well as what allows us to represent logical entities and mathematical objects, have thoughts of impossible or nonexistent entities (such as UNICORN or GOD), and have coherent and robust abstract concepts.[30] Information is encoded in language differently than in perceptual symbols, and as such, the kind and the amount of information that can be expressed in either format differs from each other. As should be apparent given the discussion above, thinking *the cat is on the mat* in natural language expresses the proposition that the cat is on the mat, whereas thinking it in simulations of perceptual symbols expresses the same content but also carries much more information—perceptual information, to be precise. And the degree to which simulations in perceptual symbols carry more information depends on the manner in which *the cat is on the mat* is simulated (e.g., depending on whether the cat is simulated to have a specific orientation; whether the cat is simulated to have a specific colour; whether the cat is simulated as a specific breed; whether the mat is simulated to have a specific colour or pattern; etc.).[31] Thus, although linguistic symbols as constituents of natural language

[29]It might be more appropriate to say that natural language words and sentences may be amodal in certain respects but not others. Word or sentence tokens that are spoken or written on a page, for instance, are certainly not amodal, as they involve the sensory modalities of audition and vision, respectively. However, it is not so clear whether employing a word or a sentence in thought and grasping its meaning involves any sensory modality.

[30]As I had indicated above, Barsalou argues that his PSS theory can account for representing abstract concepts. He uses the examples of TRUTH, DISJUNCTION, NEGATION, and ANGER. However, I do not think his theory can account for all abstract concepts. (I believe that his account of ANGER is particularly inadequate.)

[31]Cf. the discussion above in section 4.1.1 regarding what can be expressed by images as opposed to what can

can be the stuff of thought and can express some of the same information as perceptual symbols, either format will embody and express different concomitant information.[32]

It might be noticed, however, that Barsalou recognizes that natural language plays a central role in producing simulations. This is indicated in the passage cited above: "the productive nature of language, coupled with the links between linguistic and perceptual simulators, provides a powerful means of constructing simulations" (Barsalou, 1999, p. 592). More recently, Barsalou has discussed an even more prominent role for natural language (Barsalou, 2008a; Barsalou et al., 2008). For instance, Barsalou et al. (2008) assert, "In general, we assume that linguistic forms provide a powerful means of indexing simulations (via simulators), and for manipulating simulations in language and thought" (p. 252). This approaches the role of natural language assigned by Spelke and Viger, and the one advocated here. In more specific terms, Barsalou et al. believe that associations between linguistic forms generate "pointers" to associated conceptual information. They claim that humans can employ "linguistic strategies" to facilitate easy but superficial, shallow inferences. Such superficial linguistic strategies can be highly effective, but according to Barsalou et al., they do not directly involve deep conceptual processing; conceptual processing is supposed to occur only via simulations. In a little more detail:

> we assume that the syntactic structure of sentences controls the retrieval, assembly, and transformation of the componential simulations that people integrate to represent sentences and texts. Similarly, we assume that interactions between the two systems are responsible for the representation of propositions, conceptual combinations, productively produced phrases, recursively embedded structure, etc. In general, we assume that symbolic structures and symbolic operations on these structures emerge from ongoing interactions between the language and simulation systems. (pp. 272-273)

---

be expressed by propositions. It should be noted, however, that nothing I am saying here hangs on whether images can express propositions—I am not here talking about the *proposition phrase* "The cat is on the mat", but rather the *content* of the proposition.

[32] I suppose that there can be natural language simulations analogous to perceptual symbol simulations. With respect to natural language, however, simulations would occur as rehearsals of some form of "inner speech" (cf. Carruthers, 2006a). The sentences that this inner speech would produce would be various descriptions of the associated concept in natural language, and there may also be a kind of inner conversation—a talking to oneself, as it were—in which various questions are posed and answered. And, like perceptual symbol simulations, natural language simulations mainly occur unconsciously. This is mere speculation, however. None of what I have to say here depends on the details of such natural language simulations, or indeed whether they actually occur.

> [B]oth systems are probably essential for achieving the powerful symbolic abilities of the human cognitive system. Neither alone is likely to be sufficient for symbolic behaviour. ... Across many abilities, the two systems work together to achieve the power and distinctive properties of human intelligence. (p. 274)

We might therefore observe that the present account of cognition is at least consistent with Barsalou's recent writings. However, to repeat, my account of concepts allows for the integration of linguistic symbols in simulators, and I believe this is needed to account for representing abstract and nonperceptual content. I also believe that the integration of linguistic symbols in simulators aids in establishing and maintaining connections between simulators (concepts), as well as in offering a means by which certain linguistic symbols (words) can serve as labels that name or stand in for concepts. This indicates another important respect in which my account differs from Barsalou's, namely having to do with the feature of concepts that allows natural language to negotiate conceptual processing. For Barsalou and his collaborators do not share the reconciled file model in which simulators have labels in natural language, which enable the linguistic system to connect concepts and representations across domains, and exploit their content. Barsalou et al. (2008) assign natural language a substantial role in cognition, but they do not adequately explore what it is about concepts that allows natural language to connect and combine them. In fact, Barsalou et al. assume, following Barsalou (1999), that simulation mechanisms implement combinatorial symbolic operations (see Barsalou et al., 2008, p. 251), not natural language operating over linguistic data integrated in concepts as labels.

I should note that this account is meant to be neutral with respect to the language of thought hypothesis. The foregoing view is that *natural* language *augments* perceptual symbols in our conceptual apparatus in more elaborate ways than proposed by Barsalou. As such, the present account enriches Barsalou's theory of concepts. Yet, although Barsalou's theory rejects the basic tenets of the language of thought hypothesis, I have not given any arguments against the possibility that the stuff of thought is in fact language-like representations. All I have argued is for an interrelation between concepts (as simulators or files) and language.

Given this initial discussion on the account of concepts that I am adopting here, we are now in a position to understand more fully the role of heuristics in cognition, and more specifically how heuristics work. In the next section I will explain how the foregoing account of concepts fulfill the role of k-systems envisioned in the previous chapter. Examples illustrating my claims will be provided in the following chapter, wherein I discuss how some of the heuristics proposed by Kahneman and Tversky, and Gigerenzer, can be understood in terms of the theory of concepts and cognition proposed in this dissertation.

4.3 From k-systems to concepts

Recall that a k-system is an informationally rich system of knowledge or representations. K-systems encode various kinds of information in a highly organized fashion—information is encoded such that specific relations hold between items of information within systems, as well as between different systems. As explained in the previous chapter, k-systems are what facilitate the kind of reasoning we witness from humans—reasoning that is quick and efficient, and that is frugal in its computations. I claimed that without the right kind of architecture and the right kind of structure and relations among items of knowledge in our heads, we would be unable to make the inferences that we in fact do—inferences which underlie much of our cognition. I want to now claim that concepts, as explicated above, are the cognitive structures suited to the role of k-systems. More specifically, I want to claim that the way heuristics work is by exploiting concepts, and by the same token that concepts constrain heuristics.

Given the account of the nature of concepts and conceptual cognition just presented, it is not hard to see why I believe concepts to be the kind of cognitive structure that I described k-systems to be. It may be recalled that the function of k-systems is to deal with high information-load and cognitively-demanding problems, paradigmatically ill-defined problems (which are, again, the problems that we typically have to deal with). I argued in the previous chapter that humans characteristically handle such ill-defined problems successfully and solve them because we exploit the information embodied by our k-systems. In light of my explication of the nature of concepts, however, the information we exploit can be seen to be actually embodied by our concepts.

Last chapter I had discussed the role of information in the environment with respect to cognition and k-systems. Much of the burden placed on cognition can be off-loaded into the environment. Information can be stored and made to be organized in specific ways by means of notepads, books, calendars, pictures, and so on. Moreover, external devices such as calculators and computers can aid in problem-solving. These technologies can certainly mitigate cognitive costs. I claimed in turn that the requisite knowledge needed to exploit epistemic technology must be the structured and organized k-systems I described (cf. Sterelny, 2004, 2006). We can see more clearly now that one in fact requires the appropriate concepts and conceptual knowledge to suitably exploit epistemic technology. For epistemic technology can be as informationally rich as we like, but we must conceptualize the technology in certain specific ways if we hope to exploit its embodied information in the right ways. Consider, for example, a map. A map encodes information about topography, landmarks, and distances, but if one does not recognize and conceptualize a map *as* a map, then one cannot hope to exploit any of the information it embodies (cf. Sterelny, 2004). At the very least, one must have the concept MAP, with the appropriate representational content, to read a map in the right ways.

Imagine that Alice and Bob come across a piece of paper with various markings on it, mostly lines and geometric figures but no words. Alice and Bob both examine the paper and its markings. Bob cannot make sense of it, and he therefore believes that it is just a bunch of doodles. On the other hand, the markings on the paper activate MAP symbols for Alice. She thus begins to conceptualize the piece of paper as a map, and she begins to conceptualize the lines as streets, the squares as city blocks, the blue markings as bodies of water, and so on. Thus conceptualized, Alice can exploit the representational content of the map. If Alice figures out which markings on the map correspond to what landmarks, she can triangulate her location on the map, she can figure out what direction to walk to get to the shoreline, she can infer that there is a railway two blocks north, and so on. Bob on the other hand cannot exploit any such information.

Notice, however, that Alice's map concept must be *sufficiently rich* if she is to exploit the map's information. For instance, if it is not in Alice's map file that maps usually represent water with the colour blue, then she would not understand that the blue markings she sees indicate bodies of water. Or if Alice did not know that maps are often not in the business of accurately representing the colour of the land, then she might wrongly believe that the map conveys information about land colour. Thus, knowing that a piece of paper is a map is not enough to appropriately conceptualize the information embodied by the map. This is to say that our concepts must be rich enough to afford the conceptualizations necessary to exploit information possessed by epistemic technology. In fact, being able to conceptualize certain things in certain ways in the first place is what allows us to create epistemic technology *qua* objects that store and encode certain kinds of information in certain ways.

To recall the Dretskean theme pointed out in the previous chapter, what information a signal carries depends on what the receiver already knows about the source. When confronted with epistemic technology, or any object at all, our sensory experiences embody a great variety of information. Our concepts are cognitive structures that enable us to extract and exploit certain features. In this sense, learning or enriching a concept provides us with the ability to decode certain aspects of our sensory experience in such a way that we cognitively respond in certain ways. In Dretske's terms, concepts allow us to digitalize information from an analog signal.[33] A map might contain information about the topography of the land, and anyone who views the map is seeing this information. But those who do not have a rich enough concept of map to recognize and interpret topographical information are incapable of extracting or attending to such information, and therefore incapable of cognizing it or allowing it to affect one's cognitive dispositions and behaviour. In Dretske's words:

[33]Dretske believes that concepts are cognitive structures that decode analog signals to digital. In a sense, I believe Dretske is right, although perhaps not completely. It may be more accurate to claim that concepts are cognitive structures that decode some of the information in an analogue signal. What gets decoded may not be fully digital, although this can be the case—natural language, for instance, appears to be digital. (Cf. my remarks above concerning thinking in images in section 4.1.1.). At the very least, however, my position is that having the appropriate perceptual symbols and/or linguistic information is what would enable one to decode various aspects of an analog signal.

> Learning to recognize and identify daffodils, say, is not a process that requires the pickup of more information from (or about) the daffodils. ... The requisite information (requisite to identifying the flower *as* a daffodil) is getting in. What is lacking is an ability to extract this information, an ability to decode or interpret the sensory messages. What [a] child [who is learning to recognize daffodils] needs is not more information about the daffodil but a change in the way she codes the information she has been getting all along. Until this information (vis., that they are daffodils) is recoded . . . , the child *sees* daffodils but neither knows nor believes that they are daffodils. (Dretske, 1981, p. 144)

According to Dretske, developing a concept DAFFODIL is what will enable the child to identify daffodils, to know and believe that certain flowers are daffodils, and to behave appropriately.

As explained above, natural language can facilitate connections between concepts. However, deeper conceptual processing also has a role in combining and integrating the representational content of concepts (notwithstanding, however, that language makes possible many connections between concepts in the first place). Once a concept becomes established, and as knowledge is accumulated and added to the concept, one is put in an increasingly better position to make inferences about the objects or events picked out by the concept. Let us recall here Barsalou's claim that, since a concept contains an enormous amount of (multimodal) knowledge and information, one can simulate aspects that go beyond what is perceived, and thereby make various inferences about an entity. As described above, for example, the concept JET might suggest that a perceived jet contains passengers, luggage, pilots, and so on; that the jet will continue to fly horizontally and in a single trajectory, not vertically or sideways; and so on.[34] Thus, the conceptual content[35]—the representations and conceptualizations—of active or primed concepts guide inference. Furthermore, what inferences one can make, and the relative ease with which one makes inferences, is constrained by the concepts one possesses and the relations between them. This is because the associative connections that exist within and between concepts inhibit certain inferences while facilitating other inferences. This was illustrated by the example above: the ease with which inferring that your pen is in your desk drawer

[34] We should keep in mind here, however, the hesitations outlined above in section 4.1.1 with respect to this example.

[35] For the sake of brevity, I will refer to the perceptual symbols, simulations, beliefs, and other linguistically coded information (i.e., the representations and conceptualizations) embodied by concepts simply as "conceptual content", unless there is a need to be more specific.

will be much easier and more natural than inferring that your car is not in your desk drawer. Pens are the kinds of things that are typically in desk drawers, while cars are not. Thus, we typically have positive experiences of pens in desks, and these experiences and representations are stored in our memory as items in the respective concept files. On the other hand, we have no experiences of cars in desks, and so we will have no such representations in our files. You can therefore just recall from memory that your pen is in your desk drawer, whereas you have to do cognitive work to infer that your car is not in your desk drawer. In fact, one's conceptual content of our CAR and DESK concepts will generally inhibit the inference that one's car is in one's desk drawer, since (as I observed above) cars are not the kinds of things that can fit in desks. And even when we do the cognitive work to make the inference, we may still resist it: if we were to hear someone say "My car is in my desk drawer!" we would likely assume that a joke was being told or that a toy car was being referred to, rather than that someone's actual car was in a desk drawer.

4.3.1 What this means for heuristics

Now that we have done all this work on understanding the architecture of cognition, and the nature of concepts, let us see what it does for our investigation into how heuristics work. As I was saying, concepts guide and impose constraints on inference. Yet it is certainly possible that one can reason beyond the constraints imposed by the conceptual content of active and primed concepts in making inferences. The latter impose only initial constraints on inferences. If one were to thoroughly reason through a problem by fully exploring potential outcomes, solutions, or consequences, and the means to achieve them, one could make any number of inferences. Heuristic reasoning, however, does not proceed this way. Rather, heuristics are completely subject to the constraints imposed by the conceptual content of active and primed concepts. And this is what makes heuristics work the way they do. Let us recall, once again, the characterization of "heuristic" that I developed in chapter 2.

> **H_7** Heuristics are cognitive procedures that satisfice, and that require little cognitive resources for their recruitment and execution; they operate by exploiting

informational structures.

Heuristics, by their very nature, are not procedures that thoroughly explore problems (i.e., they satisfice), and so they operate according to the information that is made available (in terms of what is activated or primed) throughout the satisficing process. How much information is surveyed will depend on the heuristic used, the nature of the task, the goals of the agent, and constraints on time, resources, etc. Sometimes conceptual information will continue to get activated or primed until a goal is met—until something "good enough" (i.e., that satisfices) turns up. Other times, we may not need to go too deep into our concept files to activate conceptual information, if we arrive at our satisficing answer quickly. And we might not even need to use a heuristic if we get the correct answer right away. However, the information made available to heuristics will generally not be sufficient to make deep conceptual inferences, but sometimes (if the conceptual information is sufficiently organized) heuristics can produce good inferences based on present relations and activations. The aspiration level set for a given heuristic can entail that a certain (small) number of inferences is made before reasoning is terminated (e.g., a decision is made, or one capitulates), but the exploration of the problem will usually not go too far. Thus, the operations of such a heuristic will remain constrained by the concepts that get activated or primed throughout the satisficing process.

As explained in chapter 2, the satisficing nature of heuristics contributes to their requiring little cognitive resources for their recruitment and execution. But the way concepts constrain the operations of heuristics is what really enables heuristics to be frugal. Much of the previous chapter was devoted to showing that the organization among k-systems bears a lot of the burden entailed by high-information-load problems. Now we can understand that it is in fact the organization of the perceptual symbols and linguistic information of our concepts that bears the burden entailed by high-information-load problems. The productivity and combinatorial power of natural language is certainly a factor in mitigating high-information-load problems, as connections between concepts can be established via concept labels. In addition, the associations among perceptual symbols and linguistic information, within and between concept files, makes concepts parts of an organized infrastructure of conceptual content. Reasoning and making inferences are thereby simplified in many ways. We often do not have to spend very much cognitive resources integrating a lot of information due to the informational richness of our concepts, for either sufficient information has already been integrated for us by the simulators for our concepts, or our concepts are inherently organized in such ways that facilitate the retrieval of information. The relations between conceptual content allow easy access to associated concepts and representations. Consequently, however, the more impoverished one's concepts are (with respect to conceptual content or organization), the harder it will be to reason in certain ways with them. I will return to this point below. The point to note for now is that the organization of information within and between concepts eases the burdens associated with information processing and information search, and this in turn facilitates swift and frugal heuristic reasoning.

Thus, in a word, it is our concepts—the information they embody and extant relations within and between them—to which heuristics owe their ability to satisfice, and to require little cognitive resources for their recruitment and deployment. Concepts enable heuristics to operate quickly and efficiently; concepts enable heuristics to be robust and potent inference strategies.

This is in broad strokes how I envision heuristics to work. Some of the particulars still need to be filled in, however. In the next section I will explain in a little more detail my account of how heuristics work.

4.4 How heuristics work: Informational relations and higher-order operations[36]

I assume that the nature and content of the cognitive task initially activate and prime[37] a focused set of conceptualized concepts (i.e., specific conceptualizations of concepts, or what I am calling here "conceptual content"). Suppose, for instance, one is presented with the following task:

> Estimate whether it is likely or unlikely that there will be a plane crash within the month.

This would activate conceptual content concerning *planes*, *crashes*, *likelihoods*, *estimation*, *months*, *timeframes*, and perhaps more. Since an individual likely possesses a vast number of concepts concerning such content, the conceptual content that initially gets activated can be overwhelming. There are parameters, however, that will provide constraints on what conceptual content gets activated. More specifically, there will be two kinds of parameters: parameters with respect to the nature and content of the cognitive task (what I will call *external parameters*), and parameters with respect to the structure of one's conceptual cognitive wherewithal (what I will call *internal parameters*). In the example we are considering here, the external parameters include those that are given by the language used (e.g., using "likely" and "unlikely" rather than "probability"; using "plane" rather than "jet"; using "month" rather than "30 days"), as well as those that are suggested by the nature of the task (e.g., the task elicits a course-grained subjective likelihood assignment to a future event as opposed to, say, a fine-grained numerical subjective probability). The internal parameters will have to do with factors

[36]The following account was inspired through conversations with Chris Viger. However, I recently came across a paper by Andy Clark (2002) which advocates for a method to circumvent Fodor's frame problem by borrowing from the work of Jon Kleinberg (1999) on web search engine techniques. Though the central ideas of this subsection were arrived at independently of Clark (or Kleinberg), I roughly follow and adapt parts of Clark's account, not only because it corroborates my thesis, but because it fills in some of the details that I had not had a chance to fill in myself. Nevertheless, two important differences between my view and Clark's account are that Clark does not offer the kind of cognitive architecture I give in this dissertation, and Clark is not concerned with providing an account of how heuristics work in cognition.

[37]For simplicity, I will use "active"/"activated" to refer to both active/activated and primed content.

affecting long-term memory recall, such as which concepts an individual possesses, the relative strengths at which conceptual content is stored in long-term memory, the ease with which conceptual content is activated (perhaps based on past activations), and the existence and relative strength of extant connections between concepts and conceptual content established by past inferences. Both external and internal parameters will substantially constrain what conceptual content gets initially activated upon engaging a given cognitive task, and moreover, limits on time and cognitive resources will restrict what and how many conceptualizations occur. Yet, even with these constraints, a considerable amount of conceptual content may still get activated. Moreover, further activations can be made once the initial set of conceptualizations occur. For, according to the theory of concepts developed above, the inherent relations between concepts will have it such that invoking some conceptualization will invoke others, depending on the nature and strength of such relations. And so the set of activated concepts and conceptual content will almost invariably be more than what the nature and content of the cognitive task initially invokes (subject to the constraints just outlined).

We might distinguish here between two types of connections between conceptual content with respect to the direction in which the connection is established and maintained. In this sense, connections can be *inward* or *outward*. Inward connections of some conceptual content CC_i are those established and maintained by invoking other conceptual content CC_j; the direction of priming or activation is from CC_j to CC_i. On the other hand, outward connections to some conceptual content CC_j are those established and maintained by invoking CC_i; the direction of priming or activation is from CC_i to CC_j. For example, tokening a conceptualization of *a plane* might token a conceptualization of *people*, but tokening a conceptualization of *people* might not token a conceptualization of *a plane*. In this case, *a plane* bears an outward connection to *people*, and on the other side of the coin, *people* bears an inward connection from *a plane*; but *a plane* does not bear an inward connection from *people*. We will get a clearer sense of how this is supposed to work below when I discuss examples and present diagrams.

I suppose that these connections are established and exist in virtue of associations among conceptual content. As claimed, a conceptualization of *a plane* can likely involve conceptualizations of its passengers, and *ipso facto* conceptualizations of *people*; but it is not as likely that a conceptualization of *people* will involve conceptualizations of *a plane*. This is because there is a much weaker association from people to planes than from planes to people. Such associations can vary in degree of strength, as discussed above with respect to connections between concepts generally. In other words, there will be varying *thresholds* of association between conceptual content. In short, then, what and the extent to which conceptual content gets activated will depend in part on the types and strengths of connections borne by the concepts and conceptualizations invoked by the nature and content of the cognitive task. Let us also remind ourselves here that the connections between concepts and conceptual content are established by associations between perceptual symbols and simulations or via natural language, as explained above.

It is important to understand that, on the present account, heuristics do not primarily attend to the content of concepts. I say "primarily" since some conceptual content may figure into some heuristic processes. But my point is that conceptual content is not what heuristics generally operate over.[38] Conceptual content comes into play with the initial setup of the cognitive task (as just described), and then with deeper conceptual processing (if it comes to that), but not with respect to heuristic processing. The relations exhibited by a conceptual system embody a rich source of implicit (putative) knowledge concerning which concepts and representations are to be considered for given cognitive tasks. Heuristics utilize this implicit (putative) knowledge, not the content of the concepts concerned, by operating over the informational relations between concepts and the relational structure of the conceptual system more generally. I will describe here two types of relational structure that can arise from the conceptual information that gets activated by the nature and content of a cognitive task, and I will then explain how heuristics can exploit such structures.

One type of potential extant relational structure (or simply ER-structure) that can arise

[38]By "operate over", I mean something to the effect of doing inferential work with computations. This sort of meaning will apply throughout the rest of this dissertation.

with respect to a given cognitive task is one in which active conceptual contents bear outward connections to the conceptual content of some single concept. The number and strength of inward connections that the latter concept bears would indicate that its conceptual content is putatively pertinent to the task at hand,[39] since multiple inward connections in a sense serve as corroboration that there is something important about the concept in question with respect to the task at hand. This is similar to how a scholarly article with multiple citations is accredited as (in some sense) more important to a given topic than articles with few citations. In this way, inward connections, like citations, carry information about the pertinence or importance of a given concept.

