ABSTRACT

Title of dissertation: FRAME PROBLEMS, FODOR'S CHALLENGE, AND PRACTICAL REASON

Erich C. Deise, Doctor of Philosophy, 2008

Dissertation directed by: Professor Peter Carruthers
Department of Philosophy

By bringing the frame problem to bear on psychology, Fodor argues that the interesting activities of mind are not amenable to computational modeling. Following exegesis of the frame problem and Fodor's claims, I argue that underlying Fodor's argument is an unsatisfiable normative principle of rationality that in turn commits him to a particular descriptive claim about the nature of our minds. I argue that the descriptive claim is false and that we should reject the normative principle in favor of one that is at least in principle satisfiable. From this it follows, I argue, that we have no reason for thinking the activities of our minds to be, as a matter of principle, unmodelable. Drawing upon Baars' Global Workspace theory, I next outline an alternative framework that provides a means by which the set of engineering challenges raised by Fodor might be met. Having sketched this alternative, I turn next to consider some of the frame problems arising in practical reason and decision-making. Following discussion of the nature of emotion and its influence on practical reason and decision-making, I argue that consideration of emotion provides *one* means by which we might contend with some of the frame problem instances that arise in that domain.

FRAME PROBLEMS, FODOR'S CHALLENGE, AND PRACTICAL REASON

By

Erich C. Deise

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirement for the degree of
Doctor of Philosophy
2008

Advisory Committee:

Professor Peter Carruthers, Chair
Professor Christopher Cherniak
Professor Michael Dougherty
Professor Jeff Horty
Professor Paul Pietroski

DEDICATION

For my family, old and new.

To my parents Jerry and Sue, wife Alexis,
and son Rainer.

ACKNOWLEDGMENTS

I am particularly indebted to Peter Carruthers, whose insights, mentoring and advice have been invariably helpful and whose patience, support and encouragement have been nothing less than supererogatory. I have profited immensely from our many discussions and from the detailed comments provided by him throughout the process. Over the years, Chris Cherniak has been a source of much insight and guidance, as well. I have both enjoyed and benefitted greatly from our many discussions and thank him for, so long ago, setting me down the path that led to this essay. Both are owed special thanks for all that they have done.

I am also grateful to Jeff Horty both for helping me to better understand the formal intricacies of the planning problem and for his commentary on Chapter 1. Likewise, I would like to thank Paul Pietroski for helping me to better understand a particular facet of Fodor's argument and for his willingness to serve, on short notice, as a member of the examining committee. Thanks as well to William Gasarch of the University of Maryland Department of Computer Science both for helping me to understand some of the finer points of computational complexity theory and for assisting me in locating a version of the problem within the complexity hierarchy. I am also grateful to Georges Rey for providing extensive commentary on a number of chapters. While he endorses few, if any, of my conclusions, his suggestions and objections were invaluable in helping me to better structure the project. I also wish to thank Patricia Greenspan both for her guidance early on and for her comments on several early draft chapters.

I have benefitted from a number of discussions with other philosophers, computer scientists and cognitive neuroscientists. Among these, I would like to thank Michael Morreau, Scott James and Sotiros Kotsanis, of the University of Maryland philosophy department, David Poeppel of the NACS program, Ernie Davis of the Department of Computer Science at New York University as well as audiences at the University of Maryland. Philip Ryan deserves thanks for taking on the onerous and arduous task of proofreading the submission draft.

Finally, I am, in every way, indebted to my parents Jerry and Sue and my wife Alexis for everything else.

TABLE OF CONTENTS

INTRODUCTION

Drawing upon the frame problem of A.I., Fodor argues, quite famously, that the "interesting" activities of mind are not amenable to computational modeling. I will argue that we have no reason for thinking Fodor to have made the case for his pessimistic conclusion with respect to the unmodelability of *our* minds. Furthermore, I suggest that there are plausible alternatives available that once explicated and brought to bear might provide a means by which at least some of the "interesting" operations of mind might come to be modeled in computationally feasible terms.

In Chapter 1, I provide an overview of the frame problem. Both the original formal incarnation of problem that arose in artificial intelligence and various philosophical interpretations or variants of the problem are reviewed. I next set out the underlying structure of the puzzle, arguing that "the" frame problem is not one problem but is instead a constellation of related problem instances.

In Chapter 2, I consider in greater detail Fodor's version of the frame problem as well as his application of it to computational psychology. In presenting Fodor's argument for his pessimistic conclusion - that the interesting activities of mind are not amenable to computational modeling – I disentangle a number of distinct argument strains. In so doing, I argue that Fodor's argument relies, at base, upon a particular *normative* principle of rationality. This normative claim, I argue in turn, commits Fodor to a particular *descriptive* claim about the nature of *our* cognitive processes.

In Chapter 3, I focus on assessing Fodor's argument. Briefly, Fodor argues that our cognitive processes are unmodelable because we at least sometimes arrive at conclusions rationally. Since anything that arrives at conclusions rationally, he continues, must be capable of solving the frame problem and no system that solves that problem can be modeled, it follows, he concludes, that *our* cognitive processes are not amenable to modeling.

I argue that the normative principle of rationality undergirding Fodor's argument is untenable because it is in principle unsatisfiable. I next argue that the descriptive claim – that *we* ever arrive at conclusions rationally (in Fodor's sense) – is false. Taken together, I conclude that it is more reasonable to think that we should reject Fodor's normative rationality principle as overly demanding in favor of one that is, at least in principle, satisfiable by the likes of us. Any weakening of the demands of this rationality principle, I argue, effectively undermines Fodor's argument for the pessimistic conclusion. Specifically, by weakening the demands of the normative rationality principle, I argue that Fodor's arguments against the feasibility of the massive modularity of mind and heuristics approaches to modeling are no longer compelling. And so, I conclude that since we have no reasons for rejecting these approaches *tout court*, as Fodor suggests due to their "irrationality," we have no reason of principle for thinking the operations of mind not amenable to modeling.

There are, however, a number of in practice (as opposed to in principle) challenges raised by Fodor with which any proposed strategy for modeling the operations of mind must contend. To this end, I will consider an alternative model – namely Baars' Global Workspace framework. I will start by setting out Baars' account and turn next to considering how bringing this approach to bear might help in contending with some of the particular engineering challenges (*e.g.,* the input routing and heuristic selection problems) raised by Fodor. By way of further fleshing out Baars' proposal, I will consider both Gigerenzer's "fast and frugal" heuristic approach and Barrett's discussion of modules as metaphorical enzymatic systems. When these accounts are integrated into Global Workspace theory, a picture begins to emerge as to how a computationally modelable system might contend with the input routing and heuristic selections problems without the need for a central (and thus frame-problem-infected) executive to be posited. And so, when fleshed out, I suggest that Global Workspace theory might provide a plausible account, and an alternative unconsidered by Fodor, of how at least some of the "interesting" activities of mind might come to be modeled.

In Chapters 4 and 5, I consider in greater detail some instances of the frame problem that arise in practical reason and decision-making, ultimately suggesting that emotion might be *one* means by which we contend with some of the problems arising in this domain. I begin this discussion, however, by setting out a number of prima facie objections to the very idea of bringing emotion to bear on the frame problem.

The first objection, the "irrelevance" objection, reduces to the claim that since *all* emotions are either belief-identical or belief-dependent and since (as Fodor argues) all instances of belief-fixation are infected by/with frame problems, then bringing emotion to bear on problems that arise elsewhere would serve *only* to import a frame-problem ridden operation into some other domain. As such, the objection continues, the very idea that bringing emotion to bear might help us understand how it is that we might contend with some of the frame problems that arise in practical reason and decision-making is an obvious non-starter.

In response to this concern I begin by providing an overview of the propositional attitude approach to emotion in philosophy and its counterpart cognitive appraisal theory in psychology, both of which maintain that emotions are belief-identical or belief-dependent. Following a review of some of relevant empirical evidence in support of the conjecture, I outline a set of standard puzzles for this approach. Having outlined this, I argue that given the inability of such "cognitivist" accounts to contend with the set of puzzles discussed and given the problematic nature of much of the empirical evidence relied upon in support of cognitive appraisal theory, we have no reason for thinking that *all* emotions are belief-identical or belief-dependent. I will argue then that with respect to the irrelevance objection, if at least some of what the emotions are are belief-independent, then there should be no reason to think that *these* should, when brought to bear, serve only to import frame problems from one domain into another.