Let us call the type of ER-structure just described a *thin* ER-structure, since the property in question—namely, the number of strong inward connections to a given concept—is thin insofar as a concept can bear many strong inward connections while the concepts on the giving-end of such connections (i.e., the concepts bearing outward connections to the first) may bear no further outward connections to other potentially pertinent concepts. In other words, considering only the multiple strong inward connections to a single concept may very well miss other concepts that may be pertinent to the task at hand.

Richer informational relations between active conceptual content may give rise to a *thick* ER-structure. Richer informational relations can arise from making available (i.e., priming or activating) more conceptual content, which results in an informational structure that embodies richer connections, or more particularly, a richer informational ER-structure. Thick ER-structures exhibit not only one or a few concepts that bear many strong inward connections, but many concepts that bear strong inward connections from *sets* of concepts. The assumption behind this idea is that in a sufficiently rich structure of activated conceptual content, a pertinent concept will tend to have many sources of inward connections, and these sources will

[39]I realize that by talking about "pertinent" representations I am really talking about *relevant* representations. I use "pertinent" here not to skirt the issues concerning the relevance problems discussed in chapter 3, but to temporarily hold the issues to one side as I develop the present account of how heuristics work. As promised, I will address the epistemological relevance problem in chapter 6. Furthermore, I qualify "pertinent" with "putatively" since we have not yet established how we are to determine normative relevance (pertinence). More on this in chapter 6. In the mean time, however, I will omit "putatively" for the sake of simplicity.

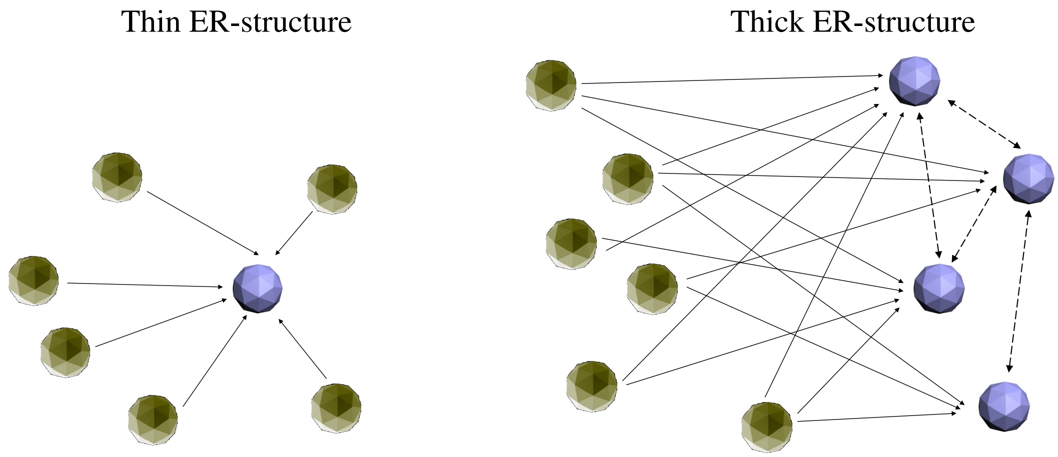


Figure 4.1: Simplified representation of *thin* and *thick* ER-structures. The nodes represent richly structured activated and primed concepts, and the arrows represent inward/outward connections. The concepts bearing the inward connections are deemed pertinent to the task at hand. With respect to the thick ER-structure represented, the concepts on the left constitute a set of reference concepts that bear multiple outward connections to the pertinent concepts on the right.

at the same time bear many outward connections to other pertinent concepts. For the sake of simplicity, let us call the said sources of inward connections "reference concepts" (RCs). It is assumed, in other words, that, since pertinent concepts (PCs) are supposed to have, in common, pertinence to the task at hand, in a rich ER-structure there should be concepts, namely RCs, that bear multiple outward connections to multiple PCs. The richer the structure, the more RCs there should be. Furthermore, we would be able to group RCs into sets according to which PCs they outwardly connect to. Each member of these sets bears multiple outward connections to multiple PCs, and thus there can be considerable overlap between sets which can also be used to pick out supersets. And, again since PCs are indeed supposed to have, in common, pertinence to the task at hand, PCs should bear inward and outward connections to and from each other as well. Returning to the article analogy, important articles will tend to have many citations from a variety of other articles, the latter of which may be grouped according to the articles (in addition to the "important article") they cite; and moreover, the important articles would tend to cite each other.

Figure 4.1 graphically illustrates, in a grossly simplified way, the thin and thick ER-structures I have described; in reality, the concepts would be more numerous and the connections messier.[40] Given these types of ER-structure, my suggestion is that a heuristic can use the information embodied by either ER-structure to pick out pertinent concepts in order to bring to bear certain representations to the task at hand, or to make predictions, or to make inferences, depending on the nature of the task and the function of the heuristic. With respect to thin ER-structures, strong connections that bear inwardly and in large numbers to given concepts would inform a heuristic that the latter concepts are pertinent to the task at hand, and that they therefore have a special role to play. There very well may be more than one concept that bears such numerous and strong inward connections, and if so then each will be considered pertinent. But, by their very nature, thin ER-structures typically do not exhibit rich structures, and so they generally should not exhibit relations that pick out more than a few pertinent concepts. Thus, when a heuristic has a thin ER-structure over which to operate, it is possible that more conceptual content will be activated in order to do more processing, if some requisite goal is not met by the thin structure alone. After more conceptual content is activated, the ER-structure might get thicker with richer connections. It is important to remember that heuristics engage in satisficing processes, and how much conceptual content is surveyed will depend on the heuristic, the nature of the task, the constraints on the agent, and the aspiration level to be met. And so, again, conceptual information will continue to get activated until the goal is satificed (i.e., the set level of aspiration is met). In terms of the present discussion, this means that a thin ER-structure may get thicker as a heuristic operates over it. Yet this will not necessarily be the case. A solution may be found in a thin ER-structure, or a heuristic might simply be satisfied with whatever answer is produced given a thin ER-structure, in which case no more content will get activated, and the heuristic will thus deliver the solution based on the thin structure.

Heuristics will often have thick ER-structures to operate over, however. This is because, as argued in the previous chapter, the cognitive tasks we face usually have a high cognitive

[40]The figure is adapted from Kleinberg (1999).

load, and as such, navigating these tasks invokes large amounts of informationally rich representations and conceptualizations. Heuristics that exploit the implicit knowledge embodied by thick ER-structures thereby have a lot more information over which to operate; heuristics that operate over thick ER-structures will be informed of which concepts are to be brought to bear or have a special role to play in inference or prediction.

To illustrate, let us briefly consider the task presented above, viz. estimating whether it is likely or unlikely that there will be a plane crash within the month. If, for example, the connectional structure among conceptual content invoked by this task is such that there are a number of inward connections to the conceptualizations concerning plane crashes, then this will indicate that plane crashes are in some sense pertinent, and it will therefore have a significant role to play in the inference. A heuristic may thus bring to bear such conceptualizations of plane crashes on the task at hand, and this might suggest to the individual that it is likely that a plane crash will occur within the month. On the other hand, if one possesses more conceptual information regarding plane crashes and their frequency, then the ER-structure invoked may be thick. Some of the representations and conceptualizations may bear multiple outward connections to such concepts as PLANES, CRASHES, FREQUENCY, and perhaps others. Depending on the conceptual content of these latter concepts, heuristics operating over such a thick ER-structure may still suggest that it is likely that a plane crash will occur within the month. If one conceptualizes FREQUENCY in the appropriate ways, however, then one would be able to infer that the frequency of plane crashes is low. A heuristic would therefore deliver a prediction that it is unlikely that a plane crash will occur within the month. This is just a very brief illustration. More detail on how this works with respect to concrete tasks and heuristics will be given in the next chapter.

We might note, however, that one need not accept what a given heuristic offers. That is, one need not make a decision based on whatever inferences or judgments are delivered by a given heuristic. However, what a heuristic does offer is made available for the taking in a quick and easy way. This is most likely why we are all so tempted to base our decisions on whatever our heuristics deliver without thinking the matter through. We might say that we are cognitively lazy. Furthermore, by the same token, one's judgments can be mitigated by further knowledge or cogitation. Much like how one can train oneself to think more thoroughly through logical problems to ensure that the correct logical rules are applied, one might train oneself to remember that recalled instances of plane crashes do not indicate whether an actual plane crash is likely. But remember, according to the foregoing account, the representations and conceptualizations activated by a context and cognitive task, and the associations among them, serve as a network of information that guides heuristic processes. Heuristic inference is also thereby constrained because heuristics are satisficing procedures that operate on the information made available throughout the process—deep conceptual processing would be nonheuristic. Thus, if the conceptual content of one's active concepts exhibit a certain ER-structure, a given heuristic procedure will facilitate certain inferences in the absence of further conceptual cogitation.

Let us observe that the picture I am presenting here of heuristics operating over ER-structures is consistent with the characterization of cognitive heuristics offered in chapter 2. I should make clear, however, that thick and thin ER-structures are not meant to be understood as determining what heuristic is to be employed. An ER-structure will determine what information a given heuristic can operate over, but it is a different matter as to what a heuristic *does with* the implicit knowledge embodied by the ER-structure. The latter will be determined by the nature of the task and the function of the heuristic. Further, I do not intend to suggest that the thick and thin properties I have described are in any way exhaustive of the structural properties that heuristics exploit; my contention is that heuristics operate over structural relations between concepts and their conceptual content, and that there are cases in which such structural relations are thick as well as cases in which they are thin, as I have described these properties. The important point to stress is that heuristics are thereby understood to operate by generally ignoring almost all of the actual content of the active concepts and their conceptual content. Rather, the structures over which heuristics operate are the relational structures—structures that implicitly embody *higher-order* information (or knowledge) about active concepts and their content. In this way, active conceptual contents work together as collections that point to a very small set of pertinent or salient concepts (in some cases the set may consist of only one concept as a member). Such pertinent or salient concepts will be those that bear numerous strong inward connections from other active content. I argued in the previous chapter that heuristics exploit the informational richness of concepts, but remain computationally cheap. It should be all the more apparent now how this is so: Heuristics do not compute over conceptual content, for that would be computationally taxing. Instead, the informational richness exploited by heuristics is implicit in the ER-structures—the informational richness is the higher-order, metainformation *about* conceptual content—which carry much of the computational load.[41]

At the same time, we should note that there will be instances when the relational structure embodied by the conceptual content of concepts may not (or maybe cannot) guide heuristic inference. These will be instances when the available ER-structures are very thick—when, for instance, too much conceptual information is activated. Such very thick ER-structures would not be conducive to heuristic processes since the extant informational relations would not exhibit anything that a heuristic can exploit.[42] Thus, some nonheuristic process may need to be employed to aid in reasoning through these sorts of structures. On the other hand, we may very well need a heuristic to navigate such a very thick structure, since sorting through all the conceptual information may be computationally taxing.

Another case when an ER-structure will not be sufficient to guide heuristic inference is when one's relevant concepts are too impoverished, or maybe even do not exist at all. In the previous chapter, I had claimed that impoverished k-systems can negatively affect the way inferences are made. The same reasoning applies here. When one's relevant concepts are too impoverished, heuristics can still be deployed, but they might operate according to the

[41] This idea is in some ways similar to Fodor's (1975) thoughts about what heuristics do with respect to assigning meaning to sentences. Rather than computing the grammatical relations exhibited by the sentence, Fodor believed that heuristics can infer meaning from background knowledge considerations of the lexical content of the sentence. See the brief discussion in chapter 2 above (section 2.2.1).

[42] This is along the same lines as what Goldstein and Gigerenzer (2002) have in mind when they discuss the "less-is-more effect" with respect to the Recognition heuristic: "The recognition heuristic leads to a paradoxical situation in which those who know more exhibit lower inferential accuracy than those who know less" (p. 79). See section 5.2.1 in the following chapter for further discussion.

informational structure of the context and cognitive task as opposed to the conceptual content of and relations between the concepts involved. On the other hand, cognition might turn to default heuristics in such cases, or maybe no heuristics at all. Thus, we might say that an ER-structure must be *thick enough* for a heuristic to operate over it to yield a certain solution (though, as we saw, it cannot be too thick).

Instances in which one's relevant concepts are impoverished may also lead to the "mistakes of reason" emphasized and researched by the heuristics and biases tradition. Alternatively, one's concepts may be constructed such that the conceptual information that tends to get initially activated when confronted with a certain type of problem may embody certain information or bear certain informational relations such that mistakes and biases tend to ensue. This may not mean that one's concepts are impoverished, or that erroneous inferences result from erroneous beliefs. Rather, one may simply need to conceptualize the relevant concepts in different ways, which would in turn activate different conceptual content, and this might facilitate correct solutions. I will discuss some of these issues in more detail in the next chapter.

But heuristics do more than just operate over specific informational structures. Though this is an important part of how heuristics work, what heuristics do with the information they operate over is key to their function and usefulness. Heuristics are typically employed to aid in prediction, decision, or problem-solving (or simply "inference"). There is, of course, a computational cost in making inferences. We saw in chapter 2 that there may be GCO or optimization procedures that would deliver guaranteed or correct inferences, but only if the problem in question is well-defined. The other class of problems—the ill-defined problems—do not have such GCO or optimization procedures, and so making an inference in such contexts can be computationally expensive. For not only will there be too many factors to consider, the problem-spaces for these sorts of problems cannot be adequately known, understood, or characterized. (And even if the problem is well-defined, we saw that there are some cases where, although there is a GCO or optimization procedure in principle, it is computationally expensive to discover or execute in practice.) As indicated, heuristics offer a way to manage and navigate these problems by effectively ignoring most of the conceptual content of the relevant representations,[43] and instead looking to the extant relations among activated concepts. These extant relations will indicate what conceptual content to consider. But remember, heuristics also mitigate the computational costs of inference by satisficing—seeking a "good enough" solution (i.e., a solution that meets some level of aspiration). Therefore, how much of the informational relations between activated concepts will be considered, how much conceptual content will be considered, and what concepts actually get activated throughout the satisficing process, will depend on the nature of the task and constraints on the agent, but most especially on the set aspiration level.

A heuristic might thus be understood as partly operating as a metacognitive judgment on how much effort will be invested in solving a problem (cf. Sperber, 2005). In this light, a heuristic evaluates and estimates likely dividends by processing the informational relations exhibited by active conceptual content, rather than considering conceptual content; if it does not meet the aspiration level further activations are made along with further evaluations. Thus, sometimes a satisficing answer is arrived at with the initial activated ER-structure, but other times more conceptual information will need to get activated to satisfice (meet the aspiration level). But a heuristic will deliver a judgment on whether to continue computing, or to capitulate and let other (perhaps deeper conceptual) processes take over. If the aspiration level is met, however, the heuristic is able to deliver an inference based on the given conceptual informational structure. More specifically, what will be brought to bear on the task will be the conceptual content exhibited by the concepts that bear a sufficient number of strong inward connections, and this is because it was determined that basing the inference on such information is good enough for the task. As we look at particular examples of heuristics in the next chapter, we will get a clearer idea of all of this.

[43] More on what makes representations relevant in chapter 6.

4.5 Concluding remarks

Let us briefly review. Following Barsalou's theory of concepts, concepts are conceived to be simulators that embody perceptual symbols, where these symbols are understood to be collections of neurons that are activated in the perceptual centres of the brain upon initial perceptual processing of entities in the concept's extension. I argued, however, that Barsalou's theory benefits from being augmented by including natural language. Specifically, I argued that concepts (simulators) should be conceived to also include linguistic symbols (collections of active neurons in the language centres of the brain), and moreover, following the file metaphor of cognition, that concepts have specific flagged linguistic symbols that serve as a name or label that expresses or refers to them. Natural language can thus operate over these names or labels without necessarily invoking conceptual content. Drawing from the work of Spelke and Viger, the productive and combinatorial properties of natural language enable us to establish connections between concepts which may not have been otherwise possible.

Given this initial setup of the architecture of concepts and cognition, I suggested that heuristics are cognitive procedures that operate by exploiting the informational structures exhibited by concepts and their conceptual content (conceptualizations and representations). Such informational structures also act as constraints that enable heuristics to satisfice, and require little cognitive resources for their recruitment and execution. The account of how heuristics work developed here involved describing thin extant relational structures (or ER-structures) that exhibit relatively simple relational properties among primed and active concepts, as well as thick ER-structures that exhibit richer relational properties among primed and active concepts. As recently stressed, the important point is that heuristic procedures do not operate over most of the conceptual content, but over the higher-order information (or knowledge) about the concepts in question implicit in the ER-structure, which makes some content salient. This allows heuristics to mitigate computational costs of inference, but heuristics also produce metacognitive judgments on how much effort will be invested in solving a problem. Hence, as argued in previous chapters, heuristics exploit the informational richness of cognitive structures, and moreover, heuristics are good candidates to mitigate the computational burdens of cognition, although it is really our conceptual wherewithal that shoulders such burdens.

In the next chapter I will review some of the empirical data on heuristic reasoning to illustrate the foregoing view of heuristics, concepts, and cognition. Specifically, I will illustrate that some of the heuristics proposed and studied by Kahneman and Tversky, and Gigerenzer, can be understood in terms of heuristic operations over thick or thin ER-structures.

Chapter 5

How Heuristics Work: Considering Empirical Evidence

Having explained the nature of the informational structures that heuristics exploit, I can now corroborate my thesis by illustrating with concrete examples how heuristic reasoning is facilitated. The easiest way to do this is by considering empirical evidence with respect to individual heuristics and showing how they can be understood as processes that exploit the informational relational structure exhibited by active concepts and content. Thus, by working through the present chapter we will get a better sense of how my thesis explains a some of the empirical data on heuristic reasoning.

In this chapter I will discuss two of Kahneman and Tversky's most developed heuristics—Availability and Representativeness—and two of Gigerenzer's most developed heuristics—the Recognition heuristic and Take the Best—and show that they can be understood as procedures that operate over such ER-structures. I will also illustrate some other aspects of the general theory of concepts and cognition developed in this dissertation. Specifically, I will use Kahneman and Tversky's cab problem (to be described below) to illustrate that one's concepts and conceptualizations play a crucial role in how problems are conceived and subsequently solved.

For the most part I have tried to be brief in my discussions. I should point out that the analyses I give are by no means complete, and more work is necessary to fill in many of the details. Furthermore, I pass over much of the empirical research regarding the heuristics I analyze. However, I present enough to show that my view can account for several phenomena, and is therefore a plausible theory of how heuristics work in cognition.

5.1 Heuristics and biases

In this section I will discuss two of the more popular heuristics researched by the heuristics and biases program—Representativeness and Availability. Although a lot of research has been conducted with respect to these heuristics since Kahneman and Tversky first began theorizing about them in the early 1970s, I will mainly be concerned with Kahneman and Tversky's original work and their original experiments and examples. This evidence remains valid, and it provides the paradigmatic cases that are widely cited and discussed throughout the literature.[1] Before I begin, however, it would be instructive to make some brief observations about Kahneman and Tversky's general view of human cognition.

Kahneman and Tversky had at times analyzed people's judgments under uncertainty within the framework of *mental models* (e.g., Johnson-Liard, 1983). Although there are different accounts of mental models and their role in reasoning, inference, and judgment, Kahneman and Tversky (e.g., 1983) appear to subscribe to the idea that mental models are mentally represented *prototypes* or *schemata* that can be mentally manipulated according to one's (perceptual) knowledge (cf. Minsky, 1975). Judgments about a particular object or event are made relative to the mental model one constructs of the class of objects or events in question. Kahneman and Tversky's purpose was not to advance any theories regarding mental models (or any alternative hypothesis), so they never discuss such a view of cognition in any detail. Nevertheless, it is clear that they believed that mental model theory underlies the operations of the Availability heuristic and the Representativeness heuristic.

There is a similarity between mental models and Barsalou's theory of concepts, and Barsa-

[1]Kahneman has recently reviewed and offered new analyses of his and Tversky's research (Kahneman, 2003; Kahneman & Frederick, 2002). According to Kahneman's new analysis of heuristics and biases, a common process of *attribute substitution* explains how judgment heuristics work. Attribute substitution is a generic heuristic process in which an "individual assesses a specified *target attribute* of a judgment object by substituting a related *heuristic attribute* that comes more readily to mind" (Kahneman, 2003, p. 707). The idea is that a substituted heuristic attribute will generally be more accessible (i.e., comes more easily to mind) than a target attribute, and thus the heuristic attribute (rather than the target attribute) will be used to make a judgment. "For example, the representativeness heuristic is the use of representativeness as a heuristic attribute to judge probability" (ibid.). As we shall see presently, the account that I offer is consistent with Kahneman's updated model of how heuristics work, but we shall also see that my account provides more detail with respect to the cognitive structures that guide heuristic inference and judgment. However, we should note that Kahneman admits two things. First, he admits that his new model of heuristics by attribute substitution does not perfectly coincide with the original conception of heuristics offered by Tversky and Kahneman (1974). Second, he admits that his new model excludes anchoring effects, i.e., effects "in which judgment is influenced by temporarily raising the accessibility of a particular value of the target attribute, relative to other values of the same attribute" (Kahneman, 2003, p. 707), which is an instance of Anchoring and Adjustment (see (13) in chapter 2, p. 42). The reason why Kahneman's new model cannot explain anchoring effects is because there is no attribute substitution in anchoring, but rather an alteration of perceived salience of the target attribute. Nevertheless, I believe that the account of how heuristics work that I have developed in this dissertation enjoys an advantage of covering anchoring effects, although I do not show this below. In fact, I believe that my account covers heuristic processes in general, but I lack the space here to adequately show this. Here in brief is how I envision my theory accounts for anchoring effects: The concepts and ER-structure that get activated upon being confronted with a given cognitive task can act as an anchor, as Tversky and Kahneman describe, by being more accessible relative to other ER-structures (that may subsequently get realized), and by pointing to certain conceptual content that is consequently perceived to be salient. Hence, activated ER-structures can influence inferences and judgments in the way that anchors do, and adjustments can be made to this anchor as one deliberates through the problem.

lou recognized this. As he remarks, "mental models are roughly equivalent to ... simulations of specific entities and events" (Barsalou, 1999, p. 586). He continues, however, by noticing a difference between his notion of simulators and the idea of mental models: "Mental models tend not to address underlying generative mechanisms that produce a family of related simulations" (ibid.).

Given that the account of concepts developed in this dissertation is based on Barsalou's theory of concepts (or simulators), there is likewise a relation between the foregoing account and mental models. There are important differences, however, including not least that linguistically coded information is available for simulation or manipulation in addition to perceptual symbols. According to my account, a mental model might be conceived to consist of the active conceptual content of a concept—i.e., active conceptual representations and conceptualizations. Nevertheless, simulations constructed from the active conceptual content of a concept are unlike mental models insofar as mental models are usually understood to be constructions of representations of objects or events generally, whereas my account applies more narrowly to concepts and the information stored in their files. Another difference that will be exhibited in the present chapter is that heuristics are supposed to operate over ER-structures according to my account, whereas no such mechanisms or structures are indicated in the mental models theory of cognition (as Barsalou observes).

However, let us not get caught up in the details of the similarities and differences between mental models and the foregoing theory of concepts. Suffice it to say that my theory provides a framework that is compatible with mental models and *ipso facto* with Kahneman and Tversky's view of cognition.