Next, I consider the support for automated appraisal theories of emotion. In contrast to the cognitivist account, automated appraisal theory maintains that emotions are undertaken by innate, automated, autonomous and plausibly modularly realized learning and appraisal operations – the "basic emotion" or "affect programs." Following a review of the evidence in support of this account, I argue that while automated appraisal theory fails to provide a complete answer to the question of "what the emotions are" it does provide a partial response to this question. That is, we have reason, or so I will argue, for thinking that *at least some* of what the emotions are are undertaken by automated, autonomous and modularly realized learning and appraisal operations. As modularly realized, these activities should be amenable to modeling. If so, then if emotion does influence and inform practical reason and decision-making, we have no reason to think that it should serve *only* to import prior frame problem infected operations into this other domain. It remains, of course, to be seen whether emotion actually informs practical reason and decision-making in ways relevant to helping us contend with the frame problems that arise in that domain.

In Chapter 5, I consider the second of two broad challenges to the idea of bringing emotion to bear. This objection reduces to the claim that even if emotion is (even in part) undertaken by operations that are uninfected by the frame problem and even if emotion actually does influence practical reason and decision-making, we have no reason to think that these influences should be beneficial. At worst, the influences of emotion would be detrimental to the proper operations of practical reason and decision-making and at best, these influences would afford on balance no more benefit than detriment. And so, the objection continues, even if emotion does influence practical reason and decision-making and does so without importing frame problems into this domain, we have no reason to think that such influences should help in any way to expedite or increase the accuracy of the these operations (and at least prima facie reason to think them detrimental or neutral).

With respect to this second broad challenge, I set about in Chapter 5 providing reasons for thinking that emotion might help us in contending with some of the frame problems that arise in practical reason and decision-making by helping to both expedite and increase the accuracy of these operations (*i.e.,* helping us to contend with *both* horns of the dilemma). To this end, I begin by presenting an overview of some representative empirical findings and detailed discussion of Bechara & Damasio's somatic marker hypothesis. Having set out the general claims of Bechara & Damasio's account, I integrate this model with the existing Global Workspace framework discussed in Chapter 3. When integrated, I will argue that emotion might be *one* means by which we contend with *some* of the frame problems that arise in practical reason and decision-making. Specifically, I will argue that emotion appears to influence and inform practical reason and decision-making in ways relevant to helping us to contend with the attentional direction, problem-sequencing, ends-selection, meta-planning and Hamlet's problem instances of the frame problem.

I next argue that the emotion-based heuristics outlined are simple and noncompensatory (relying on only one *cue*). As such, we have every reason to think, or so I will argue, that these heuristics should be far less computationally burdensome and thus far more expeditious than the "rational" alternative. From this, I conclude that bringing emotion to bear on these instances of the problem should help us in contending with the tractability horn of the dilemma posed by the frame problem.

I next consider the question of how normatively "good" we might expect such heuristics to be. I begin by noting a prima facie objection with respect to the normative goodness of any simple heuristic. Drawing upon discussion by Gigerenzer, I argue that we have no reason for thinking that just because an heuristic is simple that it must necessarily also be inaccurate. That is, we have no reason to think that all "quick" heuristics must be "dirty" (*i.e.,* inaccurate). As such, I conclude we have no reason in principle for thinking that the emotion-based heuristics proposed must be too inaccurate to satisfy a suitably weakened normative rationality principle. Rather, following further discussion of Bechara and Damasio's findings I argue that the emotion-based heuristics proposed increase the accuracy of practical reason and decision-making and are thus sufficiently normatively "good." As such, we have reason, I conclude, for thinking that these heuristics should satisfy (a suitably weakened) normative rationality principle and thus the second horn of the dilemma posed by the frame problem.

I then argue that at least some of what the emotions are are undertaken by automated, autonomous and plausibly modularly realized operations of emotional learning and appraisal. Furthermore, I will argue that we have reason to think that these inform and influence practical reason and decision-making in ways that appear to help us in contending with *both* horns of the dilemma posed (*i.e.,* both in expediting and increasing the accuracy of these operations). It follows from this, I argue, that bringing emotion to bear on the puzzle might provide *one* promising approach (of likely a number of others) by which to come to understand how we contend with some of the frame problems that arise in the practical domain.