It is also interesting to note that the present account is consonant with and adds detail to Tversky and Kahneman's (1983) notion of "natural assessments". Recall from chapter 2 (section 2.3.3) that Tversky and Kahneman envisioned that heuristics "rely on" natural assessments, that natural assessments "play a dominant role" in producing judgments and inferences, although they failed to specify precisely what these assessments are and in what sense they are "natural". Tversky and Kahneman did assert, however, that "natural assessments include computations of similarity and representativeness, attributions of causality, and evaluations of the availability of associations and exemplars" (p. 294). With the present account, we can understand natural assessments as the relations between active concepts and conceptual content realized when presented with a problem. Thus, computations of similarity and representativeness, attributions of causality, and evaluations of the availability of associations and exemplars, are all present upon the initial realization of the ER-structure invoked by the problem in question. And what makes these assessments "natural" is that they are entailed by the inherent structures that exist within and between concepts and their conceptual content. Understood as such, heuristics do not so much "rely on" natural assessments as they are, again, constrained and guided by natural assessments.

5.1.1 Availability

Tversky and Kahneman write,

> Life-long experience has taught us that instances of large classes are recalled better and faster than instances of less frequent classes, that likely occurrences are easier to imagine than unlikely ones, and that associative connections are strengthened when two events frequently co-occur. Thus, a person could estimate the numerosity of a class, the likelihood of an event, or the frequency of co-occurrences by assessing the ease with which the relevant mental operation of retrieval, construction, or association can be carried out. (Tversky & Kahneman, 1973/1982a, pp. 163-164)

These observations are what support the Availability heuristic:

(12) The frequency of a class or the probability of an event is assessed according to the ease with which instances or associations can be brought to mind.

According to Tversky and Kahneman, an assessment of availability, or the ease with which instances or associations can be brought to mind, mediates probability judgments. For example, one may predict the rate of plane crashes by recalling recent plane crashes; or one may estimate the probability that a struggling student will fail an exam by assessing the strength of association between students struggling in a class and failing exams. However, Tversky and Kahneman assert that it is not necessary that one actually mentally perform retrievals or constructions to assess availability. Rather, "It suffices to assess the ease with which these operations could be performed, much as the difficulty of a puzzle or mathematical problem can be assessed without considering specific solutions" (ibid.).[2] We can see here an indication of the metacognitive judgments that heuristics offer on how much effort will be invested in solving a problem.

The assumption is that frequent events are in general easier to recall or imagine than infrequent events, and so availability assessment will tend to covary with frequency or higher probability. Yet, as Tversky and Kahneman emphasize, availability is influenced by various factors other than actual frequency, and therefore the use of the Availability heuristic can lead to systematic and predictable biases. For example, Tversky and Kahneman posed the following problem to research participants:

> Suppose one samples a word (of three letters or more) at random from an English text. Is it more likely that the word starts with *r* or that *r* is the third letter?

They found that people tend to approach this task by trying to recall words that begin with *r* (e.g., rational) and words that have *r* appear as the third letter (e.g., cart). Since it is easier to recall words that begin with *r* than it is to recall words that have *r* as the third letter, people judge that the former case is more likely than the latter, even though there are in fact more words in the English language that have *r* as the third letter than there are words that begin with *r*. This is a simple example involving little more than memory recall. However, the Availability heuristic is supposed to influence more complex cognitive operations.

[2]It is possible that what is doing the work is not *the ease with which instances or associations can be brought to mind*, but *the number of instances that has been brought to mind*. The difference here is between assessments of relative ease in performing cognitive operations such as retrieval or construction, and actually carrying out retrievals or constructions and thereafter surveying what has been retrieved or constructed. In Schwarz and Vaughn's (2002) terms, availability can consist in *ease of recall* or *content of recall*, each of which represents distinct sources of information. Schwarz and Vaughn show, however, that experimental evidence supports Kahneman and Tversky's original formulation of the Availability heuristic, viz. the frequency of a class or the probability of an event is assessed according to the ease with which instances or associations can be brought to mind.

According to the account proposed in this dissertation, availability, or the ease with which instances or associations can be brought to mind, depends upon and is facilitated by the conceptual content of one's concepts and the existence and relative strengths of the extant relations between them. That is, the strength of association upon which availability relies is borne out by the network of activated representations and conceptualizations of the concepts one is currently entertaining.

Suppose one was given the following task.

> Estimate whether it is likely or unlikely that a person with mental illness will exhibit violent behaviour.

People with mental illness are generally not violent (especially considering the various kinds of mental illness). But one might assign a high likelihood if one's concept-file for MENTAL ILLNESS contains beliefs about mentally ill people being violent—maybe from reading certain stories in newspapers; or perhaps one had a personal experience with a violent person who happened to be mentally ill, and this experience is vividly simulated upon MENTAL ILLNESS being activated. Alternatively, one's MENTAL ILLNESS concept might exhibit certain strong associations with other concepts and conceptual content that imply violence. For example, one's MENTAL ILLNESS concept might be strongly associated with PSYCHOPATH; and since psychopaths are often violent, and one's concept-file for PSYCHOPATH may contain information having to do with psychopaths being violent, one might estimate that a person with mental illness will very likely exhibit violent behaviour.

The use of the Availability heuristic in making judgments regarding, for example, the likelihood that a person with mental illness will exhibit violent behaviour can be understood in terms of the heuristic operating over a thin ER-structure, as described toward the end of the previous chapter. I here mentioned two cases that can lead one, via the Availability heuristic, to estimate that it is likely that a person with mental illness will exhibit violent behaviour—namely, (a) if one's concept-file for MENTAL ILLNESS contains beliefs about mentally ill people being violent, or (b) if one's MENTAL ILLNESS concept exhibits certain strong associations with other concepts and conceptual content that imply violence. If invoking *mental illness* simulations (or opening one's MENTAL ILLNESS file) results in priming or activating easily recalled conceptual content, established and built up from news stories or personal experiences regarding violent mentally ill people, such conceptual content *ex hypothesi* outwardly connects to one's VIOLENT and MENTALLY ILL PEOPLE concepts; by the same token, VIOLENT and MENTALLY ILL PEOPLE bear multiple inward connections. Assuming that these inward connections are relatively numerous and strong, the ER-structure would indicate that these concepts are pertinent, and a heuristic operating over such a structure would estimate a relative likelihood according to the extant connections borne by VIOLENT and MENTALLY ILL PEOPLE. If, on the other hand, one's MENTAL ILLNESS concept exhibits associations with other concepts and conceptual content that imply violence, then VIOLENT alone may bear the requisite inward connections over which a heuristic would operate, and in a similar fashion, the heuristic would estimate a likelihood according to the extant connections. Thus, in either case, (a) or (b), one tends to judge that it is likely that a person with mental illness will exhibit violent behaviour based on an estimation delivered by a heuristic which operated over thin relational connections exhibited by the given ER-structure.

I consider the Availability heuristic as operating over thin ER-structures (rather than thick ER-structures) since it seems that what is "available"—what instances or associations are brought to mind—bears sufficient inward connections to facilitate the heuristic.[3] In addition, the Availability heuristic does not seem to operate over thick connectional properties. With respect to the present example, for instance, there does not seem to be sets of concepts invoked, the members of which bear multiple outward connections to multiple other concepts. This, of course, is a claim that can stand or fall according to empirical evidence. But even if it is the case that the Availability heuristic in fact operates over thick ER-structures, my general point still applies, which is that the heuristic operates over the implicit information embodied by the

[3]Recall the discussion at the end of chapter 4 where I explained that an aspiration level might not be met by a given thin ER-structure, and more conceptual information would thereby get activated, increasing the thickness of the ER-structure. I had also claimed, however, that this need not be the case—that no more conceptual information would be activated—if aspiration level is met by the given thin ER-structure.

relational structures existing between primed and activated concepts and conceptual content. (And this same point applies to the rest of the claims that I make in this chapter regarding the other heuristics I discuss.)

It is instructive to note that the account I am offering here is consistent with what Tversky and Kahneman envisioned for the Availability heuristic, viz. "It suffices to assess the ease with which these operations could be performed, much as the difficulty of a puzzle or mathematical problem can be assessed without considering specific solutions" (Tversky & Kahneman, 1973/1982a, pp. 163-164). Tversky and Kahneman appear here to be intimating that the Availability heuristic does not consider conceptual content, since their point is that the heuristic does not actually perform constructions to assess availability. This is along the same lines as what I had stressed in the previous chapter, namely that heuristics do not operate over the content of concepts or conceptualizations, since this would involve computations in which heuristics are not in the business to engage. Again, on my account heuristics operate over the higher-order, metainformational relations between active content. This may very well be the sort of idea that Tversky and Kahneman had in mind for the Availability heuristic, although they never attempted to describe the mechanics of the heuristics they studied.

5.1.2 Representativeness

Let now turn to the Representativeness heuristic:

> (11) Probabilities are evaluated by the degree to which one thing or event is representative of (resembles) another; the higher the representativeness (resemblance) the higher the probability estimation.

The probabilities referred to are subjective probabilities one assigns to the event(s) concerning a given inferential task.

Kahneman and Tversky never fully characterize "representativeness" (or "resemblance"), claiming that it is like "similarity", which is easier to assess than to characterize (Kahneman & Tversky, 1972). Nevertheless, they offer a partial characterization, and rely on Tversky's (1977) account of the psychological principles of judgments of similarity. According to Tversky and

Kahneman (1983), "Representativeness is an assessment of the degree of correspondence between a sample and a population, an instance and a category, an act and an actor or, more generally, between an outcome and a model" (p. 295). More formally, one judges an instance or a subset *X* of a class *C* as representative of *C* to the extent that *X* possesses features that are believed to be essential to and shared by the members of *C*, and does not possess many distinctive features that are not essential to or shared by members of *C* (Tversky & Kahneman, 1982b, p. 86).

The cases in which Tversky and Kahneman are most interested are those where representativeness affects people's judgments of probability. Their interest in such cases is motivated by the fact that probability theory presents a set of normative rules against which to evaluate personal judgments, and hence, in certain circumstances there is a sharp "contrast between the extensional logic of probability theory and the psychological principles of representativeness" (Tversky & Kahneman, 1983, p. 296).[4] As Tversky and Kahneman go on to explain, representativeness tends to covary with frequency, since frequent instances or events are generally more representative of a class than unusual instances and rare events, but there are circumstances in which this is not the case. This is because judgments of representativeness are not determined by frequency (ibid.). It is generally in such circumstances that errors and biases occur, for this tendency to covary often leads people to overestimate the actual correlation between frequency and representativeness.

Tversky and Kahneman believe that people tend to rely on representativeness, despite the fact that it can lead to systematic biases and judgment errors (with respect to probability), because representativeness is easy to evaluate and readily accessible. As they explain, "research on categorization . . . suggests that conceptual knowledge is often organized and processed in terms of prototypes or representative examples. Consequently, we find it easier to evaluate the representativeness of an instance to a class than to assess its conditional probability" (p. 89).

[4]There is an ongoing debate regarding the normativity of the probability calculus and heuristic reasoning (e.g., Kahneman & Tversky, 1996; Gigerenzer, 1996). However, I do not want to enter into or address this debate here. It will suffice for my purposes here to simply acknowledge that there is sometimes a disconnect between probability theory and heuristic reasoning.

This corroborates the account of how heuristics work and the role of concepts I have been developing throughout this dissertation. I will illustrate here that the Representativeness heuristic can be understood to be a cognitive response to the relational structure exhibited by active and primed conceptual content of one's concepts. To see this more clearly, let us consider some experiments conducted by Kahneman and Tversky which illustrate the Representativeness heuristic.

5.1.2.1 The Linda problem

Kahneman and Tversky presented a number of groups of undergraduate and graduate students with the following problem:

> Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.
>
> Please rank the following statements by their probability, using 1 for the most probable and 8 for the least probable.
>
> (a) Linda is a teacher in elementary school.
> (b) Linda works in a bookstore and takes Yoga classes.
> (c) Linda is active in the feminist movement.
> (d) Linda is a psychiatric social worker.
> (e) Linda is a member of the League of Women Voters.
> (f) Linda is a bank teller.
> (g) Linda is an insurance salesperson.
> (h) Linda is a bank teller and is active in the feminist movement.

As Tversky and Kahneman report, participants consistently ranked (h) higher than (f)—i.e., participants believed that it is more probable that Linda is a bank teller and is active in the feminist movement than that Linda is just a bank teller. According to Tversky and Kahneman, participants relied on the Representativeness heuristic and judged the degree to which Linda resembles the typical member of the class picked out by (h) to be greater than the degree to which Linda resembles the typical member of the class picked out by (f). But such a judgment of *resemblance* "is neither surprising nor objectionable ... If, like similarity and prototypicality, representativeness depends on both common and distinctive features (Tversky, 1977), it should be enhanced by the addition of shared features. … [T]he addition of feminism to the profession of bank teller improves the match of Linda's current activities to her personality" (Tversky & Kahneman, 1983, p. 297). Nevertheless, Tversky and Kahneman found "more surprising and less acceptable" that participants generally judged that (h) is *more probable* than (f). The reason why this is problematic is that, according to basic probability theory, a conjunction cannot be more probable than any of its conjuncts alone. (Formally: for any two events A and B, $\Pr(A\&B) \leq \Pr(A)$ and $\Pr(A\&B) \leq \Pr(B)$.) And since (h) is a conjunction of (c) and (f), judging (h) to be more probable than (f) violates this rule. This result has since been referred to as the *conjunction fallacy*.

Stephen Jay Gould is often quoted with respect to this problem:

> I am particularly fond of [the Linda] example, because I know that the [conjunction] is least probable, yet a little homunculus in my head continues to jump up and down, shouting at me—"but she can't just be a bank teller: read the description." … Why do we consistently make this simple logical error? Tversky and Kahneman argue, correctly I think, that our minds are not built (for whatever reason) to work by the rules of probability. (Gould, 1991, p. 469)

The claims that I will offer are (i) that the homunculi in our heads shout as they do because they are guided by ER-structures (which is just another way of saying that the Representativeness heuristic operates over ER-structures); and (ii) that it may not be so much that our minds are not built to work by the rules of probability, but that people tend not to activate the appropriate conceptual content with respect to their concepts of logic and probability, and thus they rely on a heuristic rather than logical or probabilistic reasoning.

The description given of Linda provides various conceptual information. If participants were asked to predict Linda's occupation or her current activities based on the description alone, without any options, any number of responses is possible, including but not limited to (a)-(h). However, the statements given by (a)-(h) focus the cognitive task and provide content for it. Each statement provides a different background against which to analyze and conceptualize the description of Linda, and the statements therefore act as constraints on the initial setup of the task (cf. section 4.3 of the previous chapter). The statement "Linda is a teacher in elementary school" activates the concept ELEMENTARY SCHOOL TEACHER with various conceptual content regarding things such as elementary school teachers' general demeanor, level of intelligence, typical major in university, and so on, based on one's past experiences and knowledge. The description of Linda also activates various conceptual content, such as representations of *outspoken* and *philosophy major*, and judgments of the extent to which Linda's description matches a given statement are made based on the associations between the respective conceptual representations and conceptualizations. We can thus see how internal and external parameters effect the initial setup of the ER-structure. We can also see how the Representativeness heuristic is thereby constrained by active conceptual information as one makes a judgment about the relative probability that Linda is in fact an elementary school teacher, given the description provided.

Similarly, the statement "Linda is active in the feminist movement" activates the concept FEMINIST with various conceptual content. With respect to this statement, however, it appears that in the vast majority of cases one's representations and conceptualizations of FEMINIST are highly associated with the description of Linda—that is, the description of Linda is representative of FEMINIST. This is evinced by the fact that the majority of participants surveyed by Kahneman and Tversky most often ranked (c) higher than all other statements (Tversky & Kahneman, 1982b, p. 92). Hence, people tend to believe that it is more probable that Linda is active in the feminist movement than any other description provided.

On the theory I am advocating here, the concepts and representations activated by the presentation of the Linda problem will be many. This is because the problem provides lots of information, including that given by Linda's description plus the eight statements that are to be ranked. This will, in turn, give rise to a complex ER-structure. The description of Linda will activate conceptual information that forms an initial set of representations. As one conceptualizes each of the statements given, from (a) to (h), additional sets of representations and conceptualizations will be activated. The Representativeness heuristic would then operate over the number, type, and strength of connections that exist between the concepts activated by the description-set and each statement-set. The connections between the concepts activated by the description-set and statement-sets will reveal an ER-structure with thick properties. Most relevant to our concerns about the Representativeness heuristic is that the concepts comprising the description-set will bear multiple outward connections to multiple concepts invoked by "Linda is active in the feminist movement", and these connections will be the most strong and most numerous among all the statement-sets. A heuristic that operates over these thick structural properties—namely, the Representativeness heuristic—would thereby deem the concepts invoked by "Linda is active in the feminist movement" as more pertinent than the other concepts constituting the other statement-sets. This would explain why people tend to rank (c) higher than all other statements. This would also explain why (h) "Linda is a bank teller and is active in the feminist movement" is consistently ranked higher than (f) "Linda is a bank teller", and indeed why one's homunculus shouts, "but she can't just be a bank teller: read the description". For, given what has just been said, the concepts comprising the description-set will bear multiple outward connections to multiple representations invoked by (h) *in virtue of its "active in the feminist movement" part*. This ER-structure will not be exhibited when the concepts invoked by (f) are set against the description-set, since we might expect that the conceptualizations of the concepts of (f) will bear less numerous and less strong inward connections from the description-set concepts, or perhaps there will be negative or inhibitory connections between the conceptual contents invoked by (f) and the description-set.

The issue that Kahneman and Tversky emphasize is the violation of probability theory that occurs when people tend to judge (h) to be more probable than (f). This phenomenon is certainly problematic, so long as probability theory is taken to be normative (but see footnote 5; also see below). But it is not entirely clear that participants are understanding *probable* or *probability* univocally, or even coherently (cf. Gigerenzer, 1996, 2007; Gigerenzer, Hertwig, van den Broek, Fasolo, & Katsikopoulos, 2005). The ongoing debate over the nature of probability shows that even the professionals are not clear on these concepts. "Probability" can mean anything from *degree of belief*, to *frequency*, to *propensity* (von Plato, 1994), and it can also be interpreted vaguely or ambiguously, as many laypeople do, in terms of *likely*, or *likelihood*, or *chance*. This is not to say that people believe probability is meaningless, and that people do not understand at all what is being asked of them when they are asked to make probabilistic predictions (*pace* Gigerenzer, 1991, 1993; cf. Samuels et al., 2002). Rather, the point is that people's ideas of what probability is, or what probabilities are, may not agree with one another, or may not be robustly or wholly understood.

The naïve and novices with respect to probability theory will likely not have a rich or robust PROBABILITY concept, and *ipso facto* such people would not activate the appropriate probability conceptualizations when confronted with a cognitive task involving probabilities. Nevertheless, the use of the Representativeness heuristic in the Linda problem is robust across people at all skill levels—even experts tend to rank (h) higher than (f) (Kahneman & Tversky, 1982). Of course, once revealed, the experts immediately understand their fault. Yet something interesting must be going on if even the experts are fooled by their homunculi. Contrary to Gould's sentiment, I am not convinced that humans are lost when it comes to reasoning with probabilities, for indeed many people know and understand very well how to carry out probabilistic calculations. My hypothesis is that people generally tend not to activate the appropriate conceptualizations with respect to their PROBABILITY concept. Instead, the thick ER-structure described above and the heuristic procedure that operates over it supersede, and in some ways repress, the activation of the appropriate conceptualizations that would allow people—especially experts—to rank the statements given in the Linda problem in accordance with the basic axioms of probability theory. And I suspect that this is a consequence of how the information is presented in the Linda problem. More specifically, the way the information is presented facilitates the establishment of the given ER-structure, but does not at all facilitate the activation of many *probability* conceptualizations. As a result, the naïve, novices, and experts alike tend to rely on a heuristic procedure—the Representativeness heuristic—that operates over the ER-structural properties established in cognition when reasoning about the Linda problem, unless one cogitates beyond this initial ER-structure.

It is also of interest to note that the Linda problem phenomenon can be couched in terms of relevance. Specifically, what seems to be going on is that the description of Linda is relevant to our judgments about her occupation, beliefs, lifestyle, etc. The very fact that participants are presented with a description of Linda indicates something about the relevance of the description, namely that it is of importance who Linda is, what Linda does, etc. (i.e., that it is of importance to the task of ranking the statements provided). In this context, ranking (a)-(h) in terms of the probability calculus simply appears to be the wrong norm, for participants may view the task not in terms of probability assignments (notwithstanding the wording of the task) but in terms of matching—matching the description to how (a)-(h) are conceptualized. Although this and related matters deserve serious attention, I refrain from discussing them any further with respect to the Linda problem in particular for a number of reasons. One reason is that introducing issues about the norms of reason will take us too far afield (though in the next chapter I discuss normative issues of a different sort). A second reason is that I return to the topic of relevance in the next chapter, and a more general discussion will be saved until then (see my comments below with respect to relevance and the Dick problem).

A third reason why I refrain from digressing into matters to do with relevance is that I have already discussed the Linda problem at length, and the point of the present chapter is not to explain the phenomenon in terms of relevance, but to show how the account of heuristics developed in this dissertation explains a range of phenomena (and as I have shown, it explains the Linda problem). To continue, then, I will show that the hypothesis I have suggested also explains further phenomena exhibited by other problems discussed by Kahneman and Tversky as they illustrate the Representativeness heuristic. I will briefly discuss two other such problems, namely *the Dick problem* and what has become known as *the cab problem*.

5.1.2.2 The Dick problem

To investigate the extent to which people's judgments were affected by prior probabilities, or base-rates (i.e., the probability of an event unconditioned on evidence), Tversky and Kahneman

(1974) presented participants with brief personality descriptions of several individuals, and they were told that the descriptions were chosen at random from a group of 100 professionals comprised of engineers and lawyers. For each description, the participants were asked to state whether it was of an engineer or of a lawyer, and to assign a probability. There were two experimental conditions, one in which participants were told that the group from which the descriptions had been selected consisted of 70 engineers and 30 lawyers, and the other in which participants were told that the group consisted of 30 engineers and 70 lawyers. Given this information, the probability that any given description is of an engineer rather than of a lawyer should be higher in the first condition where there is a majority of engineers, than in the second condition where there is a minority of engineers.

However, Tversky and Kahneman found that the same probability judgments were made in both experimental conditions, which is a violation of probability theory. As they report, "Apparently, subjects evaluated the likelihood that a particular description belonged to an engineer rather than to a lawyer, by the degree to which this description was representative of the two stereotypes, with little or no regard for the prior probabilities of the categories" (p. 1125). This fallacy has become known as *base-rate neglect*, since the participants allegedly ignored the base-rates (or prior probabilities) in the problem. On the other hand, when descriptions were omitted, and participants were asked to give a probability that a person chosen at random would be an engineer or a lawyer, they made their judgments according to the prior probabilities.

Here is a sample of one of the descriptions they gave participants:

> Dick is a 30 year old man. He is married with no children. A man of high ability and high motivation, he promises to be quite successful in his field. He is well liked by his colleagues.

As Tversky and Kahneman report, the information given by the descriptions were supposed to have presented uninformative information: "This description was intended to convey no information relevant to the question of whether Dick is an engineer or a lawyer. Consequently, the probability that Dick is an engineer should equal the proportion of engineers in the group, as if no description had been given" (p. 1125). However, participants tended to assign a probability of 0.5 to the statement that Dick is an engineer. This, then, is an example in which the Representativeness heuristic overrides consideration for base-rates.

According to my account the setup and content of the Dick problem naturally invokes a certain ER-structure, and a heuristic procedure operates over this structure, assessing its connectional relations to decide the relative pertinence of representations. Given the empirical results, it seems likely that the ER-structures exhibited by one's ENGINEER and LAWYER concepts were not decisive between whether Dick is an engineer or lawyer. As in the case of the Linda problem, the ER-structure invoked by the nature and content of the task supersede consideration of the base-rates. More specifically, the conceptual content of the Dick problem is not conducive to and interferes with activating the appropriate *probability* conceptualizations, and thus the probability assigned to the statement that Dick is an engineer is made via a heuristic procedure rather than via probabilistic calculations. This hypothesis is consistent with the fact that when descriptions were omitted participants made their judgments in accordance with probability theory—a situation in which there was no information to interfere with exciting the appropriate *probability* conceptualizations. In the absence of descriptions, one's ENGINEER and LAWYER concepts may be active, but the content that would influence one to ignore the base-rate is not there.