Finally, I consider a set of open questions. I argue that emotion is quite automatically exploited to influence and direct manifold operations, including those of memory encoding and search, non-trivial planning and assent or opinion fixation. These automated influences, I suggest, should help to expedite the operations of memory encoding, memory search, complex and multi-stage planning and opinion-fixing. In any case, relying on these heuristics, I argue, should be far less computationally burdensome and thus far more expeditious than the rational alternative. I provide reason for thinking that relying on these heuristics should be a sufficiently normatively "good" strategy as well - thus helping to increase the accuracy of the operations. However, while I provide reason for thinking this likely, there is, as of yet, no direct empirical evidence bearing on the question. As such, the question must remain open.

It should be noted at the outset that I am not claiming to have found a way to "solve outright" the frame problem – whatever that might mean. Nor am I suggesting that emotion "solves" the frame problem. I am not providing a complete computational model of mind, nor even an exhaustive account of all of the ways that emotion might influence cognition. Rather, my aims are modest. My first aim to provide reason for thinking Fodor's pessimism to be unwarranted insofar as we have no reasons for thinking that the operations of mind are necessarily mysterious and unmodelable. Nor do we have reason for thinking that the modules and heuristics approaches to modeling should be rejected *tout court*. Put somewhat differently, I argue that nothing could solve the frame problem as interpreted by Fodor. This suggests, of course, that we don't solve *that* problem either. If so, then we must be contending with the problem in some other way. I argue that we might successfully do so by exploiting heuristics. My second aim is to suggest that *one* of the ways that we contend with some of the instances of the frame problem that arise in practical reason and decision-making is by heuristically exploiting emotion. And so, I am claiming only that consideration of emotion helps us understand how we might contend with *some* instances of the frame problem that arise in practical reason and decision-making.

CHAPTER 1: INTRODUCING THE FRAME PROBLEM

Suppose that we consider a few objects arranged on the desk in front of us. For simplicity, only four objects will be considered: a coffee cup, a book, a note pad and a magazine. As things stand the coffee cup is on the desk and has nothing on top of it. The note pad, which has nothing on top of it, is resting atop the book, which is itself only on top of the desk. Finally, the magazine is resting on the desktop and has nothing on top of it. Pictorially the scene looks something like this:

	Note Pad	
Coffee cup	Book	Magazine

Suppose now that we are set the task of making it so that the book is on top of the magazine (perhaps we want to jot down a note or tidy the desk). That is, we want to the resultant situation or scene to look something like this:

		Book
Coffee cup	Note Pad	Magazine

Determining a sequence of actions (*i.e.,* a plan) that will rearrange the contents of the desktop to the above configuration is a rather trivial task. Intuitively, one need only unstack the note pad from the book and place the book on top of the magazine. Suppose however, that one wished to treat the situation and task more formally. For example, suppose that one wished to fully automate the process so that it might be accomplished by a machine (*i.e.,* a robot).[1]

[1] The following presentation is drawn from various formal discussions of the frame problem. Most notably the accounts presented by McCarthy & Hayes (1969), Janlert (1987), Shanahan (1997), Morgenstern (1996), Horty (2001) and Russell & Norvig (1995). Most commentators, in explicating the formal structure of the puzzle, relate strikingly similar sets of axioms, frame axioms and situations and rely on a seemingly standard "blocks world" presentation from which the above example is drawn.

THE SITUATION CALCULUS

In following McCarthy and Hayes, let us attempt to provide a more formal treatment of the problem by introducing the apparatus of the *situation calculus* (SC). Before so doing, however, some terminology must be introduced. The ontology of the situation calculus includes (1) *situations* defined as "the complete state of the universe at an instant of time," (McCarthy & Hayes, 1969 p.18) (2) *fluents*, which are defined as a "function whose domain is the space of situations," (McCarthy & Hayes, 1969 p.19) and which are perhaps more readily understood to be the situation-dependent properties of objects, and (3) *actions*, the results of which alter the values of *fluent*. The language of the situation calculus includes the predicate *Holds* and the function *Result*. (McCarthy & Hayes, 1969 p.22) That a particular facts obtains in a situation *s* is articulated by the expression *Holds* (ϕ, *s*). Likewise, *Result* (α, *s*) expresses the situation resulting from the execution of an action (action-sequence or plan) α in situation *s*.