In contrast to Tversky and Kahneman's assertion that the descriptive information of Dick is irrelevant to the problem, I believe that such information is in fact relevant. At the very least, the active conceptual content, and the extant relations between such information, proves to be *informative* insofar as ER-structures tend to reflect the way information is structured in the real world—one's concepts and conceptual content are, after all, heavily influenced by empirical learning. Again, however, I will not pursue this matter any further at the moment. I mention this point here only to intimate some of what I will be arguing in the next chapter.

5.1.2.3 The cab problem

Another example that exhibits base-rate neglect is the following cab problem.

> A cab was involved in a hit and run accident at night. Two cab companies, the Green and the Blue, operate in the city. You are given the following data:
> (a) 85% of the cabs in the city are Green and 15% are Blue.
> (b) A witness identified the cab as Blue. The court tested the reliability of the witness under the same circumstances that existed on the night of the accident and concluded that the witness correctly identified each one of the two colors 80% of the time and failed 20% of the time.
>
> What is the probability that the cab involved in the accident was Blue rather than Green knowing that this witness identified it as Blue?

This problem is unlike the Dick problem, since the latter illustrates how supposed irrelevant information invites the use of the Representativeness heuristic, causing one to ignore prior probabilities. The cab problem, on the other hand, is supposed to illustrate that the Representativeness heuristic can be triggered even while cogitating relevant information. Here, (a) and (b) are both relevant to determining the probability that the cab involved in the accident was Blue, and the correct answer is arrived at through Bayes' theorem: the probability is 0.41.[5] Thus, despite the witness' report, the hit-and-run cab is more likely to have been Green, and this is because the prior probability that the cab is Green is higher than the credibility of the witness. However, Tversky and Kahneman (1980) report that participants who are presented with this cab problem typically answer that the probability that the cab was Blue is about 0.80, which happens to coincide with the credibility of the witness. Tversky and Kahneman interpret this result as participants employing the Representativeness heuristic, taking the witness' credibility as representative of the probability that the hit-and-run cab was in fact Blue.

I am not convinced that Tversky and Kahneman have correctly described the processes underlying people's judgments on this problem. Rather, I believe that the cab problem simply reinforces my claim that most people—namely the naïve and novices—do not *know* how to solve probability problems involving base-rates. According to my account, the incorrect responses are the result of an epistemic deficiency with respect to treating probabilistic information. However, I suspect that, unlike the Linda problem and the Dick problem, if the cab

[5]Where H = the hypothesis that the cab was Blue, and E = the reliability of the witness' testimony:

$$\Pr(H|E) = \frac{\Pr(E|H)\times\Pr(H)}{[\Pr(E|H)\times\Pr(H)]+[\Pr(E|\neg H)\times\Pr(\neg H)]} = \frac{0.8\times0.15}{[0.8\times0.15]+[0.2\times0.85]} = 0.41$$

problem was presented to knowledgable statisticians, or indeed if the problem was on a final exam for a statistics class (with bright students), incorrect responses would not be the majority. In the terms I have used above, the PROBABILITY concepts possessed by the participants in Tversky and Kahneman's study did not bring to bear the appropriate conceptual content to facilitate the correct answer. On the other hand, statisticians or students in the context of an exam on statistics will likely bring to bear the appropriate conceptualizations of (probabilistic) information presented in the cab problem. For not much other conceptual information is invoked by the cab problem to enable the proposed interference effects observed in the Linda and Dick problems.

Nevertheless, I bring up the cab problem for discussion because it can be used to illustrate a different point, namely that people's responses to the cab problem can be manipulated by the format in which probabilistic information is presented. Probabilistic information can be presented in a number of ways: as percentages, as fractions, numerically, pictorially (e.g., in pie-charts), and so on. Notice that the manner in which probabilistic information is presented in the cab problem is as numerical percentages. According to Gigerenzer and Hoffrage (1995, 1999), the human brain is not equipped to reason with numerical percentages, and I tend to agree. More precisely, I believe that humans generally do not possess the conceptual wherewithal to appropriately conceptualize probabilistic information in terms of percentages. Gigerenzer and Hoffrage illustrate, on the other hand, how probabilistic reasoning can be facilitated when probabilistic information is presented in what they call "natural frequencies", i.e., unnormalized frequencies observed by encountering instances in a population sequentially. As they illustrate, natural frequencies carry information about base-rates, whereas normalized frequencies or probabilities do not. Thus, when one reasons with natural frequencies, one need not bother with understanding how to appropriately integrate base-rate information in making inferences or predictions, since such information is already built in.[6]

[6]Despite Gigerenzer's frequent accusations that Kahneman and Tversky focus too much on the mistakes and biases that occur as a result of employing heuristics, Kahneman and Tversky have in fact investigated how reasoning is affected by the format in which probabilistic information is presented, and they have shown that probabilistic reasoning can be thereby improved. See Kahneman and Tversky (1979, 1973a); Tversky and Kahneman (1980,

To express the probabilistic information in the cab problem in natural frequencies, (a) and (b) would be replaced by the following:

(a′) 85 out of 100 are Green, 15 out of 100 are Blue.
(b′) A witness identified the cab as Blue. The court tested the reliability of the witness under the same circumstances that existed on the night of the accident and concluded that the witness correctly identified each cab 8 of every 10 times and failed 2 of every 10.

The question posed would then be:

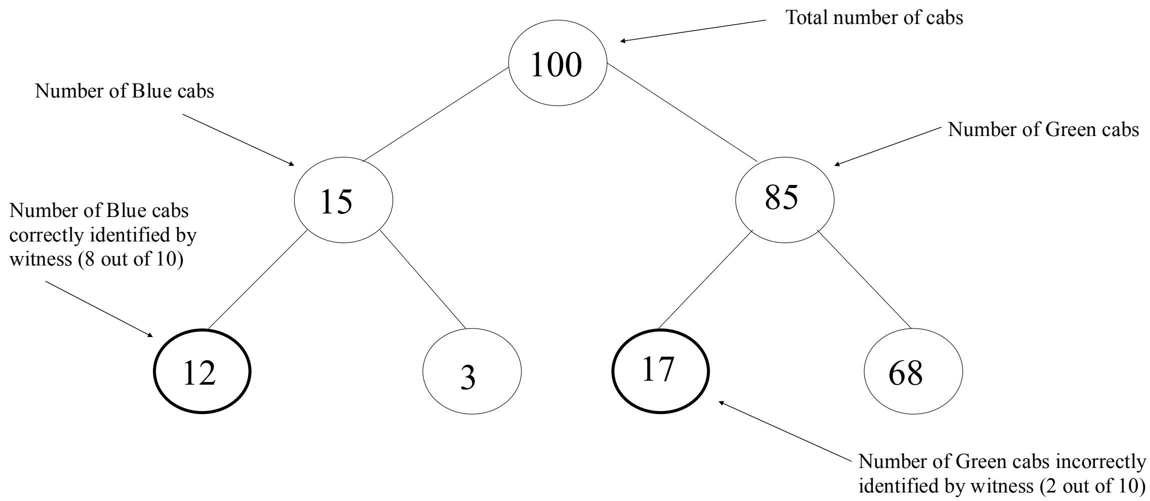
The witness will identify _______ out of _______ cabs as Blue.

Reasoning would proceed as follows. Of the 15 cabs that are Blue, the witness would correctly identify 12 (8 of every 10). Of the 85 cabs that are Green, the witness would incorrectly identify 17 (2 of every 10); this means that on these occasions the cab was actually Green. Thus, the witness would identify 12 cabs as Blue of a possible 12 + 17 cases in which the cab is not actually Blue. So the answer is $\frac{12}{(12+17)} = 0.41$, which is the correct answer. Gigerenzer and Hoffrage (1995) performed experiments, the results of which show that people have an easier time arriving at the correct answer to the cab problem when the probabilistic information is presented as given here rather than as in the original (although not everyone arrives at the correct answer).

The natural frequency information can be also be represented graphically, as presented in Figure 5.1 (cf. Gigerenzer & Hoffrage, 1995). When the information is represented in this way, it is easily seen that the witness will have correctly identified 12 out of 12 + 17 cabs as Blue.

This illustrates that the manner in which information is (re)presented affects the way it is cognized. In terms of the thesis of this dissertation, changing the format in which information is presented will activate different conceptualizations; and the different conceptualizations will facilitate different ways to reason about the information given. Specifically, when the same information is represented in natural frequencies, one is able to conceptualize it in ways that

1983).



The witness will identify ______ out of ______ cabs as Blue.

$$\frac{12}{12 + 17} = 0.41$$

Figure 5.1: A graphical representation of the cab problem expressed in terms of natural frequencies.

make it easier to cognize than when the same information is presented in percentages; and when natural frequencies are represented graphically, the task may become even easier.

There are three things to note here. First, in addition to the fact that natural frequencies embed base-rate information, when the cab problem is presented in Gigerenzer and Hoffrage's reformulation, the question that participants are expected to answer (viz. "The witness will identify _____ out of _____ cabs as Blue") is not equivalent to the question that participants are expected to answer in the original problem (viz. "What is the probability that the cab involved in the accident was Blue rather than Green knowing that this witness identified it as Blue?"). Gigerenzer and Hoffrage's formulation asks participants to give a fraction with respect to the number of cabs that *the witness* will identify as Blue, whereas Kahneman and Tversky's formulation asks participants to give *their own* probability that the cab involved in the accident was indeed Blue. The actual problem is therefore different between the two formulations. This is an important point, for what is asked of the participants will affect the way in which the problem itself is conceptualized; and different conceptualizations of information will be more or less helpful to different conceptualizations of the problem. Here, the way Gigerenzer and Hoffrage present the probabilistic information in the cab problem is conducive to the manner in which they present the information to be predicted. I suspect, however, that their participants would not have as easy a time arriving at the correct answer if the question posed was that of the original (viz. "What is the probability that the cab involved in the accident was Blue rather than Green knowing that this witness identified it as Blue?") or something close (e.g., "What is the probability that the cab involved in the accident was Blue?").

A second (and related) thing to note is that the manner in which the problem is conceptualized between the two formulations affects the cognitive task elicited from participants. Kahneman and Tversky's formulation elicits a task of determining a probability based on numerical information, whereas Gigerenzer and Hoffrage's formulation does not elicit such a task, but a categorization task. In chapter 3, I cited Sterelny as claiming, "the practice of marking a trail while walking in the bush converts a difficult memory problem into a simple perceptual problem" (Sterelny, 2006, p. 225). The same sort of thing is happening here between Gigerenzer and Hoffrage's formulation of the cab problem on the one hand, and Kahneman and Tversky's formulation on the other. Gigerenzer and Hoffrage's formulation converts a difficult probability problem into a simpler categorization problem. When probabilistic information is presented in natural frequencies, the task is changed from predicting a probability to grouping together the number of instances when the witness identified cabs as Blue, and comparing it to the number of instances when the witness was actually correct. This categorization task is made even easier when represented graphically as in Figure 5.1. However, in Kahneman and Tversky's formulation, the task is much more difficult because none of these resources is available to solving a probability problem as such. And, as I am claiming here that the conceptual content of people's PROBABILITY concept tends to be impoverished, people will generally have a hard time solving the problem in Kahneman and Tversky's formulation.

The third thing to note is that the way in which we solve the cab problem does not have much to do with heuristic procedures as I have been understanding them here. For the way in which the cab problem is solved—in either Kahneman and Tversky's or Gigerenzer and Hoffrage's formulation—is not by exploiting any metainformational structures such as those embodied by ER-structures. Rather, the cab problem is essentially a problem in formal probability which requires certain skills and knowledge to arrive at the correct answer. Thus, contrary to what Kahneman and Tversky claim, I claim that the cab problem does not exhibit the use of the Representativeness heuristic, or any other heuristic. I discussed the cab problem at length, however, to illustrate that conceptualizations of a problem and conceptualizations of the information relevant to solving the problem or making predictions interact in specific and important ways. The format in which information is presented, and the manner in which questions or problems are posed, therefore affect one's ability to reason about and solve the problem. Moreover, to corroborate the general picture of concepts and cognition advanced in this dissertation, I wanted to illustrate that one's conceptual wherewithal plays a crucial role in reasoning insofar as it is equipped to produce the kinds of conceptualizations to facilitate certain ways of reasoning about the problem. I suspect that the account I am offering here with respect to the cab problem can be applied more generally to *framing effects*,[7] however I will refrain from exploring this matter here.

5.2 Fast and frugal heuristics

Let us now turn to Gigerenzer's fast and frugal heuristics. I will here comment on two further heuristics: the Recognition heuristic and Take the Best. As it will presently become evident, these heuristics can likewise be understood as procedures that exploit and are constrained by extant, informationally rich structures between active concepts and content.

[7] The framing effect is a phenomenon wherein the same information presented in different ways elicits different and contradictory responses.

5.2.1 The Recognition heuristic

The Recognition heuristic is perhaps the simplest of Gigerenzer's heuristics. The way it basically works is one chooses the alternative she recognizes:[8]

(14) (i) Search among alternatives; (ii) stop search when one alternative is recognized; (iii) choose the recognized object.

However, this heuristic (like any other) is not successful and accurate in all environments. It works only when just one alternative is recognized, while the others are completely unrecognized.[9] Moreover, "The single most important factor for determining the usefulness of the Recognition heuristic is the strength of the relationship between the contents of recognition memory and the criterion [being predicted]" (Goldstein & Gigerenzer, 2002, p. 78). What is being referred to here is the *recognition validity*, which is the proportion of instances when a recognized object has a higher criterion value than an unrecognized object (in the same class of objects). The Recognition heuristic is defined as successful when the recognition validity is greater than 0.5 (i.e., when there is an above-chance correlation between recognition and the criterion being predicted).

Goldstein and Gigerenzer (1999, 2002) hypothesize how the Recognition heuristic works. They note that criteria that are being predicted are not always readily available or accessible, and so a person has to make inferences about it. Inferences, of course, are based upon an individual's previous knowledge of the class of objects in question, some of which comes

[8]The Recognition heuristic is really supposed to apply only when there are two alternatives, and no more. I do think, however, that this heuristic can be generalized to apply to situations in which there are more than two alternatives.

[9]Goldstein and Gigerenzer (2002) use the term "recognition" to "divide the world into the novel and the previously experienced" (p. 77). The way I use the term "completely unrecognized" coincides with their understanding of being confronted with a novel object. To have previous experience with an object means that it is recognized, and this experience can be vaguely remembered (a situation Goldstein and Gigerenzer call "mere recognition") or richly remembered (a situation where the object is recognized *and* additional information can be provided about it). Though I do not want to enter into any kind of discussion on the matter here, I believe that when an object is vaguely remembered, one possesses conceptual content about the object, some of which may even be primed upon vague remembering, but such information does not make in into conscious thought. Thus, as far as possession of knowledge about an object is concerned, I do not think that there is much of a difference between the object being vaguely remembered and it being richly remembered.

from "mediators". Mediators, according to Goldstein and Gigerenzer, are third-party entities that indirectly carry information about the criterion in question. As we shall soon see, being mentioned in newspapers can serve as a surrogate for information about relative city size, and in this way it is a mediator for the latter criterion. An individual can make accurate predictions on some criterion provided that there is a correlation between the criterion and the information carried by the mediator (what Goldstein and Gigerenzer call the "ecological correlation"), and provided that there is a correlation between one's recognition memory and the information of the mediator (what they call the "surrogate correlation"). Recognition can therefore carry information about a given criterion via a mediator.

Despite its simplicity and frugality, Gigerenzer and his colleagues have touted the Recognition heuristic to make more accurate predictions on a number of tasks than a variety of standard rational and statistical methods of decision, such as multiple regression (for the most discussion, see Gigerenzer et al., 1999). Gigerenzer (2007) goes on to argue that the Recognition heuristic can be used to explain a wide variety of phenomena, varying from which product a consumer is more likely to buy, to predicting which football team will win a game, to which symphony will be favoured—the recognized brand will most likely be purchased by a consumer, the recognized team will be predicted to win, the recognized symphony will be favoured.

Let us consider Goldstein and Gigerenzer's (2002) experiment which tested the use of the Recognition heuristic. They presented a group of American undergraduates with pairs of names of German cities. The task was to choose the larger city in each pair, and to indicate whether they recognized the names of the cities. Goldstein and Gigerenzer found that in the vast majority of instances where one city was recognized but not the other, participants predicted that the city they recognized was the larger of the two. Hence, Goldstein and Gigerenzer concluded that the Recognition heuristic was being employed. In addition, they found that the Recognition heuristic produced, on the whole, accurate predictions. They explain such accuracy in terms of mediators: as recently mentioned, being mentioned in newspapers is a good surrogate of information about the sizes of foreign cities, for as Goldstein and Gigerenzer show, there is a correlation between city size and the number of times the city was mentioned in newspapers.

A more startling result that Goldstein and Gigerenzer discuss is that the American students were better at predicting the relative sizes of German cities than predicting the relative sizes of American cities. They call this phenomenon the "less-is-more effect": "The Recognition heuristic leads to a paradoxical situation in which those who know more exhibit lower inferential accuracy than those who know less" (p. 79). They explain this instance of the less-is-more effect by claiming that the American students were unable to use the Recognition heuristic on American cities, since both cities of many pairs were recognized, reducing their responses to guesses if no other knowledge aided in their predictions. Of course, Goldstein and Gigerenzer recognize that knowing less is not always going to produce better predictions. Knowing neither cities in a given pair is going to produce only chance guesses; and knowing lots about each city may produce more accurate predictions, especially if one possesses a piece of knowledge which directly indicates which of two cities is larger. Thus, the less-is-more effect really only occurs when one relies on recognition to make an inference, but one's knowledge does not exceed recognition validity.[10]

However, a more general way to explain the use and success of the Recognition heuristic is through the exploitation of the connectional structures among active conceptual information. The recognized object will in most cases activate more concepts and conceptual information than the unrecognized one, thereby imposing initial constraints on heuristic reasoning. For example, suppose of Berlin and Augsburg, Berlin is the recognized city while Augsburg is completely unrecognized. Upon reading or hearing "Berlin", a number of conceptualizations contained within or related to one's BERLIN-file may be activated, such as those having to do with the Berlin Wall, or the fact that Berlin is the capital of Germany, or reading about Berlin

[10]On the present account of heuristics, knowing too little or too much about objects (without specifically knowing the criterion) creates informational structures over which it is difficult for heuristic procedures to navigate. On the one hand, if too little is known, the ER-structure would be impoverished, making heuristic operations over its structural relations difficult. On the other hand, if one knows quite a lot (without specifically knowing the criterion), then there may be too many connections over which a heuristic must operate; in this case, employing a heuristic may not be an option. See the discussion at the end of chapter 4 (section 4.4).

in the news quite a bit lately, and so on. However, Augsburg, by virtue of it being completely unrecognized—by virtue of not possessing any knowledge of it—will invoke minimal conceptual information.[11] Thus, there will be nothing in one's (newly created) AUGSBURG-file to suggest one way or the other whether it is a large or small city. On the other hand, there is ample conceptual information in one's BERLIN-file that suggests that it is, or is likely to be, a large city. For the representations and conceptualizations activated by entertaining BERLIN is saturated with information that can indicate that it is a large city: a famous wall that divided a city post-World War II would probably not have been erected in a small city, but a large and important one; capital cities of countries tend to be large; small cities usually do not make the news frequently, but large cities do; and so on. All this information is embedded in or entailed by one's BERLIN concept, and hence, LARGE CITY would bear a number of inward connections from the conceptual content activated by opening one's BERLIN-file. Again, these structural relations will not be exhibited by the conceptual content activated by AUGSBURG (all of which comes from the context within which Augsburg is mentioned; see footnote 11). A heuristic procedure operating over such thin relational structures will therefore indicate that Berlin is a large city, while no such metainformational structures indicate this for Augsburg. One would thereby predict that Berlin is the larger of the two.

We can agree with Goldstein and Gigerenzer that mediators exist and carry information about certain criteria. This is simply an empirical fact, uninteresting as it is. What is more interesting, however, is what information is actually carried by mediators, how we cognitively manage such information, and what we do with it. In the above example, the fact that Berlin once had a famous wall dividing the city, the fact that Berlin is the capital of Germany, and the fact that Berlin has been in the news lately, all served as mediators. But mediators themselves assume a great deal of conceptual knowledge. Newspapers, for example, carry information about relative sizes of cities, but this information is caught up in a web of knowledge con-

[11]Of course, it will not be the case that *no* conceptual information will be invoked. For in the context of the (thought) experiment, one would understand that Augsburg is a city, and that it is supposed to be a city in Germany. Thus, even if one has not heard of Augsburg before, a file would be created for it from what one can gather from the context.

cerning *why* such cities would be in the news—it is no accident, for instance, that small towns are rarely, if ever, mentioned in the *New York Times*. As should be obvious by now, I suggest that such webs of knowledge play important roles in heuristic inference and reasoning insofar as they embody higher-order information about what concepts and conceptualizations are pertinent and therefore useful in making predictions.

Suppose, for instance, that instead of Berlin and Augsburg, Alice was presented with Wittenberg and Augsburg, and was asked to predict which is larger. Let us further suppose that Alice does not recognize Augsburg but recognizes Wittenberg only because she recalls Bob once mentioning in passing that he had once visited that city. Basing her inference on recognition, Alice predicts that Wittenberg is the larger city. This prediction is incorrect, however—Augsburg is in fact the larger of the two. Goldstein and Gigerenzer might blame Alice's incorrect response on a poor ecological correlation between, on the one hand, the relative sizes of Wittenberg and Augsburg, and on the other hand, Bob's mentioning that he had visited Wittenberg; or they might blame it on a poor recognition validity. But ecological correlation and recognition validity are simply factors that contribute to the description of why the Recognition heuristic may or may not work in a given instance. In contrast, the theory I am advancing here tells us what is actually doing the work when we rely on recognition in our reasoning. With respect to Alice, if remembering that Bob had visited Wittenberg did not excite for her any information that bore specific relations to certain representations having to do with the relative size of the city, then, since there would be little or no structural information to guide heuristic inference, she would have simply guessed (wrongly) that Wittenberg was larger based *ex hypothesi* on the mere fact that she recognized the name "Wittenberg". On the other hand, remembering that Bob visited Wittenberg might have actually invoked for Alice a specific ER-structure among activated concepts and content. For example, she might have activated conceptualizations having to do with remembering that Bob detests visiting small cities, or having to do with the fact that tourists tend to visit bigger cities. Such conceptualizations might thereby bear outward connections to her LARGE CITY concept with respect to her conceptualizing *Wittenberg*, few (and maybe weak) as these connections may be. In this case, Alice's prediction that Wittenberg is the larger city would have resulted from the Recognition heuristic operating over the said thin connectional relations. And so contrary to what Gigerenzer would assume, Alice's prediction would have been based on more than just recognition.

Further phenomena can be explained by my theory that cannot be explained by the Recognition heuristic alone. Daniel Oppenheimer (2003) conducted studies similar to Goldstein and Gigerenzer's. In one study Oppenheimer asked a group from Stanford University to report which of Chernobyl and Heingjing is larger, and in another he asked a different group to report which of Milpitas and Heingjing is larger. Chernobyl is widely known for its nuclear disaster in 1986, though this information is supposed to be irrelevant to its size.[12] Milpitas is a small city of about 62,000 inhabitants,[13] situated just north of San Jose, which was presumed to be known to the participants. Heingjing, however, is a fictitious city; the name was made up by Oppenheimer for the purposes of his experiments. The results are interesting: Neither group, on the average, relied on the Recognition heuristic. Instead, they predicted that Heingjing is larger than Chernobyl, as well as Milpitas. Oppenheimer uses these results to challenge Gigerenzer's claim that usage of the Recognition heuristic is ubiquitous and widespread.

Gigerenzer (2007, 2008c), however, retorts by asserting that there are actually two steps to employing heuristics. For the Recognition heuristic, the first step is recognition: whether one alternative is recognized, and the others not. This first step determines whether the Recognition heuristic is applicable. The second step is an evaluative one: an assessment of the ecological rationality of the heuristic. If by this evaluation it is judged that the heuristic is not ecologically rational, it will not be used, and a search for some other (perhaps heuristic) strategy commences. Hence, Gigerenzer explains Oppenheimer's results by claiming that the participants did not use the Recognition heuristic after its ecological validity was evaluated. Gigerenzer asserts that the participants knew that the nuclear disaster in Chernobyl has nothing to do with

[12] But see my remarks above regarding the Representativeness heuristic and the supposed irrelevance of information.

[13] At the time of the experiment.

population size, and that they also knew that Milpitas is a small city. He goes on to claim that this knowledge overrode the Recognition heuristic: "In both cases, many participants did not follow the Recognition heuristic and answered 'Heingjing.' They recognized a city for reasons independent from population (which invalidated the heuristic in this situation) or because they had direct knowledge about the criterion (population)" (Gigerenzer, 2008c, p. 25).

Nevertheless, the results of Oppenheimer's studies are explained more naturally by my theory of how heuristics work. In both of Oppenheimer's studies, being presented with the name "Heingjing" not unlikely produced representations and conceptualizations having to do with China and cities in China. Most people know that China is a densely populated country, containing the largest population on the planet. This knowledge would have excited conceptualizations that would almost certainly represent Heingjing as a large city. When this representational information is compared to the representational information activated for CHERNOBYL (such as information having to do with the nuclear disaster, and whatever else bears extant relations to this) and MILPITAS (such as information having to do with its being a relatively small city), it suggests that Heingjing is likely larger than both Chernobyl and Milpitas. This of course presumes that neither CHERNOBYL nor MILPITAS induced conceptual content that exhibited structural relations that embody appropriate connections to one's LARGE CITY concept. If this were the case, or if the suggested informational structure did not exist for HEINGJING, then Heingjing may not be predicted to be larger.

In this situation I am agreeing with Gigerenzer that those who predicted that Heingjing is the larger city did not employ the Recognition heuristic. What I am suggesting, however, is that Gigerenzer fails to recognize, or at least appreciate, the crucial role played by our concepts and the extant relations between conceptual content in heuristic reasoning. Recognition carries information (via mediators) about the thing recognized, but it is not recognition *per se* that is doing the work when we employ the Recognition heuristic. Indeed, in such instances our inferences are based on a lot more information than mere recognition,[14] for it is the structure of

[14] I am not here using "mere recognition" as Goldstein and Gigerenzer (2002) do. See footnote 9.

relations between activated concepts and conceptual content that guides our heuristics. Again, in this way activated concepts and conceptual content constrain heuristics. Thus, Gigerenzer does not need to offer *ad hoc* revisions to the operations of the Recognition heuristic (viz. claiming that there are two steps). Instead, if my account is accepted, the extant ER-structure will indicate that some heuristic other than the Recognition heuristic will be applied.

5.2.2 Take the Best

As mentioned above, when more than one object in a group is recognized, the Recognition heuristic cannot be applied. However, if additional information about the recognized objects can be recalled, Gigerenzer (2001; Gigerenzer & Goldstein, 1996) claims that an individual may rely on the Take the Best heuristic:

> (15) (i) Search among alternatives; (ii) stop search when one alternative scores higher on some predetermined criterion; (iii) choose the object identified in step ii.

Take the Best assumes that an individual has a subjectively ranked order of beliefs about cues which may discriminate between objects and which are considered sequentially when making predictions.[15] The highest ranked cue which discriminates is termed "the best", and the objects that have a positive value with respect to this best cue will be chosen or predicted to possess some criterion. This idea behind Take the Best was briefly explained in chapter 2 (section 2.4.2.2) and in a little more detail in chapter 3 (section 3.4.2).

Gigerenzer's examples and experiments that illustrate Take the Best are much like the ones given for the Recognition heuristic.[16] For example, suppose you had to choose which of a pair of cities is larger. Both are recognized, but you believe that having a university is the best cue that discriminates on the criterion of city size, and you know that one city has a university while the other does not. You would thus use Take the Best, and infer that the city with the university

[15] Like the Recognition heuristic, Take the Best is supposed to apply only when there are two alternatives. But, again like the Recognition heuristic, I believe that Take the Best can be generalized to apply to situations in which there are more than two alternatives.

[16] In fact, the Recognition heuristic might be understood as a special case Take the Best, if recognition is taken to be a cue.

is the larger of the two.

Here again we can explain this heuristic in terms of my theory of concepts and heuristics, for Take the Best, too, exploits informationally rich conceptual structures. Indeed, as I had indicated in chapter 2 (section 2.4.2.2), significant amounts of conceptual information is typically entailed by the very belief that a cue discriminates. For instance, I had claimed that certain beliefs are implied about the cue upon which the choice in question was made—at the very least, we can infer that the said cue was believed to be the "best" to make the choice in question, and that the cue is believed to be the "best" implies certain things about the conceptual content of the cue (as possessed by the chooser), as well as certain things about how the cue fits within the chooser's conceptual structure. In chapter 3 (section 3.4.2), we saw in a bit more detail what we might infer about the conceptual content of the "best" cue as possessed by the chooser. For example, if one believes that having a university is a good predictor of relative city size, one must possess a concept UNIVERSITY which is rich enough to support such a belief, or at least to support the belief that having a university is a better predictor of relative city size than some other cue. And the conceptual information built into one's UNIVERSITY concept might include, in addition to various representations and conceptualizations that are required to understand what a university is, information about a university's function in society; how it houses a number of faculty, staff, and students; perhaps the relative social classes of members of a community typical of such faculty, staff, and students; maybe the ways in which having a university relates to the economy of a city; and probably much more. Indeed, it is such information that generally *makes one believe* that having a university is a good predictor of relative city size. (We see here again hints of Dretske: A cue will be treated as a discriminating cue only if one possesses the appropriate conceptual wherewithal to do so.)

The process by which I understand Take the Best to operate is similar to the one outlined for the Recognition heuristic. The nature and content of the cognitive task of predicting which of two cities is larger, for example, will activate and prime various concepts and conceptual content. If only one of the two cities happens to have a university, then conceptualizations pertaining to a university's function in society, how it houses a number of people, etc., will be among what is activated or primed. These latter representations will bear strong outward connections to one's LARGE CITY concept, which is just to say that LARGE CITY will bear a number of strong inward connections. Thus, a heuristic procedure operating over these thin ER-structural connections will suggest that having a university is indicative of a large city, and as a result, the city with the university will be predicted to be larger than the other.

On this view, we might thus understand one's belief that having a university is a good predictor of relative city size to be implicit in the metainformational ER-structure; that is, beliefs about discriminating cues need not be explicitly represented. Assuming that we explicitly represent pre-ranked sets of cues is implausible, in any case. For we must not only have a pre-ranked set of cues for predicting relative city size, but a set for any potential decision task that we are faced with, or at least any potential decision task concerning objects that we have some familiarity with. This must be the case if Take the Best is available for use for any such decision task. If I am to predict which car would be faster, I must have pre-ranked cues for that task; if I am to predict which student will score higher on the final exam, I must have pre-ranked cues for that task; if I am to predict which apple will taste better, I must have pre-ranked cues for that task; and so on. Yet carrying around all sorts of pre-ranked sets of cues in our heads is cognitively demanding in terms of storage space—the human brain is far too limited to be able to store all these pre-ranked sets of cues. To be sure, the number of sets of pre-ranked cues would have to be unbounded, since there is an unbounded number of possible decision tasks. This is an independent reason why we should doubt the framework Gigerenzer proposes for Take the Best.[17]

Notice that on the view I am offering there is no need to presume that each of us has a pre-ranked set of potentially discriminating cues, even if they are implicitly represented. We might agree with Gigerenzer that we have sets of cues which we may rely upon to predict which object

[17]As noted in footnote 32 in chapter 3, I suppose that one can produce cue rankings on the fly, as the situation arises. However, according to Gigerenzer, we must have predetermined cue rankings stored in memory (Gigerenzer & Goldstein, 1996, 1999).

scores highest on some criterion (based on rich conceptual information), but on my proposal we do not have to assume that the cues are in any way ranked or otherwise ordered. Rather, whatever concepts and conceptual content are activated and the ER-structure will determine which, if any, cues are treated as discriminating cues. So, for example, suppose one must predict which of Kitchener and London$_o$ (Ontario) is larger. One's KITCHENER and LONDON$_o$ concepts would be entertained, and their respective files searched to see if either possesses information that discriminates with respect to relative size. Once it is found that London$_o$ has a university while Kitchener does not, the ER-structure just outlined would obtain; and if no other discriminating information is revealed—if no further representations are invoked which would alter the ER-structure to suggest otherwise—then one might predict that London$_o$ is larger than Kitchener. This does not require that one has already determined that having a university is a good predictor of relative city size. Instead, if there is no other discriminating information that comes to mind after consulting one's KITCHENER and LONDON$_o$ files, one can simply make an inference on whatever the heuristic delivers according to the information available.

Notice furthermore that, whereas Gigerenzer's account entails that one and only one cue is used in making a heuristic inference (via Recognition or Take the Best), my account does not entail this. Instead, a number of relevant cues may be considered, and a heuristic inference may be made based on the conceptual information and ER-structure activated by them. Thus, in addition to doing away with the presumption of pre-ranked sets of cues, we do not need to suppose that cues are considered in a sequential manner. Suppose again that one had to predict which of Kitchener and London$_o$ is larger, that one believes that London$_o$ has a university while Kitchener does not, and that one (implicitly) believes that having a university is a good predictor of city size. This time, however, suppose that one is told that Kitchener has a greater number of residential zones than London$_o$.[18] As before, no other discriminating information is revealed. In this case, the way in which one predicts which city is larger would depend on the conceptual content embedded in UNIVERSITY and RESIDENTIAL ZONES concepts. For example,

[18] I am not sure if this is in fact true. I would assume that it is not. But the truth of the matter is irrelevant.

one's RESIDENTIAL ZONES concept might invoke representations and conceptualizations that make one (implicitly) believe *more residential zones means more people, which means larger city*, and this belief may be weighed against one's beliefs regarding universities and the implications for relative city size. Indeed, one may come to believe that the number of residential zones is a *better* predictor of relative city size, despite the fact that having a university is still believed to be a good predictor. If this is the case, then one might infer that Kitchener is larger than London_o based on the ER-structure that would be established by activating UNIVERSITY and RESIDENTIAL ZONES and conceptualizations thereof.

Given the story that is being told here, I would assume that, in general, when the task is to choose between recognized objects, one will almost always activate more conceptual content than what Gigerenzer assumes. This is to say that it is unlikely that one will only invoke a handful of cues to decide; instead one will, so to speak, search through the concept-files of each concept activated by the nature and content of the cognitive task. I therefore believe that the ER-structure in question will often be thick, since I assume that the conceptual content invoked by the nature and content of the task can very well be substantial, especially if one knows a good deal about each object in question. For instance, one might invoke conceptualizations concerning not only universities and residential zones, but also concerning whether a major highway runs through one or both cities in question; whether one or both cities have a professional sports team; whether one has heard one city in the news more often than the other; whether there is a famous shopping centre in one or both cities; how many airports there are in either city; whether one knows of a major river in either of the two cities; whether one has any friends or relatives who live in or have been to either city; and much more. And all the activated concepts and conceptualizations would bear various inward and outward connections to one another.

I presume that a close analysis would reveal that a number of such active concepts could be grouped according to whether and the extent to which they bear multiple outward connections to specific concepts (these are included among the reference concepts, as described in the pre-

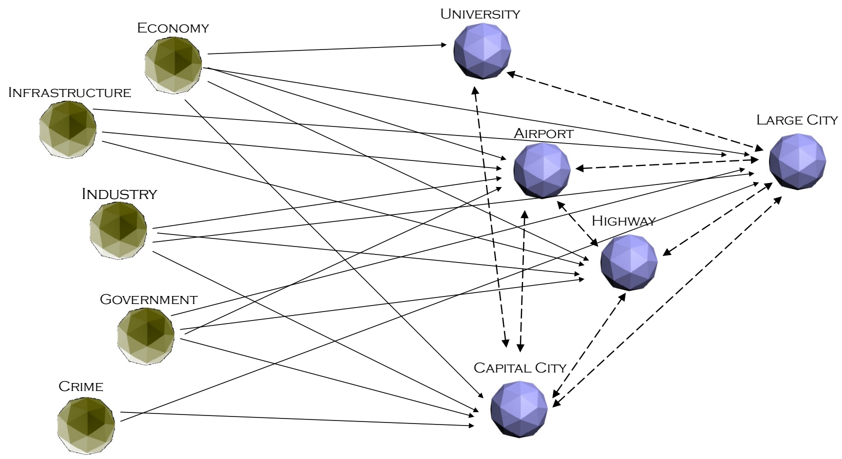


Figure 5.2: Representation of an ER-structure arising from Goldstein and Gigerenzer's (1999) task of predicting which of two cities is larger. My claim is that Take the Best operates over a structure like the one presented here.

vious chapter), and which concepts are bearing the inward connections (these are the pertinent concepts, as described in the previous chapter). For instance, of the many concepts excited by the task we are discussing, some may be grouped into sets according to whether and the extent to which they bear outward connections to UNIVERSITY, AIRPORT, HIGHWAY, CAPITAL CITY, SPORTS TEAM, and MAJOR RIVER; and these concepts implicate individually, and corroborate together, the concept LARGE CITY. Potential sources for such outward connections are ECONOMY, INFRASTRUCTURE, INDUSTRY, GOVERNMENT, and CRIME. The representation in Figure 5.2 would suggest that having a university, being a capital city, having one or more airports, and having a major highway run through the city are implicitly believed to indicate relative city size. Take the Best can thus be viewed as operating over the suggested ER-structure to deliver an inference that the city in question is the larger of the two under consideration.

To gain a better appreciation of how my theory offers a richer and more plausible explanation of how we reason and how Take the Best really works, consider the following. Goldstein and Gigerenzer (1999) report an experiment in which a group of American undergraduates were presented a number of facts (cues) about a variety of German cities, and were coached in understanding which of them were good predictors of city size. The students were then asked to choose which of two German cities were larger. Goldstein and Gigerenzer show that many of the students ignored a good predictor cue—namely, having a major league soccer team—and as a result they performed poorly on many of their predictions. Goldstein and Gigerenzer claim that the participants were instead simply following the Recognition heuristic, despite the availability of a good predictor cue. Nevertheless, I offer an alternative account here: Such "ignored" predictor cues were simply not among the representations that were deemed pertinent by the students' ER-structures invoked by the task. There are many reasons why this might have been so. One possible explanation is that, although the students were coached in understanding which cues were good predictors of city size, they were not sufficiently coached in the *pertinence* of these cues. In order to play the role of a discriminating cue, a cue must be *conceptualized as* a discriminating cue; and this requires that one's conceptual information for the cue is rich enough and bears the right kinds of relations to other concepts and information, as explained above with respect to UNIVERSITY. Only with rich enough content can a representation fulfill the role of a discriminating cue.

This is not to say that Goldstein and Gigerenzer's coaching had no bearing on the American students' predictions. Rather, it is to say that other conceptual information had greater bearing on their reasoning, resulting in incorrect inferences. It is not certain exactly why having a major league soccer team did not bear in a significant enough way on these students' reasoning. However, according to my theory, it likely had something to do with the students' background knowledge of German cities and their MAJOR LEAGUE SOCCER or MAJOR LEAGUE SOCCER TEAM concept. I suspect, for instance, that since the participants were American students, the fact that a certain city has a major league soccer team was not very pertinent because soccer is not among the major American-centred sports, such as baseball and (American) football,[19] and hence they have comparatively little conceptual knowledge having to with the pertinence

[19]As of late, soccer has been gaining popularity as a sport in the USA. Goldstein and Gigerenzer's experiment was conducted before 1999. Nevertheless, I still do not think that soccer would be considered a major American-centred sport.

of having a major league soccer team to predicting relative city size. And even though the students were instructed that having a major league soccer team is a good indicator of relative city size, it is likely that their existing conceptual structures (like anyone's conceptual structure) were so entrenched that they overwhelmed the pertinence of that information. In other words, their MAJOR LEAGUE SOCCER or MAJOR LEAGUE SOCCER TEAM concept were not rich enough to bear the appropriate connectional structures that embody the implicit belief that a major league soccer team is useful in predicting relative city size. As such, such concepts would not be rich enough to support an ER-structure that exhibits the requisite connectional relations over which Take the Best can operate to deliver a prediction involving the cue that a city has a major league soccer team.

If the same experiment was conducted with European participants, I would predict that they would utilize the fact that a city had a major league soccer team differently (it would probably be used as a discriminating cue), since their conceptual understanding of MAJOR LEAGUE SOCCER and MAJOR LEAGUE SOCCER TEAM would likely be adequately rich. In general, I suspect that if the conceptual content of one's concepts were somehow elicited from someone, accurate predictions can be made about what inferences she would make regarding some subject matter concerning the concepts in question. With respect to Goldstein and Gigerenzer's participants, an initial assessment of the American students' concepts would indicate something about their implicit beliefs about the cues' respective pertinence to the given problem. If the American students were given a cue that is a good predictor of relative city size *and* that is congruent with their background conceptual knowledge, the American students would likely implicitly believe such a cue to be useful in predicting relative city size, and it would therefore bear significantly on their reasoning and their resulting predictions. Such cues may, again, include a city being the nation's capital, or the number of airports a city has, or whether a city has a university. Each concept—CAPITAL CITY, AIRPORT, UNIVERSITY, respectively—would probably be rich enough for the American students to implicitly understand and utilize the cue as a discriminating cue, and thereby produce an ER-structure to facilitate correct answers via a heuristic like Take the Best.

To repeat, this is not to say that a discriminating cue would be the only piece of information these students would consider, for other information may very well have a bearing on their reasoning. Rather, this is to say that such a cue will have greater pertinence for these students (i.e., have an important place in their ER-structure) than a cue that is a good predictor of city size but is not congruent with the background conceptual knowledge of the individual.

5.3 General remarks

Gigerenzer has complained on more than one occasion that Kahneman and Tversky's proposed heuristics are "vague labels", and that a given cognitive process can appear to be one heuristic under one description, but appear to be a different heuristic under a different description (e.g., Gigerenzer, 1996; Gigerenzer & Todd, 1999). Gigerenzer may very well be right. Kahneman and Tversky's heuristics certainly bear a certain level of vagueness insofar as there is no determinate measure of representativeness or availability. Nevertheless, what I hope was made clear in this chapter is that, notwithstanding that a given cognitive process can appear to be one heuristic under one description but appear to be a different heuristic under a different description, under any description a heuristic process operates over and is constrained by the ER-structures borne by activated concepts and their conceptual content.

Although much more detail can be given as to the manner in which the proposed heuristics discussed in this chapter operate, I believe that the analyses that I have provided is sufficient to see that such heuristics are grounded in the heuristic procedures described at the end of the previous chapter, and more generally in the account of concepts and cognition developed in this dissertation. More work needs to be done to furnish a complete account of the empirical evidence of the use of heuristics. Indeed, there are many other proposed heuristics in the literature that this chapter did not even mention, including those that Simon discusses and a wide variety of methodological heuristics (see chapter 2). However, what has been presented here suggests that my theory of concepts and cognition, and how heuristics work, accounts for enough phenomena to be considered plausible.

It is also important to note that many of the claims made in this chapter are empirically testable. In general, I suspect that with the appropriate experimental set-up, we may be able to infer (to some extent) the conceptual content of one's concepts and the relations between them (perhaps based on association tasks); and if we were able to tease this out, we would be able to make predictions about what ER-structures would get invoked with given inference tasks. We would then be able to make predictions about what heuristics one would employ, and what inferences one would thereby make, based on the account presented here.

Chapter 6

Relevance and the epistemological relevance problem

Throughout this dissertation I have advanced a theory that claims that heuristics work in cognition by operating over the organization of and connections between our concepts and their conceptual content. However, one last important issue remains to be discussed, namely the epistemological relevance problem, which was introduced in chapter 3. As promised, this concluding chapter is devoted to addressing this aspect of the frame problem. This matter will be by no means fully detailed, but I will use my theory of concepts and cognition to sketch a way that the epistemological relevance problem might be circumvented; and this will indicate what role heuristics have, if any, in solving the problem.

In chapter 3 I discussed how heuristics are commonly invoked by various researchers as a way to circumvent relevance problems in cognition. However, I argued that invoking heuristics to avoid the computational relevance problem says nothing of whether heuristics can circumvent the epistemological relevance problem, which is the problem of knowing or determining what is and is not relevant in our cognitive processing. Returning to this matter, I will begin in section 1 by reminding us of the epistemological guise of the frame problem and why it is a problem. I will then discuss in section 2 how to define relevance, for it seems that how and whether information is determined to be relevant depends on what property *relevance* picks out. I will briefly expound the influential account of relevance developed by Dan Sperber and Deirdre Wilson (1986/1995), and then argue for a reinterpretation of their theory within the framework of cognition I developed in this dissertation. This will allow me to assess the role heuristics have in making determinations of relevance, as well as to explain the reasonable levels of success exhibited by humans in judging what is and is not relevant in their cognitive tasks.

In section 3 I conclude that our conceptual wherewithal predetermines what I call *de facto* relevance (cf. Gabbay & Woods, 2003). Two important theses are entailed by this position. One is that heuristics do not actually make determinations of relevance, and so they actually have no role in circumventing the epistemological relevance problem. But this thesis is not as severe as it may appear, for I will also argue—which is the other important thesis—that humans do not in any case solve the epistemological relevance problem; so heuristic solutions *ipso facto* are empty. Contrary to what Fodor believes (see chapter 3), it really does not matter much whether we solve the epistemological relevance problem.[1] Nevertheless, it is important to see the headway my theory makes in confronting the frame problem in general. I will illustrate how my account can help us to better understand how humans solve a different aspect of the frame problem, namely the representational frame problem (as described in chapter 3), which enables us to exhibit our characteristic reasonable levels of success in determining what to bring to bear on our cognitive tasks.

In chastising recent attempts to offer a heuristic solution to the frame problem, Fodor (2008) has remarked that "The rule of thumb for reading the literature is: If someone thinks that he has solved the frame problem, he doesn't understand it; and if someone thinks that he does understand the frame problem, he doesn't; and if someone thinks that he doesn't understand the frame problem, he's right" (pp. 120-121). This concluding chapter might be viewed as a response in the following way. I claim that I think I understand the frame problem, but I do not think to have solved it. At the same time, however, I believe that the account of concepts and cognition offered in this dissertation helps us to understand how we determine what to bring to bear on our cognitive tasks. And insofar as I am successful at this, I hope to show that my account of concepts and cognition (which underwrites my theory of how heuristics work) explains what we actually do in our reasoning.