In order to provide a formal description of our toy situation outlined above, two fluents need to be introduced. Let *On*(x,y) express that object x is on top of object y, and let *Clear*(x) express that object x is free to have something stacked on top of it. And so, the following provides a minimal description of the initial state of our desk.

THE INITIAL SITUATION s_0

Holds[*On*(Cup, Desk), s_0]
Holds[*On*(Magazine, Desk), s_0]
Holds[*On*(Book, Desk), s_0]
Holds[*On*(Note Pad, Book), s_0]

Holds[*Clear*(Cup), s_0]
Holds[*Clear*(Note Pad), s_0]
Holds[*Clear*(Magazine), s_0]
Holds[*Clear*(Desk), s_0]

ACTIONS

Some formal account much also be given of the actions of moving one object from atop another and of moving one object onto another. While this could be accomplished by using only one action *Move* (x,y) to express both stacking and unstacking, discussion will be far clearer if two distinct actions are allowed.[2] Let *Unstack*(x,y) express the action of unstacking object x from atop object y and let *Stack*(x,y) express the action of stacking object x on y. Given the actions of *Stack* and *Unstack,* the consequences of the execution of these actions may be expressed by the following axioms.

The *Stack* axiom:

$\forall$x,y,s(*Holds*[*Clear*(x),s] & *Holds*[*Clear*(y),s] & ~(x=y)) $\rightarrow$ *Holds*[*On*(x,y), *Result*(*Stack* (x,y),s)]

> If, in the initial situation, it is the case that object x has nothing on top of it, object y has nothing atop it, and the objects are distinct, then the consequence of stacking object x upon object y will be that object x is on top of y.

The *Unstack* axiom:

$\forall$x,y,s(*Holds*[*On*(x,y),s] & *Holds*[*Clear*(x),s] & ~(x=y)) $\rightarrow$ *Holds*[*Clear*(y), *Result*(*Unstack* (x,y),s)]

> If, in the initial situation, it is the case that the object x in on y and there is nothing on top of x, then the consequence of unstacking object x from object y is that there will be nothing atop y.

THE PLANNING PROBLEM

With the above in place, let Γ be composed of: 1) the seven sentences describing the initial situation s_0 and 2) the above axioms describing the consequences of stacking and unstacking objects. Further, let ϕ represent the goal state or proposition. In this

[2] That is, Horty's (2001) presentation of the problem in which two actions *stack* and *unstack* are used in place of Shanahan's (1997) more concise but less accessible *move* action.

case, ϕ would be the aim of having it be the case that the book is on top of the magazine - *Holds*[*On*(Book, Magazine)]. The *planning problem*, Horty explains, is the "problem of finding an action sequence α whose execution in the initial state *s* can be proved from the information in Γ to yield a state in which the goal proposition ϕ holds."(Horty, 2001 p. 339) That is, the planning problem is the problem of determining a sequence of actions – in this case the execution of a sequence of actions of unstacking and stacking on the initial and subsequent situations for which it can be established that $\Gamma \vdash$ *Holds*[ϕ, *Result*(α, s)].

At the beginning of the discussion, when presented with the task of rearranging the items on the desk, we were able to determine a simple sequence of actions α that would result in the desired goal state ϕ (the book being on the magazine). Specifically, by unstacking the note pad from the book and then stacking the book on the magazine the desired arrangement is achieved. More formally, we conjectured that the following would hold

Holds[*On*(Book, Magazine), *Result*(<*Unstack*(Note Pad, Book),*Stack*(Book, Magazine)>, s_0)]

Given the discussion of the planning problem, it should be possible to determine the feasibility of such a plan by establishing that the above sequence of actions and subsequent results are derivable from the contents of Γ. That is, it should be possible to establish that:

$\Gamma \vdash$ *Holds*[*On*(Book, Magazine), *Result*(<*Unstack*(Note Pad, Book),*Stack*(Book, Magazine)>, s_0)]

A serious difficulty now arises. Γ, given its contents, cannot yield such a result. From the contents of Γ, specifically, that in the initial situation both *On*(Note Pad, Book) and *Clear*(Note Pad) hold, and equipped with the *unstack* axiom it can be established that the consequence of unstacking the note pad from the book in the initial situation will result in the book being clear. (Horty, 2001) Formally, Γ |– *Holds*[*Clear*(Book), *Result*(*Unstack*(Note Pad, Book), s_0)]

That Γ contains the sentences *Holds*[*Clear*(Cup) s_0] and *Holds*[*Clear*(Magazine) s_0], when taken in conjunction with the fact presented above (that unstacking the note pad from the book will result in the book being clear) appears to provide us with sufficient resources to achieve the goal state of *Holds*[*On*(Book, Magazine)]. And so, given that it is already known that the coffee cup and magazine are clear and that the unstack axiom establishes the fact the after unstacking the book is clear, we appear to be but one stack away from realizing our goal. Since we need only stack the book on top of the magazine to complete the task, it should be possible to use the stack axiom to establish that:

Holds[*On*(Book, Magazine), *Result*(*Stack*(Book, Magazine), *Result*(*Unstack*(Note Pad, Book), s_0))]

And so, it should be possible to establish that the consequence of unstacking the note pad from the book and stacking the latter on the magazine will result in the book being on the magazine. A difficulty arises McCarthy and Hayes note, when we notice that while the magazine was known to be clear in the initial situation s_0 the system does not know that it remains clear in the state, call it s_1 which results from the unstacking of the note pad from the book in s_0. For ϕ to be satisfied would require that the system be capable of determining that the magazine remains clear in the state following the unstacking of the note pad from the book. More formally, it needs to be established that:

Holds[*Clear*(Magazine), *Result*(*Unstack*(Note Pad, Book), s_0)]

That the magazine remains clear in the resulting state, however, cannot be established with only the available contents of Γ. As Horty explains,

> [T]his intermediary step seems perfectly natural from the standpoint of one's ordinary reasoning about actions… In fact, however, nothing in **Γ** allows this intermediary step to be derived – and indeed, the step should not be derivable as a matter of logic, for it is always possible, at least, that [the unstacking of one object from another will interfere with the fluent status of the third.] (Horty, 2001 p. 340)

That is, it is quite consistent with the contents of Γ that unstacking the pad from the book *could* result (somehow) in the coffee cup mysteriously coming to be atop the magazine. As Shanahan notes, "Although [Γ is able to] capture what does change as a result of an action, [it fails] to represent what *doesn't* change."(Shanahan, 1997 p.4)

Frame axioms

Perhaps this problem might be solved by the addition to Γ of supplementary axioms that would allow the value of certain fluents unaffected by an action to be made explicit. Let us then include in Γ the following *frame axioms*.

For the *Clear* fluent in our toy scenario, the following frame axioms should suffice:

$\forall x,y,z,s\ \mathit{Holds}[\mathit{Clear}(x), s_0) \ \& \ {\sim}(x=z) \rightarrow \mathit{Holds}[\mathit{Clear}(x), \mathit{Result}(\mathit{Stack}(y,z), s_0)]$

> If, in the initial situation, object x has nothing on it and object x and z are distinct, then the consequence of stacking object y on object z is that object x is (*i.e.,* remains) clear. More simply put, object x will remain clear unless something is stacked on it.

$\forall x,y,z,s\ \mathit{Holds}[(\mathit{Clear}(x), s_0)] \rightarrow \mathit{Holds}[\mathit{Clear}(x), \mathit{Result}(\mathit{Unstack}(y,z), s_0)]$

> If, in the initial situation object x is clear and object x and z are distinct, then the consequence of unstacking object y from z is that object x is (*i.e.,* remains) clear. Put another way, object x will remain clear even if other objects are unstacked.

Taken together these two axioms establish that clear objects remain clear unless something is stacked upon them. For completeness two additional frame axioms should be incorporated that will capture the persistence of objects once stacked. For example, if x is stacked on y and w is stacked on z, axioms are needed to guarantee that x will remain on y even if w is removed from z or if another object is put on w.