6.1 The epistemological relevance problem

Let us begin by reviewing the epistemological relevance problem, and the role that is commonly ascribed to heuristics in circumventing it. The cognitive systems that are paradigmatically responsible for general reasoning and decision-making—i.e., central systems—admittedly allow for free exchange of information. A dream of a snake biting its own tail, for example, bore in interesting and important ways on Kekulé's theorizing of the benzene molecule. What is en-

[1] In any event, Fodor's understanding of the problem places unreasonable, unrealistic demands on what is required of central cognition. See chapter 3, section 3.1.2.

tailed by such a free exchange of information, however, is that, provided an appropriate set of background beliefs, any representation held by an agent can *in principle* be relevant to a given cognitive task in central cognition. Who won tonight's football game is *prima facie* irrelevant to whether there is beer in your friend's fridge. But if you believe that your friend's favourite football team played tonight, and that your friend usually overindulges in beer consumption whenever his favourite team wins, then your belief about who won tonight's game actually is relevant to your belief about beer in your friend's fridge. Since relevance can be determined only with respect to one's background beliefs, there is no way *a priori* to circumscribe the subset of representations that are relevant in a given occasion of reasoning or inference. And this is what gives rise to various aspects of the frame problem, as expounded in chapter 3 (section 3.1).

The computational relevance problem, it may be recalled, is one aspect of the frame problem. It is a problem of how a cognitive system tractably delimits (i.e., frames) what gets computed in a cognitive task. On the other side of this coin, the epistemological relevance problem is a problem of how a cognitive system not only ignores most of what it knows, but ignores the right sorts of things in particular, namely what is irrelevant. In chapter 3 I expressed this problem thus:

> **The epistemological relevance problem** The problem of how a cognitive system considers only what is (mostly) relevant, or equivalently, how a cognitive system knows what is relevant.

Humans seem to determine what is relevant in their cognitive tasks quickly and rather easily. This is not to say that we always know and consider what is relevant. For we often fail to do so, especially when cognitive demands are high and/or when cognitive resources are low (cf. Samuels, 2005, forthcoming). Nevertheless, humans characteristically exhibit reasonable levels of success at identifying representations that are relevant to the task at hand. Such reasonable levels of success cannot be ascribed to chance or luck. Therefore, we are left to explain how humans (seem to) solve the epistemological relevance problem, short of considering the totality of one's beliefs (Fodor, 1987, 2000).

As I had pointed out in chapter 3, heuristics are often invoked as a solution to relevance problems, or at least to circumvent relevance problems. I argued that heuristics may very well relieve the computational burden put on a system in delimiting the set of representations to be considered for a given cognitive task—i.e., circumventing the computational relevance problem—notwithstanding that the informational richness and extant relations between concepts are what heuristics owe their utility to. As interesting as this may be, however, this says nothing of whether or the extent to which heuristics can circumvent the epistemological relevance problem. That is, thus delimiting the amount of representations that get considered does not ensure that what does get considered will in fact be relevant to the task at hand. And as I had indicated, the arguments often offered to alleviate the worries of relevance problems (e.g., Carruthers, 2006a, 2006c; Samuels, 2005, forthcoming) do not in fact address this epistemological worry.

Nevertheless, the account of heuristics, concepts, and cognition that I have developed in this dissertation does address the epistemological relevance problem, at least in a certain respect. To fully appreciate this, however, we must first understand that the inquiry into how relevance is determined is constrained by how relevance is defined. In more explicit terms, whether and the extent to which relevant representations are picked out and brought to bear on a cognitive task (by heuristics, for example) will depend on what property we are concerned with. Thus, after relevance is defined we might understand how the present view offers a psychological/cognitive account of the conditions under which heuristics manage to pick out relevant representations.

What I will do in the following section is provide a cursory overview, and somewhat of a critical assessment, of what is arguably the most influential account of relevance—Sperber and Wilson's *Relevance Theory*. My aim is not to offer a full critique of Sperber and Wilson's theory, but to provide an assessment such that I am able to use the theory as a basis from which to synthesize a working definition of relevance that conforms to my view of concepts and cognition. The overarching goal for this working definition is to explain the phenomena of how we determine relevance in reasoning, and this will have implications for the role heuristics have, if any, in bringing to bear relevant representations on the task at hand.

6.2 Relevance

6.2.1 Sperber and Wilson's Relevance Theory

Sperber and Wilson (1982, 1986/1995) developed their Relevance Theory in the context of communication and pragmatics. Humans tend to have an easy time communicating with each other, despite the fact that the meanings of utterances are enormously underdetermined. A simple example: Alice says to Bob, "Isn't that cute?" while nodding toward a chipmunk scurrying up a tree; Bob knows that by "that" Alice is referring to the chipmunk, and not to the birds in the other branches, the tree itself, the landscape, the sound of a horse whinnying, or whatever else was within his perceptual field at the time of her utterance. According to Sperber and Wilson, Bob understands that Alice was referring to the chipmunk because the stimulus of the chipmunk running up the tree was *relevant* (or at least more so than any other present stimulus). Although Sperber and Wilson's Relevance Theory is mainly concerned with verbal/ostensive communication and comprehension in particular, they claim that their theory can be extended to the spontaneous, unconscious inference processes of ordinary thinking (1986/1995, p. 67; p. 75). Sperber and Wilson are able to make this claim because they ground their account of relevance in a fundamental and general view of human cognition.

Sperber and Wilson's view of human cognition is a familiar one derived from general biological principles, and is very similar to some of the tenets on which Gigerenzer and his colleagues base their fast and frugal heuristics research program, as well as to some of the principles which guided Simon's thinking (see chapter 2, section 2.3). Simply put, biological organisms are limited entities with finite resources, and adaptive pressures along with other evolutionary forces compel such organisms to carefully manage their resources. Typically this means that a system is not wasteful in its resource consumption. But more specifically, this means that the potential benefits of performing a task—whatever the goals are—must in some sense be worth the resources invested and used in achieving them. In this sense, the benefits must exceed, or at least in some sense balance off, the resources expended in the pursuit and attainment of the goal. Biological organisms must therefore allocate their resources in such a way that *positively contributes* to performance and survival; or, what amounts to the same thing, they must be *efficient*, operating with efficient mechanisms.

Sperber and Wilson understand cognition to be a biological function, and as such they make a general claim that cognition always tends toward efficiency—maximizing gains and minimizing costs. But Sperber and Wilson make a more particular claim, namely that human cognition succeeds in increasing its efficiency by having a tendency toward *maximizing relevance*. Relevance, for Sperber and Wilson, is a technical term referring to a property of inputs (assumptions) to cognitive processes. According to their Relevance Theory, the relevance of an input is a function of *cognitive effect* on the one hand, and *processing effort* on the other. Hence, maximizing relevance, according to Sperber and Wilson, is a matter of best managing one's cognitive resources to achieve cognitive effects without investing too much processing effort.

Cognitive effects are best understood as contextual effects achieved by a cognitive system. A context is simply just a set of assumptions within which informational input can be processed. Contextual effects are the results of processing input within a context that could not have been effected by the input alone or by the context alone. These are characteristically implications (yielding inferences, and resulting in new assumptions), contradictions (compelling the system to abandon one or more assumptions), or strengthenings (increasing the system's degree of belief in given assumptions).

Sperber and Wilson (1995) make the point that a cognitive system, or more specifically an individual human, is not interested in contextual effects *per se*, but only insofar as such contextual effects contribute to achieving the system's (the individual's) cognitive goals, or otherwise fulfilling its functions. For indeed, there may be contextual effects that are not worth having, or that contribute negatively to the individual's cognitive functions or goals. Thus, Sperber and Wilson define a *positive cognitive effect* as a contextual effect that contributes to the cognitive goals or functions of the individual in a positive way. This is typically achieved by alterations in the individual's beliefs. In short, positive cognitive effects are cognitive benefits; they yield positive cognitive utility; they produce an "epistemic improvement" (Sperber & Wilson, 1995, p. 266); they make a "worthwhile difference" (Wilson & Sperber, 2006, p. 608) to the cognitive functioning of the individual.

Processing effort, on the other hand, incurs a cost on the cognitive system in terms of resources expended (time, energy, and other cognitive resources). Effort is required not only to process inputs and achieve contextual effects—to draw inferences and acquire new beliefs, revise old beliefs, add to and retrieve information from memory stores, and so on—but it is also required to construct an appropriate context within which to process inputs. The expenditure of resources bears negatively on a cognitive system, and therefore bears negatively on relevance.

It should be obvious that relevance, assessed in terms of cognitive effect and processing effort, comes in degrees, and that this is a direct result of both effect and effort being matters of degree. What is more, some input may yield greater cognitive effects on some occasions and less effects on others. This may be due to varying limitations on accessing certain information at certain times, and on constructing appropriate contexts within which to process the input. Or, depending on circumstances related to fatigue or stress, the same input may be more or less easy to process at different times. And, of course, such factors widely vary between individuals, and so one individual may process some input with greater ease and with greater effects while another may find it difficult to process the same input with the same effects. Or one individual may process some input with little effort but with little effect, while another may process the same input with great effort but with greater effect. And so on. What all this means, in short, is that relevance is a relative property—relative to an individual (the cognitive system within which the context is brought to bear on the input in question) and to a time. Hence, Sperber and Wilson (1986/1995) provide the following two conditions for relevance:

(19) *Ceteris paribus*, the greater the cognitive effects achieved by an individual by processing some input within a context, the more relevant that input is to that individual at that time.

(20) *Ceteris paribus*, the greater the effort expended by an individual by processing some input within a context, the less relevant that input is to that individual at that time.

To illustrate this notion of relevance, let us consider a toy example (Sperber & Wilson, 1996): You purchased a lottery ticket and you know there are three prizes: \$100, \$500, and \$1,000. After the lottery has been drawn, you receive a phone call informing you that you have won a prize. The informant can tell you any of the following three things:

(i) You have won \$500.
(ii) You have won \$100, \$500, or \$1,000.
(iii) Either you have won \$500 or, if I'm speaking right now then $\sqrt{2} = 7$.

Of these three statements, all are relevant in the sense characterized by (19) and (20). However, (i) is *most* relevant: (i) is more relevant than (ii) since the former entails the latter, and so (i) yields all the cognitive effects of (ii), plus some more specific effects without greater processing effort. (i) is also more relevant than (iii) since the cognitive effects yielded by both are the same—for (i) and (iii) are logically equivalent—but such effects are easier to derive from (i) than from (iii).

According to Sperber and Wilson, the tendency to allocate cognitive resources to processing available inputs—from the environment or from memory—so as to maximize expected cognitive effects for the least expected effort is the consequence of a general principle which governs cognition:

The Cognitive Principle of Relevance Human cognition tends to be geared to the maximization of relevance.

Thus, when presented with competing inputs, cognition will tend to process what is relevant.[2]

[2]Sperber and Wilson's Relevance Theory makes claims about human cognition in general, but an important consequence of their Cognitive Principle, and one that is the basis for their work in pragmatics, is the Communicative Principle of Relevance: *Every ostensive stimulus conveys a presumption of its own optimal relevance*. This principle is, of course, based on Paul Grice's (1975) conversational maxim of relation (or relevance), which really inspired and motivated Sperber and Wilson's Relevance Theory in the first place. The idea behind Sperber and Wilson's Communicative Principle is that a communicator will have a wide range of stimuli with which to

It is instructive to note at this point the kind of cognitive architecture that Sperber and Wilson envision. In the original edition of *Relevance* (1986), Sperber and Wilson subscribed to a Fodorian cognitive architecture of the human mind, consisting of a variety of specialized input modules individually dedicated to processing perceptual information of a specific domain (e.g., visual, auditory, linguistic, etc.), in addition to a central processing system (central cognition) which makes inferences by integrating (transduced) information received from the perceptual modules or retrieved from memory. Thus, following Fodor, Sperber and Wilson understood the central processing system as nonmodular in nature, and as an inferential centre where many kinds of information come together, get combined or otherwise manipulated.

By the second edition of *Relevance* (1995), Sperber and Wilson had disavowed such a central systems architecture in favour of the massive modularity hypothesis. Though I note this only in passing, I submit that I think certain complications arise for their Relevance Theory within a massively modular architecture. This is a matter to be discussed on some other occasion. However, to avoid these complications, I will continue to refer to Sperber and Wilson's views as if they still subscribed to a central systems architecture; this type of cognitive architecture is closer to the architecture advanced in this dissertation, at any rate.

Against the background of their original cognitive architecture, Sperber and Wilson saw cognitive efficiency largely as a matter of achieving "an optimal allocation of central processing resources" (1986, p. 48). At least this is what they saw as the way by which the mind might achieve efficiency in its inferential processes. Of course, input modules do not face such a challenge of achieving an optimal allocation of resources, for if Fodor (1983, 2000) is right and they are informationally encapsulated computational mechanisms, then there is no challenge for them to appropriately allocate processing resources—a sufficiently small proprietary

communicate her informative intention. To ensure success of communication, it is in the best interests of the communicator to choose a stimulus that will require the least processing effort on behalf of the addressee. This Communicative Principle of Relevance is really the centerpiece of Sperber and Wilson's Relevance Theory, being a significant contribution and highly influential to studies in pragmatics. However, since communication and pragmatics are not the focus of this dissertation, the Communicative Principle will not be discussed any further. Instead, it is the Cognitive Principle that bears significance here, since this principle is supposed to govern general reasoning and cognition.

database guarantees that the system will not have to search for relevant information or need to make computations of relevance (see chapter 3, section 3.1.2).

Sperber (2005) has recently suggested that cognitive efficiency, in terms of maximizing relevance, is achieved biologically—specifically, by optimally allocating energy in the brain. His proposal is to think of maximizing cognitive effects and minimizing effort in terms of noncognitive physiological processes. Sperber asks us to consider the way in which we manage energy consumption in our muscular movements. The use of our muscles depend on chemical reactions involving oxygen, nutrients, and the removal of lactate. Efficiency in muscle movements depends on converting the energy produced by these chemical reactions to work, and letting as little of it as possible degrade into heat. Thus, the way in which muscle efficiency is achieved is not by cognitively representing effort,[3] but by regulating effort noncognitively by physiological processes (e.g., producing the right amount of energy in the right tissues, appropriately adjusting blood flow, removing sufficient lactate, etc.). Sperber believes that cognitive efficiency is achieved by analogous physiological means. He writes,

> There may be physiological indicators of the size of cognitive effects in the form of patterns of chemical or electrical activity at specific locations in the brain. ... Suppose that these physiological indicators locally determine the ongoing allocation of brain energy to the processing of specific inputs. These indicators may be coarse. Nevertheless, they may be sufficient to cause energy to flow toward those processes that are likely to generate relatively greater cognitive effects at a given time. (p. 65)

At the very least, this seems to be a biologically plausible way that the brain might manage cognitive efficiency. Sperber developed this idea within the purview of a massively modular cognitive architecture, but I believe it can equally apply in a central systems architecture.

6.2.2 Relevance in light of the foregoing theory of concepts and cognition

Sperber and Wilson's Relevance Theory seems consistent with the account of cognition and how heuristics work developed in this dissertation. Nevertheless, there are a number of ways

[3]Of course, muscular effort is cognitively represented when the muscle is fatigued. But, as Sperber points out, it is only above this fatigue threshold that muscular effort is cognitively represented, and even then it is in a very coarse manner (Sperber, 2005, p. 64).

in which Relevance Theory is inadequate. For instance, relevance is a property of cognitive inputs according to the theory, but more specifically it is a property of *assumptions*. Assumptions, for Sperber and Wilson, are what get processed within contexts, and assumptions also comprise contexts themselves. According to Sperber and Wilson, assumptions are conceptual representations that are treated as true by the individual (Sperber & Wilson, 1986/1995, p. 2). This would be fine as it is, except that the theory of concepts that Sperber and Wilson adopt is along the lines of the language of thought hypothesis. That is, they believe that assumptions are composed of amodal, atomic conceptual representations over which computations sensitive to their formal semantic and syntactic properties are defined (see pp. 83-93). This runs counter to the foregoing view of concepts which, borrowing from Barsalou's work, envisions concepts as partly comprised of multimodal perceptual information (i.e., perceptual symbols).

It is interesting to note, however, that Sperber and Wilson subscribe to an account of concepts that appears to be just the file model presented above. More specifically, Sperber and Wilson

> assume that each concept consists of a label, or address, which performs two different and complementary functions. First, it appears as an address in memory, a heading under which various types of information can be stored and retrieved. Second, it may appear as a constituent of a logical form, to whose presence ... deductive rules may be sensitive. ... [W]hen the address of a certain concept appears in a logical form being processed, access is given to the various types of information stored in memory at that address. (p. 86)

According to Sperber and Wilson, the types of information that can be stored "under" a concept's address are *logical* (a set of deductive rules that apply to the logical form of the concept), *encyclopedic* (information about the extension of the concept), and *lexical* (information about the natural language counterpart that expresses the concept). As Sperber and Wilson have it, encyclopedic entries of concepts consist of further assumptions subject to the rules set out by the corresponding logical entries. This picture can be made compatible with the present account of concepts if perceptual symbols are allowed in. Indeed, Sperber and Wilson's account of concepts appears to especially lend itself to the view I developed in chapter 4. For the labels or addresses that Sperber and Wilson refer to function in a way compatible with that which I proposed for file labels. As opposed to being forms subject to deductive rules, however, I envisioned labels as lexical items over which natural language can operate to connect, combine, and inform our concepts. The lexical items stored under a concept's address on Sperber and Wilson's account can also be understood as analogous to the linguistic symbols I described.

But even if multimodal perceptual information is allowed to be stored under concepts' addresses, and their account of concepts made compatible with mine, Sperber and Wilson's view of computational cognition would need to be modified, for they understand computational cognition strictly in terms of deduction—processes performed by a "deductive device" that is supposed to "model the system used by human beings in spontaneous inference" (p. 94). Perceptual symbols will not fall under the scope of a concept's logical entries, and more generally, will not be subject to the operations of a so-called deductive device.

This brings us back to the point I initially made, which was that, according to Relevance Theory, relevance is a property of assumptions. We can now see why their view requires this, namely because assumptions are the type of thing that can serve as input to the deductive device they believe to be constitutive of central cognition. And moreover, the deductive device is just the type of thing that can process an assumption within a context to produce contextual effects. Yet, if the foregoing account of concepts and cognition is accepted, Sperber and Wilson's view needs to be adjusted in more than one respect. For there would have to be a way to define relevance as a property of multimodal perceptual information inasmuch as a property of assumptions, but in the absence of a language of thought hypothesis.

Nevertheless, there is something intuitively right about the idea that relevance has something to do with positive cognitive effects. For indeed, it seems as if processing irrelevant information would not generally yield positive cognitive effects. By positive cognitive effects, however, I am not referring to Sperber and Wilson's understanding of it as contextual effects occurring in cognitive systems. Rather, I mean it in their broader and looser sense, generally as cognitive benefits, positive cognitive utility, "epistemic improvement", or a "worthwhile difference" to the cognitive functioning of the individual. But is there a more precise understanding of positive cognitive effect without invoking contextual effects or deductive inference, but that is somehow tied to relevance? Or, what would be better, does the foregoing account of concepts and cognition provide a more precise understanding of positive cognitive effect that is tied to relevance? I think it does.

To see this, let us recall once again the Dretskean theme that has played a role in the present account. As I had argued above (chapter 4, section 4.3), in a Dretskean sense, our concepts are cognitive structures that enable us to extract and exploit certain information in the environment. Learning or enriching a concept provides us with the ability to decode certain aspects of our sensory experience in such a way that we are able to cognitively respond in certain ways which we would not have been able to otherwise. Having the concept DAFFODIL enables one to see a daffodil *as* a daffodil, and thus allows one to have daffodil-thoughts and to cognize daffodil-stimuli in certain ways. Importantly, then, what information one can decode from stimuli crucially depends on what one already knows about the stimuli, or in the terms of the present account, on what and the way in which information is coded in one's concepts and thus conceptualized.

A natural account of relevance follows from this picture in terms of the amount of information received from a source (such as a stimulus). More specifically, the greater the amount of information received, the greater the relevance of that information. For example, suppose that Alice and Bob are on a nature walk. Alice is a botanist, whereas Bob never cared for plant science. As Alice and Bob gaze upon the flora of the forest floor, they both cognitively extract a vast amount of information from their respective perceptual scenes. However, Alice's conceptual knowledge is so rich that she is able to extract more specialized information than Bob does or even can, having to do with the various kinds of plants that they come across. In this way, the perceptual scene carries more information for Alice than for Bob. Of course, they both process the *same* information, but Alice can *cognitively extract* more information. Whereas Bob simply sees a plant, Alice sees Blindwood ivy; whereas Bob simply sees flowers, Alice sees daffodils. Importantly, certain information in Alice's and Bob's perceptual scene is very *relevant* to Alice but not so relevant to Bob. And the information from the perceptual scene that Alice finds relevant is just that information she is able to extract via her conceptual knowledge. On the other hand, such information is not as relevant to Bob because he cannot represent the information in the same ways, since he lacks the conceptual wherewithal to do so.[4]

Therefore, I suggest that the relevance of a stimulus to a given cognitive system (or agent) depends on the amount of information received from that stimulus. Since the amount of information received depends on one's conceptual wherewithal to attend to and code specific information in certain ways, whether and the extent to which something is relevant is dependent on the conceptual content of one's concepts. But this is not the entire story, since relevance will also depend on the context and cognitive task. Suppose, for instance, that both Alice and Bob are botanists, but Alice is interested in finding a rare flower while Bob is interested in seeing a specific species of ivy. Both Alice and Bob can code the same information in the same ways, but because of their different goals and cognitive tasks, Alice will find flower information more relevant than Bob, and Bob will find ivy information more relevant than Alice.

This can be easily accounted for, however, once it is understood that, in setting up one's goals and preparing for one's cognitive task, one requisitely activates a number of concepts with specific conceptualizations and representations. This will in a sense serve as a filtering mechanism for focusing attention. Alice will thus be (cognitively/conceptually) geared to attend to specific information related to a specific flower whereas Bob will be (cognitively/conceptually) geared to attend to specific information related to ivies, as each will have prepared a set of concepts, conceptualizations, and representations (conceptual content) upon embarking on their

[4]In the limit, information is irrelevant if the amount of information received is equal to 0. For this to happen, however, one would have to be unable to extract and conceptualize any information from the source in question (cf. Gabbay & Woods, 2003). Unlike Dretske, however, I do not suppose that a signal carries 0 information if one already knows the message the signal carries. Dretske's example is of a neurotic person who constantly has to check to see if his door is locked. According to Dretske, this neurotic person is receiving no information by checking to see if his door is locked after the first instance. I disagree. Upon checking to see if the door is locked, the neurotic person still receives the information that the door is *still* locked, or that the door *has not come unlocked*. Processing information one already knows may be *redundant*, but it may also be useful to strengthen or reaffirm one's beliefs. Thus, when I say here that information is irrelevant if the amount of information received is equal to 0, I am not claiming that one is processing irrelevant information when the information is what one already knows.

respective cognitive tasks.[5] The conceptual content that gets activated when one prepares for a given cognitive task will tend to be relevant, although this may not always be the case. What gets activated will depend on previous experience, past activations, and the extant relations and connections among the activated conceptual content. Since experience would indicate what was and was not useful in the past, one would tend to bring to bear similar conceptual content to bear on similar problems. And since usefulness is an indicator of relevance, such conceptual content will tend to be relevant, although what gets activated is essentially based on (what might be thought of as) guesses. Nevertheless, this will set up certain relations among the activated concepts and conceptualizations, and these relations will constrain and guide inference, as explained in chapters 4 and 5 (especially with respect to heuristic inference). The relations among activated information can be understood as a context in which information-processing occurs, in a sense similar to Sperber and Wilson's Relevance Theory. But on the view I am suggesting here, the degree to which information is relevant is a matter of the *informativeness* of such information within such a context. Understood this way, relevance is not just a property of information (or assumptions or inputs) *per se*—not just a matter of what information gets processed—but a matter of *how information gets processed*.