$\forall$x,y,z,w,s $\mathit{Holds}[(\mathit{On}(x,y), s_0)\ \&\ {\sim}(x=z)] \rightarrow \mathit{Holds}[\mathit{On}(x,y), \mathit{Result}(\mathit{Stack}\ (z,w), s_0)]$

> If in the initial situation object x is on object y and x and z are distinct then the consequence of stacking object z on w is that object x is (*i.e.,* remains) on top of y.

$\forall$x,y,z,w,s $\mathit{Holds}[(\mathit{On}(x,y), s_0)\ \&\ {\sim}(x=z)\ \&\ {\sim}(y=w)] \rightarrow \mathit{Holds}[\mathit{On}(x,y), \mathit{Result}(\mathit{Unstack}(z,w), s_0)]$

> In the initial situation, if object x is on y and object x and v and y and w are distinct, then the consequence of unstacking z from w is that object x is (*i.e.,* remains) on top of y.

The addition of these frame axioms to Γ appears to adequately solve the toy problem set out earlier. That is, given the addition of these frame axioms it may now be established that:

$\Gamma \vdash \mathit{Holds}[\mathit{On}(\text{Book, Magazine}), \mathit{Result}(\mathit{Stack}(\text{Book, Magazine}), \mathit{Result}(\mathit{Unstack}(\text{Note Pad, Book}), s_0))]$

Since the addition of these four frame axioms, while admittedly cumbersome, has sufficiently solved this simple toy problem, there does not appear to be much of a problem after all. Suppose, however, that only one additional action and one extra fluent are added to the original scenario. Shanahan provides an accessible and colorful example. (Shanahan, 1997 p.9)

Let *Color*(x,c) express that object x has the color c and let *Paint*(x,c) represent the action of painting object x the color c. As with stacking and unstacking the consequences of painting an object can be expressed by the formalism:

$\mathit{Holds}[\mathit{Color}(x,c), \mathit{Result}(\mathit{Paint}(x,c))]$

Modifying only slightly Shanahan's example, let us suppose that in the initial situation all of the objects are white. And so, five additional sentences must be added to the description of s_0.

$\mathit{Holds}[\mathit{Color}(\text{Coffee Cup, white}), s_0]$

$\mathit{Holds}[\mathit{Color}(\text{Book, white}), s_0]$

$\mathit{Holds}[\mathit{Color}(\text{Desk, white}), s_0]$

$\mathit{Holds}[\mathit{Color}(\text{Note Pad, white}), s_0]$

$\mathit{Holds}[\mathit{Color}(\text{Magazine, white}), s_0]$

Intuitively, stacking and unstacking an object should not affect its color, nor should it affect the color of any other object. Additionally, painting an object should not affect the color of any other object, nor should it affect the arrangement of itself or any other object. However, given Horty's discussion, given only the contents of Γ, such intuitively obvious results cannot be established. Just as axioms are required to establish that clear objects remain clear and stacked objects remain stacked after an unrelated action is performed, so too must additional axioms be added to Γ to account for persistence after painting.

$\forall x,y,c,s\ \mathit{Holds}[(\mathit{Color}\ (x,c), s] \rightarrow \mathit{Holds}[\mathit{Color}(x,c), \mathit{Result}(\mathit{Stack}(x,y), s)]$

> If in some situation object x is c colored, then the result of stacking x on y will be that object x is (*i.e.,* remains) c colored.

$\forall x,y,c,s\ \mathit{Holds}[\mathit{Color}(x,c), s] \rightarrow \mathit{Holds}(\mathit{Color}(x,c), \mathit{Result}(\mathit{Unstack}(x,y), s)]$

> If in some situation object x is c colored, then the result of unstacking x from y will be that object x is (*i.e.,* remains) c colored.

$\forall x,y,c,s\ \mathit{Holds}[(\mathit{Color}(x, c_1), s)\ \&\ {\sim}(x=y)] \rightarrow \mathit{Holds}[\mathit{Color}(x, c_1), \mathit{Result}(\mathit{Paint}(y, c_2), s)]$

> If object x is c_1 colored and objects x and y are distinct then the result of painting object y will be that object x is (*i.e.,* remains) c_1 colored.