If this is right, then positive cognitive effect can be understood in terms of the informativeness of information that is processed—processing information yields positive cognitive effects insofar as the results are informative. But this is just to refer to the degree of relevance of the information in question, or so I am claiming. This means that positive cognitive effect is yielded by processing relevant information. We therefore have our intuitive connection between positive cognitive effect and relevance borne out by the present account: Processed information has positive cognitive effect *because* it is informative; and the degree to which it is informative is the degree to which the information is relevant.

Consider, as a final example, riddles. Riddles present an audience with *all* of the relevant

[5]This is not to rule out, however, a potential role for perceptual mechanisms. It might be the case, for instance, that nonconceptual creatures can focus attention based on simple perceptual pattern-matching. But I suspect that, at least for the human case, one's conceptual wherewithal will have a very big role to play.

information needed to solve them. But what makes riddles hard to solve is that people do not process the information in the right sorts of ways. A person can very well process the information (or assumptions) presented in a riddle with lots of cognitive effort but to little effect. On Sperber and Wilson's Relevance Theory, this would mean that the information is not relevant. But this is clearly not the case. Instead, the person who tries but fails to solve a riddle is not conceptualizing the information in a way that facilitates the correct inference—the person is not bringing to bear on her inferences the right sorts of conceptual information (in terms of perceptual symbol simulations or linguistically coded representations). Hence, what should be relevant is not conceptualized as relevant, and she therefore fails to solve the riddle. This is precisely what the present account claims. And, indeed, once one finally comes around to conceptualizing the information in the right ways, the solution to the riddle appears obvious (cf. my remarks in the previous chapter (sections 5.1.2.3 and 5.1.2.4) regarding having to conceptualize probabilistic information in the right ways in order to facilitate probabilistic inferences and judgments).

Thus, on the foregoing view, it is not that information is relevant because it produces many cognitive effects with little effort, but the other way around—information produces many cognitive effects with little effort because it is relevant (as a function of its informativeness). Cognitive effects and effort therefore do not define relevance. Rather, concepts and the extant relations within and among their conceptual content—i.e., the ER-structure—will facilitate cognitive effects and effort in processing information. Hence, it is structured concepts that deliver relevance, which in turn assists in producing cognitive effects with little effort.

6.3 Do heuristics determine relevance?

As I have been arguing, relevance depends on the conceptual content of and extant relations within and between concepts. But it is a further matter as to how cognition knows what information to code up and integrate in—i.e., what information is relevant to—a concept's simulator or file. In constructing a CAT concept, how does the brain know to code up and integrate the general shape, size, and furriness of cats rather than the way they cast shadows, the range of motion of their left hind legs, or indeed, the position of the moon relative to the earth? That is, how does cognition know what is and is not relevant in building our concepts? To this the foregoing account has no answer, nor pretends to; but this is the hard problem of epistemological relevance.

Evolutionary stories certainly can be told in which the brain evolved to pick out and integrate relevant information. Witness, for instance, Sperber and Wilson:

> As a result of constant selection pressures toward increasing efficiency, the human cognitive system has developed in such a way that our perceptual mechanisms tend automatically to pick out potentially relevant stimuli, our memory retrieval mechanisms tend automatically to activate potentially relevant assumptions, and our inferential mechanisms tend spontaneously to process them in the most productive way. (Wilson & Sperber, 2006, p. 610)

Nevertheless, such evolutionary stories come cheap, and are generally unsatisfactory for informing us on *how* cognition determines relevance.

Yet the epistemological relevance problem expounded in chapter 3 with respect to the frame problem—namely, the problem of how we *know* what is relevant to the task at hand—may be a little easier to tackle. In this regard, let us ask: Can heuristics help in solving this problem in ways that certain other philosophers (e.g. Carruthers, 2006a, 2006c; Samuels, 2005, forthcoming) have failed to consider? More specifically, can the theory of how heuristics work, as developed in this dissertation, solve the epistemological relevance problem? The answer, unfortunately, is "No". In fact, I believe that heuristic solutions fail generally for two reasons. First, of all our epistemic commitments, we in fact almost never *know* the bearing that a given commitment may have to a given cognitive task, at least not *a priori*. This is just to repeat why the epistemological relevance problem is indeed a problem. It is possible that something that presently appears to be relevant to a given task will turn out to be relevant only after the appropriate research is conducted, and once we have sufficiently built up our background knowledge. Louise Antony (2003) provides a neat example:[6] In the 1930s, an archeological dig in New

[6]Thanks to Rob Stainton for informing me of this example.

Mexico revealed a ceramic piece that graphically represented a supernova that occurred in the 11th century. According to Chinese and Japanese astrological records, the supernova in question produced a light visible in Earth's sky which lasted twenty-three days. The representation on the ceramic artifact is of a small circle with twenty-three lines radiating from it. Carbon dating indicates that the artifact originated around the time when the said supernova was to have occurred. According to some experts, this ceramic piece serves as the most certain record of the supernova discovered outside of China and Japan. In this way, then, the artwork on that particular ceramic artifact is actually relevant to astronomy, specifically to constraining astronomical theory. Yet, I believe that this serves to illustrate a broader point that we almost never can know *a priori* what is relevant to our cognitive tasks. The times when we do or can know what is relevant will likely be confined to those instances when the problem is specially circumscribed, or more particularly when the problem is well-defined (see chapter 2, section 2.2). This, it seems, is why science—the paradigmatic example of a holistic enterprise—requires teams of people doing lots of work over long periods of time; and even then, science does not always get everything right. In this sense, I agree with Fodor (2000) that solving the frame problem (or at least the epistemological aspect of concern here) is tantamount to considering the whole background of one's epistemic commitments, for this is the only way to guarantee that one has not missed anything relevant.

The second reason why I think heuristic solutions to the epistemological relevance problem fail is that there is nothing to suggest that heuristics, or cognition generally, is in the business of guaranteeing that nothing relevant has been missed (cf. Samuels, 2005, forthcoming). In other words, there is nothing to suggest that cognition attempts to solve the epistemological relevance problem in the first place. In which case, it is not really a problem; we already know how to solve it, namely the hard way indicated above, notwithstanding that this is not practical in most cases. At the same time, it seems that most of human cognition is content with, and seems to get by on, *satisficing* judgments of relevance. And judgments that are in some sense good enough should not be confused with solving the epistemological relevance problem. Ensuring certainty in our relevance determinations is severely cognitively taxing, requiring time and resources that we simply do not have in managing all of our day-to-day cognitive tasks. Consequently, it does not seem as if complete or holistic relevance is even an issue that deserves serious attention in our cognitive lives unless we are doing high-stake reasoning, such as in science or philosophy. I therefore believe that any heuristic solution to the epistemological relevance problem is empty in virtue of the problem being, by all practical considerations, unsolvable (at least for the majority of interesting tasks), as well as being a problem that we are typically not in the business of solving in the first place. Hence, I disagree with Fodor (2000) that our inability to understand how we solve the frame problem (or at least the epistemological aspect of concern here) is detrimental to computational cognitive science; for indeed, we do not solve the problem, and so Fodor's worries are moot.

But now the question arises: If we do not solve the epistemological relevance problem, then how do we manage our day-to-day tasks and make reasonable decisions? How do we tend to bring to bear what is relevant if we do not know what is in fact relevant? This is just the issue that I had accused Carruthers (2006a, 2006c) and Samuels (2005, forthcoming) to have skirted in their appeal to heuristics to circumscribe what gets considered in our cognitive tasks (chapter 3, section 3.2). But I will not likewise skirt the issue here. For indeed, we humans enjoy reasonable levels of success in our reasoning—as Fodor (2008) states, "witness our not all being dead" (p. 116)—and we must explain this.

We might begin to see how the foregoing account of concepts and cognition can help to explain our characteristic levels of success in our reasoning. We may not be in the business of solving the epistemological relevance problem, but, given the present account, we have more constraints on our reasoning than we know. Specifically, a certain kind of relevance is determined by the extant relations among the information embodied by our concepts, or more particularly the ER-structures described in chapter 4. The kind of relevance I have in mind is a sort of *de facto* relevance (cf. Gabbay & Woods, 2003), in which information appears to be relevant due to the architectural characteristics of cognition. More specifically, and to continue the line of reasoning in the previous section, I propose that something is *de facto* relevant if it is (more or less) informative when processed against a given set of activated concepts (along with their attending conceptualizations and representations). In this way, *de facto* relevance cannot be determined *a priori*, as should be expected. Instead, it simply arises out of the nature and structures of our concepts.

The *de facto* relevance established by the extant relations within and between activated concepts and conceptual content appears to be enough for humans to get by on. Indeed, a theme that has been running throughout this dissertation is that this is precisely what is needed to ensure tractable cognition and, more importantly, to facilitate heuristics. The kind of relevance that matters to the epistemological relevance problem, on the other hand, is an ontologically prior relevance. For lack of a better term, let us call this *objective* relevance. There will certainly be times when we fail to process objectively relevant information, or that we process information that is not very objectively relevant at all. In some cases we may end up processing some objectively *ir*relevant information. Moreover, there will inevitably be cases in which we fail to process *de facto* relevant information, due to cognitive limitations, fatigue, stress, or some other extraneous factor. Under satisfactory conditions, however, our activated concepts, with their conceptual content and extant relations, provide a network that informs cognition of what is *de facto* relevant, and constrains and guides its processing accordingly (sometimes via heuristics). The situation may not be ideal, but it is good enough for us to get by on—indeed, such is to be expected from satisficing organisms. On the other hand, when we enter into certain high-stake arenas, such as science or philosophy, we alter our standards, and *de facto* relevance is no longer good enough. In such circumstances, objective relevance is sought, and again, this is why progress and getting things right are much more difficult to achieve in these endeavours.

At the same time, however, we might understand the foregoing account of concepts and cognition as contributing to how humans manage to solve, not the epistemological relevance problem, but a different guise of the frame problem. This is something I alluded to toward the end of chapter 3 (section 3.4). To see what I mean, let us revisit an issue that I mentioned there. Recall that in discussing the frame problem, Dennett (1984) had presented an example of fixing a midnight snack. He noticed that such a mundane task requires copious amounts of knowledge: "We know trillions of things; we know that mayonnaise doesn't dissolve knives on contact, that a slice of bread is smaller than Mount Everest, that opening the refrigerator doesn't cause a nuclear holocaust in the kitchen" (p. 136). Dennett also noticed that it is implausible to think that such knowledge is stored in a human as separate declarative statements. We must therefore possess a system of representing this knowledge in such a way that it is accessible on demand. Dennett therefore believed that the frame problem for AI researchers is to design a system that is organized in such a way that achieves the efficient representation and access we observe in humans. In his own words: "A walking encyclopedia will walk over a cliff, for all its knowledge of cliffs and the effects of gravity, unless it is designed in such a fashion that it can find the right bits of knowledge at the right times, so it can plan its engagements with the real world" (pp. 140-141); thus,"[the frame] problem concerns how to represent (so it can be used) all that hard-won empirical information . . . . Even if you have excellent knowledge (and not mere belief) about the changing world, how can this knowledge be represented so that it can be efficaciously brought to bear?" (p. 140). Given Dennett's analysis, I suggested that an aspect of the frame problem can be understood as the *representational frame problem*:

The representational frame problem The problem of how a cognitive system embodies the informational organization, and enables access to the relevant information, that seems to be required for human-like cognitive performance.

I believe that the present account of concepts and cognition is precisely what enables humans to solve the representational frame problem. For as I have been at pains to illustrate, concepts are organized in such a way that the extant relations between their conceptual content facilitates access to *de facto* relevant information. Such information may not be objectively relevant, but it will almost certainly be the kind of information that is needed to guide successful action, and this is all that is needed for human-like performance.

It is interesting to note, however, that it seems that much of the information that is *de facto* relevant turns out to be objectively relevant a lot of the time. This is evident from how humans get on in the world, and the success rate of many human inferential endeavours. Again, this is certainly not a result of chance, but likely an outcome of some evolutionary story. Nevertheless, this is what explains our reasonable levels of success in bringing to bear objectively relevant information on our cognitive tasks. I admit that this is not much of an explanation. However, if we conceive of our conceptual system to have evolved to track *things in the world*, then it should not be much of a mystery why our conceptual wherewithal reflects the organized structure of information in the world, including objective relevance relations. In this way, then, *de facto* relevance is built up by systems that track objective relevance. And, just like any cognitive system that tracks stuff in the world, sometimes things work out and sometimes things go awry; and sometimes cognitive systems track truths but not all the time (such as the perceptual systems; cf. Viger, 2006b). It seems, however, that *in the main* cognition tracks truths in the world, and is quite good at it. Thus, the *de facto* relevance embodied by the relations within and between concepts and conceptual content by and large reflects objective relevance in the world. This, I propose, is what allows us to get by in the world, or in other words, what enables *de facto* relevance to be good enough for our purposes.

Assuming that all of this is right, what, then, is the role for heuristics? Heuristics make no determinations of what information is relevant to given cognitive tasks. Rather, relevance is already established in the *de facto* sense. Since heuristics are constrained and guided by what concepts and other attending information is activated, as argued above, and since these things are what determine *de facto* relevance, heuristics are in fact constrained and guided by what is *de facto* relevant. More precisely, heuristics operate over ER-structures which embody implicit information (knowledge) about what is *de facto* relevant. Thus, the proper role of heuristics is to draw inferences based on such structures that determine *de facto* relevance, and I do not think that we should expect (or even desire) more than that.

6.4 Concluding remarks

This concludes this dissertation. Let us briefly summarize: There are many different (and sometimes incompatible) ways that “heuristic” is employed in the philosophical and cognitive science literatures. I suggested that we understand heuristics as cognitive procedures that operate by exploiting (mental) informational structures. The informational structures heuristics exploit are active and primed concepts and their attending information (conceptualizations and representations). Heuristic reasoning is thereby constrained by concepts and their attending information. More specifically, heuristics operate over extant relational structures between concepts and conceptual content that embody implicit, higher-order, metainformational knowledge about the concepts in question vis-à-vis the cognitive task at hand. By operating over such structures, heuristics are enabled to satisfice, and require little cognitive resources for their recruitment and execution. Heuristics are thereby good candidates to mitigate the computational worries arising from the frame problem (i.e., the computational relevance problem), although it is really our conceptual wherewithal that is shouldering the informational burden. Our conceptual wherewithal is also what determines *de facto* relevance, and insofar as it does this, our conceptual wherewithal is what helps us to achieve our characteristic levels of success in making determinations of relevance. The structures that determine *de facto* relevance are those very structures that heuristics operate over. However, notwithstanding certain claims to the contrary, heuristics do not circumvent a different aspect of the frame problem, namely the problem of how a cognitive system *knows* what is (objectively) relevant (i.e., the epistemological relevance problem). Rather, the latter problem is generally not solved, and moreover, human cognition is not in the business of solving it in the first place. But it appears that depending on *de facto* relevance is good enough for human performance.

References

Antony, L. (2003). Rabbit-pots and supernovas: The relevance of psychological evidence to linguistic theory. In A. Barber (Ed.), *The epistemology of language* (p. 47-68). Oxford: Oxford University Press.

Ashcraft, M. H. (2002). *Cognition* (Third ed.). New Jersey: Prentice Hall.

Aspeitia, A. A. B., Eraña Ángeles, & Stainton, R. (2010). The contribution of domain specificity in the highly modular mind. *Minds & Machines*, *20*, 1927.

Bach, K. (1987). *Thought and reference*. Oxford: Oxford University Press,.

Baron-Cohen, S. (1995). *Mindblindness*. Cambridge: The MIT Press.

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

Barsalou, L. W. (2003). Situated simulation in the human cognitive system. *Language and Cognitive Processes*, *18*, 513-562.

Barsalou, L. W. (2005a). Abstraction as dynamic interpretation in perceptual symbol systems. In L. Gershkoff-Stowe & D. Rakison (Eds.), *Building object categories. Carnegie symposium series* (p. 389-431). Majwah, NJ: Erlbaum.

Barsalou, L. W. (2005b). Situated conceptualization. In H. Cohen & C. Lefebvre (Eds.), *Handbook of categorization in cognitive science* (p. 619-650). St. Louis, MO: Elsevier.

Barsalou, L. W. (2008a). Grounded cognition. *Annual Review of Psychology*, *59*, 617-645.

Barsalou, L. W. (2008b). Grounding symbolic operations in the brain's modal systems. In G. R. Semin & E. R. Smith (Eds.), *Embodied grounding: Social, cognitive, affective, and neuroscientific approaches* (p. 9-42). New York: Cambridge University Press.

Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society of London: Biological Sciences*, *364*, 1281-1289.

Barsalou, L. W., Santos, A., Simmons, W. K., & Wilson, C. D. (2008). Language and simulation in conceptual processing. In M. D. Vega, A. M. Glenberg, & A. C. Graesser (Eds.), *Symbols, embodiment, and meaning* (p. 245-283). Oxford: Oxford University Press.

Barsalou, L. W., Simmons, W. K., Barbey, A. K., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences*, *7*, 84-91.

Bechtel, W. (2010). How can philosophy be a true cognitive science discipline? *Topics in Cognitive Sciences*, *2*, 357-366.

Boyd, R. (1991). Realism, anti-foundationalism and the enthusiasm for natural kinds. *Dialectica*, *43*, 1-2,5-29.

Braunstein, M. L. (1994). Decoding principles, heuristics and inference in visual perception. In G. Jansson, S. S. Bergström, & W. Epstein (Eds.), *Perceiving events and objects* (p. 436-446). Hillsdale, NJ: Lawrence Erlbaum Associates.

Brook, A. (2004). Introduction: Philosophy in and philosophy of cognitive science. *Topics in Cognitive Science*, *1*, 216230.

Brown, H. I. (1988). *Rationality*. London; New York: Routledge.

Carey, S. (2004). Bootstrapping and the origin of concepts. *Daedalus*, *Winter*, 59-68.

Carey, S. (2009). *The origin of concepts*. New York: Oxford University Press.

Carruthers, P. (2006a). *The architecture of the mind: Massive modularity and the flexibility of thought*. Oxford: Oxford University Press.

Carruthers, P. (2006b). The case for massively modular models of mind. In R. J. Stainton (Ed.), *Contemporary debates in cognitive science* (p. 3-21). Malden, MA: Blackwell Publishing.

Carruthers, P. (2006c). Simple heuristics meet massive modularity. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The innate mind: Culture and cognition* (p. 181-198). Oxford: Oxford University Press.

Clark, A. (2002). Local associations and global reason: Fodor's frame problem and second-order search. *Cognitive Science Quarterly*, *2*, 115-140.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, *58*, 7-19.

Condry, K. F., & Spelke, E. S. (2008). The development of language and abstract concepts: The case of natural number. *Journal of Experimental Psychology: General*, *137*, 2238.

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (p. 163-228). New York: Cambridge University Press.

Crowley, K., & Siegler, R. S. (1993). Flexible strategy use in young children's tic-tac-toe. *Cognitive Science*, *17*, 531-561.

Damasio, A. R. (1989). Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition*, *33*, 25-62.

Dennett, D. C. (1984). Cognitive wheels: The frame problem of AI. In C. Hookway (Ed.), *Minds, machines, and evolution* (p. 129-152). Cambridge: Cambridge University Press.

Dennett, D. C. (1991). *Consciousness explained*. Boston: Little, Brown and Co.

Dennett, D. C. (1993). Learning and labeling. *Mind & Language*, *8*, 540-547.

Dennett, D. C. (2000). Making tools for thinking. In D. Sperber (Ed.), *Metarepresentation: A multidisciplinary perspective* (p. 17-29). Oxford: Oxford University Press.

Dennett, D. C. (2009). The part of cognitive science that is philosophy. *Topics in Cognitive Science*, *1*, 231236.

Dretske, F. I. (1981). *Knowledge and the flow of information*. Cambridge, MA.: The MIT Press.

Dunbar, K. (1998). Problem solving. In W. Bechtel & G. Graham (Eds.), *A companion to cognitive science* (p. 289-298). Malden, MA: Blackwell Publishing.

Evans, G. (1982). *The varieties of reference*. Oxford: Oxford University Press.

Evans, G. (1985). The causal theory of names. In *Collected papers* (p. 1-24). Oxford: Clarendon Press.

Evans, J. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin & Review*, *13*, 378-395.

Evans, J. (2009). How many dual-process theories do we need? one, two, or many? In J. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (p. 33-54). Oxford: Oxford University Press.

Evans, J., & Frankish, K. (2009). *In two minds: Dual processes and beyond*. Oxford: Oxford University Press.

Evans, J., & Over, D. E. (1996). *Reasoning and rationality*. Hove, U.K.: Psychology Press.

Feldman, J. (1999). Does vision work?: Towards a semantics of perception. In E. Lepore & Z. Pylyshyn (Eds.), *What is cognitive science?* (p. 208-229). Malden, MA: Blackwell Publishers.

Feldman, R. (2008). Naturalized epistemology. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Fall 2008 ed.). http://plato.stanford.edu/archives/fall2008/entries/epistemology-naturalized/.

Fodor, J. A. (1975). *The language of thought*. Cambridge: Harvard University Press.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge: The MIT Press.

Fodor, J. A. (1987). Modules, frames, fridgeons, sleeping dogs and the music of the spheres. In J. L. Garfield (Ed.), *Modularity in knowledge, representation and natural-language understanding* (p. 25-36). The MIT Press. (Reprinted from *The robot's dilemma: The frame problem in artificial intelligence*, p. 139-149, by Z. Pylyshyn, Ed., 1987, Norwood, NJ: Ablex)

Fodor, J. A. (1990). *A theory of content and other essays*. Cambridge: The MIT Press.

Fodor, J. A. (2000). *The mind doesn't work that way: The scope and limits of computational psychology*. Cambridge: The MIT Press.

Fodor, J. A. (2008). *LOT 2: The language of thought revisited*. Oxford: Clarendon Press.

Fodor, J. A., Bever, T. G., & Garrett, M. F. (1974). *The psychology of language: An introduction to psycholinguistics and generative grammar*. New York: McGraw-Hill Companies.

Fodor, J. A., & Pylyshyn, Z. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*, 3-71.

Frankish, K., & Evans, J. (2009). The duality of mind: An historical perspective. In J. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (p. 1-29). Oxford: Oxford University Press.

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, *13*, 187-222.

Gabbay, D., & Woods, J. (2003). *A practical logic of cognitive systems, Volume 1. Agenda relevance: A study in formal pragmatics*. Amsterdam: Elsevier.

Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond 'heuristics and biases'. *European Review of Social Psychology*, *2*, 83-115.

Gigerenzer, G. (1993). The bounded rationality of probabilistic models. In K. I. Manktelow & D. E. Over (Eds.), *Rationality: Psychological and philosophical perspectives* (p. 284313). London: Routledge.

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky. *Psychological Review*, *103*, 592-596.

Gigerenzer, G. (2000). *Adaptive thinking: Rationality in the real world*. New York: Oxford University Press.

Gigerenzer, G. (2001). The adaptive toolbox. In G. Gigerenzer & R. Selten (Eds.), *Bounded rationality: The adaptive toolbox* (p. 37-50). Cambridge: The MIT Press.

Gigerenzer, G. (2004). Fast and frugal heuristics: The tools of bounded rationality. In D. J. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (p. 62-88). Malden, MA.: Blackwell Publishing.

Gigerenzer, G. (2006). Bounded and rational. In R. J. Stainton (Ed.), *Contemporary debates in cognitive science* (p. 115-133). Malden, MA.: Blackwell Publishing.

Gigerenzer, G. (2007). *Gut feelings: The intelligence of the unconscious*. New York: Viking (the Penguin Group).

Gigerenzer, G. (2008a). Moral intuition = fast and frugal heuristics? In W. Sinnott-Armstrong (Ed.), *Moral psychology, volume 2: The cognitive science of morality: Intuition and diversity* (p. 1-26). Cambridge: The MIT Press.

Gigerenzer, G. (2008b). *Rationality for mortals: How people cope with uncertainty*. Oxford: Oxford University Press.

Gigerenzer, G. (2008c). Why heuristics work. *Perspectives on Psychological Science*, *3*, 20-29.

Gigerenzer, G. (2010). Moral satisficing: Rethinking moral behavior as bounded rationality. *Topics in Cognitive Science*, *2*, 528-554.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*, 650-669.

Gigerenzer, G., & Goldstein, D. G. (1999). Betting on one good reason: The Take the Best heuristic. In G. Gigerenzer, P. M. Todd, & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (p. 75-95). New York: Oxford University Press.

Gigerenzer, G., Hertwig, R., van den Broek, E., Fasolo, B., & Katsikopoulos, K. V. (2005). "a 30-percent chance of rain tomorrow": How does the public understand probabilistic weather forecasts? *Risk Analysis*, *25*, 623-629.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve bayesian reasoning without instruction: Frequency formats. *Psychological Review*, *102*, 684704.

Gigerenzer, G., & Hoffrage, U. (1999). Overcoming diffi culties in bayesian reasoning: A reply to Lewis and Keren (1999) and Meilers and McGraw (1999). *Psychological Review*, *106*, 425430.

Gigerenzer, G., & Todd, P. M. (1999). Fast and frugal heuristics: The adaptive toolbox. In G. Gigerenzer, P. M. Todd, & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (p. 3-34). New York: Oxford University Press.

Gigerenzer, G., Todd, P. M., & the ABC Research Group (Eds.). (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.

Gilovich, T., & Griffin, D. (2002). Introduction—heuristics and biases: Then and now. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (p. 1-18). New York: Cambridge University Press.

Goldstein, D. G., & Gigerenzer, G. (1999). The recognition heuristic: How ignorance makes us smart. In G. Gigerenzer, P. M. Todd, & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (p. 37-58). New York: Oxford University Press.

Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, *109*, 75-90.

Gould, S. J. (1991). *Bully for brontosaurus: Reflections in natural history*. New York: W.W. Norton & Company Inc.

Gouteux, S., & Spelke, E. S. (2001). Children's use of geometry and landmarks to reorient in an open space. *Cognition*, *81*, 119-148.

Grice, H. P. (1975). Logic and conversation. In P. Cole & L. Morgan (Eds.), *Syntax and semantics: Speech acts, volume 3* (p. 41-58). New York: Academic Press.

Hempel, C. G. (1966). *Philosophy of natural science.* Englewood Cliffs, NJ: Prentice-Hall.

Hermer, L., & Spelke, E. S. (1994). A geometric process for spatial reorientation in young children. *Nature*, *370*, 57-59.

Hermer, L., & Spelke, E. S. (1996). Modularity and development: The case of spatial reorientation. *Cognition*, *61*, 195-232.

Hermer-Vazquez, L., Moffett, A., & Munkholm, P. (2001). Language, space, and the development of cognitive flexibility in humans: The case of two spatial memory tasks. *Cognition*, *79*, 263-299.

Hermer-Vazquez, L., Spelke, E. S., & Katsnelson, A. S. (1999). Sources of flexibility in human cognition: Dual-task studies of space and language. *Cognitive Psychology*, *39*, 3-36.

Hirschfeld, L. A., & Gelman, S. A. (1994). *Mapping the mind: Domain specificity in cognition and culture*. Cambridge: Cambridge University Press.

Horgan, T., & Tienson, J. (1998). Rules. In W. Bechtel & G. Graham (Eds.), *A companion to cognitive science* (p. 660-670). Malden, MA: Blackwell Publishing.

Hutchinson, J. M., & Gigerenzer, G. (2005). Simple heuristics and rules of thumb: Where psychologists and behavioural biologists might meet. *Behavioural Processes*, *69*, 97-124.

Johnson-Liard, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge: Harvard University Press.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, *58*, 697-720.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (p. 4981). New York: Cambridge University Press.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982a). *Judgment under uncertainty: Heuristics and biases*. Cambridge, U.K.; New York: Cambridge University Press.

Kahneman, D., Slovic, P., & Tversky, A. (1982b). Preface. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (p. xi-xiii). Cambridge, U.K.; New York: Cambridge University Press.

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, *24*, 175-219.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, *3*, 430-454.

Kahneman, D., & Tversky, A. (1973a). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*, 207-232.

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

Kahneman, D., & Tversky, A. (1979). Intuitive prediction: Biases and corrective procedures. *TIMS Studies in Management Science*, *12*, 313-327.

Kahneman, D., & Tversky, A. (1982). On the study of statistical intuitions. *Cognition*, *11*, 123-141.

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review*, *103*, 581-591.

Kaiser, M. (1998). The perception of dynamical constancies. In V. Walsh & J. Kulikowski

(Eds.), *Perceptual constancy: Why things look as they do* (p. 436-454). Cambridge: Cambridge University Press.

Kim, J. (1988). What is naturalized epistemology? In J. E. Tomberlin (Ed.), *Philosophical perspectives, volume 2* (p. 381-406). Asascadero, CA: Ridgeview Publishing Co.

Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the Association for Computing Machinery*, *46*, 604-632.

Kulkarni, D., & Simon, H. A. (1988). The process of scientific discovery: The strategy of experimentation. *Cognitive Science*, *12*, 139-176.

Lakatos, I. (1963-64). Proofs and refutations, I-IV. *British Journal for the Philosophy of Science*, *14, 53-56*, 1-25, 120-139, 220-245, 296-342.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Langley, P., Bradshaw, G., Simon, H. A., & Zytgow, J. (1987). *Scientific discovery*. Cambridge: The MIT Press.

Lawlor, K. (2001). *New thoughts about old things: Cognitive policies as the ground of singular concepts*. New York: Garland Publishing.

McCarthy, J., & Hayes, P. (1969). Some philosophical problems from the standpoint of artificial intelligence. In B. Meltzer & D. Michie (Eds.), *Machine intelligence* (p. 463-502). Edinburgh: Edinburgh University Press.

McDermott, D. (1987). We've been framed: Or, why AI is innocent of the frame problem. In Z. W. Pylyshyn (Ed.), *The robot's dilemma: The frame problem in artificial intelligence* (p. 113122). Norwood, NJ: Ablex Publishing.

Michalewicz, Z., & Fogel, D. B. (2000). *How to solve it: Modern heuristics*. Berlin: Springer.

Minsky, M. (1975). A framework for representing knowledge. In P. H. Wilson (Ed.), *The psychology of computer vision* (p. 211-277). New York: McGraw-Hill.

Mithen, S. (2000). Mind, brain and material culture: An archeological perspective. In P. Carruthers & A. Chamberlain (Eds.), *Evolution and the human mind* (p. 207-217). Cambridge: Cambridge University Press.

Moustakas, C. E. (1990). *Heuristic research: Design methodology, and applications*. Newbury Park: Sage Publications.

Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review*, *90*, 339-363.

Oppenheimer, D. M. (2003). Not so fast! (and not so frugal!): Rethinking the recognition heuristic. *Cognition*, *90*, B1-B9.

Pearl, J. (1984). *Heuristics: Intelligent search strategies for computer problem solving*. Reading, MA: Addison-Wesley Publishing Company.

Pericliev, V. (1990). On heuristic procedures in linguistics. *Studia Linguistica*, *43*, 59-69.

Perry, J. (2001). *Knowledge, possibility, and consciousness*. Cambridge: The MIT Press.

Pólya, G. (1957). *How to solve it: A new aspect of mathematical method*. Garden City, NY: Doubleday Anchor Books.

Popper, K. (1959). *The logic of scientific discovery*. London: Hutchinson.

Prinz, J. J. (2002). *Furnishing the mind: Concepts and their perceptual basis*. Cambridge: The MIT Press.

Prinz, J. J. (2006). Is the mind really modular? In R. J. Stainton (Ed.), *Contemporary debates in cognitive science* (p. 22-36). Malden, MA: Blackwell Publishing.

Pylyshyn, Z. W. (1999). What's in your mind. In E. Lepore & Z. Pylyshyn (Eds.), *What is cognitive science?* (p. 1-25). Malden, MA: Blackwell Publishers.

Pylyshyn, Z. W. (2003). *Seeing and visualizing: It's not what you think*. Cambridge: The MIT Press.

Quine, W. V. O. (1969). Epistemology naturalized. In *Ontological relativity and other essays* (p. 37-48). New York: Columbia University Press.

Richardson, R. C. (1998). Heuristics and satisficing. In W. Bechtel & G. Graham (Eds.), *A companion to cognitive science* (p. 566-575). Malden, MA: Blackwell Publishing.

Samuels, R. (1998). Evolutionary psychology and the massive modularity hypothesis. *The British Society for the Philosophy of Science*, *49*, 575-602.

Samuels, R. (2000). Massive modular minds: Evolutionary psychology and cognitive architectures. In P. Carruthers & A. Chamberlain (Eds.), *Evolution and the human mind* (p. 13-46). Cambridge: Cambridge University Press.

Samuels, R. (2005). The complexity of cognition: Tractability arguments for massive modularity. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The innate mind: Structure and contents* (p. 107-121). Oxford: Oxford University Press.

Samuels, R. (2006). Is the human mind massively modular? In R. J. Stainton (Ed.), *Contemporary debates in cognitive science* (p. 37-56). Malden, MA: Blackwell Publishing.

Samuels, R. (2009). The magical number two, plus or minus: Dual-process theory as a theory of cognitive kinds. In J. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (p. 129-146). Oxford: Oxford University Press.

Samuels, R. (forthcoming). Classical computationalism and the many problems of cognitive relevance. *Studies in History and Philosophy of Science*.

Samuels, R., Stich, S., & Bishop, M. (2002). Ending the rationality wars: How to make disputes about human rationality disappear. In R. Elio (Ed.), *Common sense, reasoning and rationality* (p. 236-268). New York: Oxford University Press.

Samuels, R., Stich, S., & Faucher, L. (2004). Reason and rationality. In I. Niiniluoto, M. Sintonen, & J. Wolenski (Eds.), *Handbook of epistemology* (p. 131-179). Dordrecht: Kluwer.

Santos, A., Chaigneau, S. E., Simmons, W. K., & Barsalou, L. W. (forthcoming). Property generation reflects word association and situated simulation. *Language and Cognition*.

Savage, L. J. (1954). *The foundations of statistics*. New York: Dover Publications Inc., 2nd edition.

Schwarz, N., & Vaughn, L. A. (2002). The availability heuristic revisited: Ease of recall and content recall as distinct sources of information. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (p. 103-119). New York: Cambridge University Press.

Searle, J. (1980). Minds, brains and programs. *Behavioral and Brain Sciences*, *3*, 417457.

Shafer, G. (1986). Savage revisited. *Statistical Science*, *1*, 463-485.

Shanahan, M. (2009). The frame problem. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2009 ed.). http://plato.stanford.edu/archives/win2009/entries/frame-problem/.

Simmons, W. K., & Barsalou, L. W. (2003). The similarity-in-topography principle: Reconciling theories of conceptual deficits. *Cognitive Neuropsychology*, *20*, 451-486.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63*, 129-138.

Simon, H. A. (1957). *Models of man, social and rational: Mathematical essays on rational human behavior in a social setting*. New York: John Wiley.

Simon, H. A. (1973). Does scientific discovery have a logic? *Philosophy of Science*, *40*, 471-480.

Simon, H. A. (1977). *Models of discovery*. Reidel: Dordrecht.

Simon, H. A. (1979). *Models of thought*. New Haven: Yale University Press.

Simon, H. A. (1990). Invariants of human behavior. *Annual Review of Psychology*, *41*, 1-19.

Simon, H. A., & Newell, A. (1958). Heuristic problem solving: The next advance in operations research. *Operations Research*, *6*, 1-10.

Simon, H. A., Newell, A., Minsky, M., & Feigenbaum, E. (1967). Heuristic programs and algorithms. *SIGART Newsletter*, *6*, 10-19.

Spelke, E. S. (2003a). Developing knowledge of space: Core systems and new combinations. In S. M. Kosslyn & A. Galaburda (Eds.), *Languages of the brain* (p. 239-258). Cambridge, MA: Harvard University Press.

Spelke, E. S. (2003b). What makes us smart? core knowledge and natural language. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the investigation of language and thought* (p. 276-311). Cambridge, MA: MIT Press.

Spelke, E. S. (2007). Core knowledge. *Developmental Science*, *10*, 8996.

Spelke, E. S. (2009). Sources of human knowledge. *Jean-Nicod Lectures. Centre National de la Recherche Scientifique*.

Spelke, E. S., Lee, S. A., & Izard, V. (2010). Beyond core knowledge: Natural geometry. *Cognitive Science*, *34*, 863884.

Sperber, D. (1994). The modularity of thought and the epidemiology of representations. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (p. 39-67). Cambridge: Cambridge University Press.

Sperber, D. (1996). *Explaining culture: A naturalistic approach*. Oxford: Blackwell.

Sperber, D. (2005). Modularity and relevance: How can a massively modular mind be flexible and context-sensitive? In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The innate mind: Structure and contents* (p. 53-68). Oxford: Oxford University Press.

Sperber, D., & Wilson, D. (1982). Mutual knowledge and relevance. In N. V. Smith (Ed.), *Mutual knowledge* (p. 61-85). New York: Academic Press.

Sperber, D., & Wilson, D. (1986/1995). *Relevance: Communication and cognition*. Oxford: Blackwell (2nd edition, 1995).

Sperber, D., & Wilson, D. (1995). Postface to the second edition of *Relevance: Communication and Cognition*.

Sperber, D., & Wilson, D. (1996). Fodor's frame problem and relevance theory (reply to Chiappe & Kukla). *Behavioral and Brain Sciences*, *19*, 530-532.

Sterelny, K. (2003). *Thought in a hostile world*. Malden, MA: Blackwell.

Sterelny, K. (2004). Externalism, epistemic artifacts and the extended mind. In R. Schantz (Ed.), *The externalist challenge: New studies on cognition and intentionality* (p. 239-254). Berlin; New York: Walter de Gruyter.

Sterelny, K. (2006). Cognitive load and human decision, or, three ways of rolling the rock uphill. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The innate mind: Culture and cognition* (p. 218-233). Oxford: Oxford University Press.

Todd, P. M., & Gigerenzer, G. (2000). Précis of *Simple Heuristics that Make Us Smart*. *Behavioral and Brain Sciences*, *23*, 727-780.

Treisman, A. (1982). Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology: Human Perception and Performance*, *8*, 194-214.

Treisman, A., & Schmidt, H. (1982). Illusory conjunction in the perception of objects. *Cognitive Psychology*, *14*, 107-142.

Turing, A. M. (1936-7). On computable numbers: With an application to the entscheidungsproblem. *Proceedings of the London Mathematical Society*, *42*, 230-265. (Correction *ibid. 43*, 544-546 (1937).

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*, 327-352.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *New Series, 185*, 1124-1131.

Tversky, A., & Kahneman, D. (1980). Causal schemas in judgments under uncertainty. In M. Fishbein (Ed.), *Progress in social psychology* (p. 4972). Hillsdale, NJ: Erlbaum.

Tversky, A., & Kahneman, D. (1982a). Availability: A heuristic for judging frequency and probability. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (p. 164-178). Cambridge, U.K.; New York: Cambridge University Press. (Reprinted from *Cognitive Psychology*, 1973, *4*, 207-232)

Tversky, A., & Kahneman, D. (1982b). Judgments of and by representativeness. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (p. 84-98). Cambridge, U.K.; New York: Cambridge University Press.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, *90*, 293-315.

Viger, C. (2005). Learning to think: A response to the *Language of Thought Argument* for innateness. *Mind & Language*, *20*, 313-325.

Viger, C. (2006a). Frame problem. In K. Brown (Ed.), *Encyclopedia of language and linguistics* (p. 610-613). Oxford: Elsevier.

Viger, C. (2006b). Is the aim of perception to provide accurate representations? a case for the "no" side. In R. J. Stainton (Ed.), *Contemporary debates in cognitive science* (p. 275-288). Malden, MA: Blackwell Publishing.

Viger, C. (2006c). Symbols: What cognition requires of representationalism. *Protosociology: The International Journal of Interdisciplinary Research*, *22*, 40-59.

Viger, C. (2007). The acquired language of thought hypothesis: A theory of symbol grounding. *Interaction Studies*, *8*, 125-142.

von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press.

von Plato, J. (1994). *Creating modern probability*. Cambridge: Cambridge University Press.

Wang, R. F., Hermer-Vazquez, L., & Spelke, E. S. (1999). Mechanisms of reorientation and object localization by human children: A comparison with rats. *Behavioral Neuroscience*, *113*, 475-485.

Wason, P. C., & Evans, J. (1975). Dual processes in reasoning? *Cognition*, *3*, 141-154.

Wilson, D., & Sperber, D. (2006). Relevance theory. In L. R. Horn & G. Ward (Eds.), *The handbook of pragmatics* (p. 607-632). Malden, MA: Blackwell Publishing.

Wittgenstein, L. (1953). *Philosophical investigations. The German text, with a revised English translation.* (G. E. M. Anscombe, Trans.). Oxford: Blackwell. (2001).

Wynn, K. (1990). Children's understanding of counting. *Cognition*, *36*, 155-193.

Wynn, K. (1992). Children's acquisition of the number words and the counting system. *Cognitive Psychology*, *24*, 220-251.

Curriculum Vitae

Name

Sheldon J. Chow

Education

Ph.D., Philosophy. The University of Western Ontario, London, Ontario. Spring 2011.
Dissertation Title: Heuristics, Concepts, and Cognitive Architecture: Towards Understanding How The Mind Works
Advisory Committee: Christopher Viger (supervisor), Robert Stainton

M.A., Philosophy. Dalhousie University, Halifax, Nova Scotia. Fall 2006.
M.A. Thesis Title: Zeno's Paradoxes and Problems with Physical Infinities
Supervisor: Robert Martin

B.A. (Honours, with Great Distinction), Philosophy. University of Lethbridge, Lethbridge, Alberta. Spring 2004.
textbfUndergraduate Honours Thesis Title: To Be a Fly on the Wall in The Quantum World (*On Realism in Quantum Mechanics*)
Supervisor: Kent Peacock

Publications

William Harper, **Sheldon J. Chow**, and Gemma Murray. "Bayesian Chance". To appear in *Synthese*.

Presentations

"What Cognitive Science Can Offer Social Science: Lessons from Practical Ethics". *Workshop on Cognitive Social Sciences: Grounding the Social Sciences in the Cognitive Sciences, at CogSci2010: The Annual Meeting of the Cognitive Science Society.* Portland, Oregon. August 2010.

"The Frame Problem, Heuristics, and A New Problem for Abduction". *The Canadian Philosophical Association.* Concordia University, Montréal, Quebec. May 2010.

"Keeping Darwin in Mind". [Co-authored with Alain Ducharme.] *The Canadian Society for the History and Philosophy of Science.* Concordia University, Montréal, Quebec. May 2010.

"Why Practical Ethics Should be Interested in Cognitive Science". *The Canadian Society for* *the Study of Practical Ethics.* Concordia University, Montréal, Quebec. May 2010.

"Keeping Darwin in Mind". [Co-authored with Alain Ducharme.] *150 Years After* Origin*: Biological, Historical and Philosophical Perspectives.* Institute for the History and Philosophy of Science and Technology. University of Toronto, Toronto, Ontario. November 2009.

"Decision Making and Bioethics". *The Canadian Philosophical Association.* Carleton University, Ottawa, Ontario. May 2009.

"Not Another Essay on Modularity of Mind: On the Possibility of Central Systems". *City University of New York 12th Annual Graduate Student Philosophy Conference.* City University of New York, New York City. April 2009.

"Newton's Theory of Universal Gravitation, the Cause of Gravity, and the General Scholium". *The Canadian Society for the History and Philosophy of Science.* University of British Columbia, Vancouver, British Columbia. June 2008.

"Objective Chance Need Not Guide Rational Subjective Credences: Against Lewis's Principal Principle". *The Philosophy Graduate Student Association Colloquium Series.* The University of Western Ontario, London, Ontario. November 2007.

"The Principal Principle: Who Needs It?". *The Canadian Philosophical Association.* University of Saskatchewan, Saskatoon, Saskatchewan. May 2007.

"The Principal Principle: Who Needs It?". *University of Waterloo Philosophy Graduate Student Association Conference.* University of Waterloo, Waterloo, Ontario. March 2007.

"A Quantum Logical Approach". *Griffith University Centre for Quantum Dynamics Seminar Series.* Griffith University, Brisbane, Queensland, Australia. March 2005.

Scholarships and Awards

Scholarships

⋄ Ontario Graduate Scholarship, sponsored by the Provincial Government of Ontario, 2009-2010 & 2010-2011; value: $5,000 per academic term

⋄ Canada Graduate Scholarship - Master's, sponsored by the Social Sciences and Humanities Research Council of Canada, 2005-2006; value: $17,500

⋄ Louise McKinney Scholarship, sponsored by Alberta Scholarship Programs, 2000, 2001 and 2002; value: $2,500 each

⋄ Jason Lang Memorial Scholarship, sponsored by Alberta Scholarship Programs, 2000; value: $1,000

⋄ Alexander Rutherford Scholarship, sponsored by Alberta Scholarship Programs, 1999; value: $2,500

Awards

⋄ Chinook Summer Research Award, sponsored by the University of Lethbridge Board of Governors, 2002 and 2003; value: $4,000 each

Academic Work Experience

Instructorships

◇ The University of Western Ontario
- PHIL 2254: Introduction to Logic — Winter 2011 (King's University College)
- PHIL 1020: Introduction to Philosophy — Fall 2010 and Winter 2011
- PHIL 1300: Introduction to Philosophy — Fall 2008 and Winter 2009

Tutorial Instructorships

◇ The University of Western Ontario
- PHIL 1200: Critical Thinking — Fall 2006 and Winter 2007
- PHIL 020: Introduction to Philosophy — Fall 2007 and Winter 2008

◇ University of Lethbridge
- LIBED 1001: Liberal Education — Fall 2003 and Spring 2004

Teaching Assistantships

◇ Dalhousie University
- PHIL 2805: Ethics in Health Care — Fall 2005
- LOGIC 2130: Deductive Logic — Spring 2006

◇ University of Lethbridge
- PHIL 1000: Introduction to Philosophy — Fall 2001, Fall 2002, Fall 2003
- LOGIC 1000: Introduction to Logic — Fall 2001, Fall 2002
- PHIL 3406: Business Ethics — Spring 2002, Fall 2002
- PHIL 2220: Philosophy of Mind — Spring 2003
- LOGIC 2003: Symbolic Logic — Spring 2004

Research Assistantships

◇ Griffith University (Centre for Quantum Dynamics), Brisbane, Queensland, Australia
- Professor Howard Wiseman; Foundations of Physics — Feb. 2005 - May 2005

◇ University of Lethbridge, Lethbridge, Alberta
- Professor Peter Alward; Philosophy of Language, speech act theory — Summer 2003
- Professor Peter Alward; Philosophy of Mind, mental causation — Summer 2002