The addition of the *Paint* action also requires the addition of the following frame axioms to account for, respectively, the fact that stacked objects will remain stacked and clear objects will remain clear when any object is painted.

$\forall x,y,z,c,s,\ \mathit{Holds}[\mathit{On}(x,y), s] \rightarrow \mathit{Holds}[\mathit{On}(x,y), \mathit{Result}(\mathit{Paint}(z,c), s)]$

$\forall x,y,c,s,\ \mathit{Holds}[\mathit{Clear}(x), s] \rightarrow \mathit{Holds}[\mathit{Clear}(x), \mathit{Result}(\mathit{Paint}(y,c), s)]$

Likewise, axioms should be required to account for the persistence of each object's color following the stacking, unstacking and/or painting of an unrelated object or objects. Shanahan notes of the necessity of these axioms:

> In the general, because most fluent are unaffected by most actions, every time we add a new fluent we are going to have to add roughly as many new frame axioms as there are actions in the domain, and every time we add a new action we are going to have to add roughly as many frame axioms as there are fluents in the domain. [Thus] the total number of frame axioms required for a domain of *n* fluents and *m* actions will be of the order of *n* x *m*. (Shanahan, 1997 p.5)

And so, the fact that the number of frame axioms necessary for the satisfaction of a goal proposition grows with surprising rapidity as fluents and actions are added lays the foundation of the original incarnation of the frame problem as discussed by McCarthy and Hayes. Specifically, the inclusion of frame axioms will result in success only in situations in which there are a very limited number of fluents and actions. Once a problem is presented that requires more than a few actions and fluents, such as the vast majority of problems in need of solving in the world (*i.e.,* any non-contrived and non "toy" scenario), the number of frame axioms necessary is vast and quickly becomes computationally unmanageable. Understood in this manner the original incarnation of the frame problem becomes one of how to formally contend with the problem of how to reason about the effects of actions without having to contend with an unmanageably vast set of frame axioms. (Shanahan, 1997 p.5)

THE QUALIFICATION PROBLEM

In addition to the *ramification* and *persistence* problems outlined above, there is a related difficulty of a somewhat more traditionally philosophical nature that is often incorporated into the frame problem – the *qualification* problem.

By way of introducing this variant or aspect of the puzzle, consider the action axioms of stack, unstack and paint. The stack axiom for instance, $\forall x,y,s(\mathit{Holds}[\mathit{Clear}(x),s] \,\&\, \mathit{Holds}[\mathit{Clear}(y),s] \,\&\, {\sim}(x=y)) \rightarrow \mathit{Holds}[\mathit{On}(x,y), \mathit{Result}(\mathit{Stack}(x,y),s)]$ states that given that the conditions are satisfied that object x is clear, object y is clear and x and y are distinct, the result of stacking x on y will be that x will be on y. These (pre)conditions are included to eliminate the possibility of the system attempting to lift an object that has something on top of it or to stack something on top of an object that already has something upon it as well as to rule out any attempts at stacking something on top of itself. A difficulty arises when it becomes apparent that the list of conditions incorporated into the antecedent of the axiom is, as written, incomplete and cannot be made complete. (Horty, 2001)

Just as in our toy example, the conditions of non-identity and "clearness" must be met before an object may be stacked on top of another, there are countless other conditions the satisfaction of which could be as relevant to the satisfaction of its goal. For example, it is possible that the book is particularly heavy, while the magazine is particularly fragile. The book might be so heavy that it cannot be lifted. Or it might be so heavy as to crush the magazine. It is possible that painting an object will cause it to dissolve or that some of the objects are somehow attached thus making it impossible to unstack one from another. To cite two examples common throughout much of the literature, it is possible that the one of the objects is in fact an explosive device which will detonate if picked up, unstacked or painted or perhaps the objects are so highly lubricated that any attempt to stack, unstack or paint them will be unsuccessful. Taken further, it is possible that the objects are mere holograms. And, taken to the extreme, it is possible that there is a Cartesian demon who prefers the objects white and in the initial configuration, thus possibly stymieing any attempt at painting, stacking and unstacking. Any of these possibilities might make it the case that any attempt at stacking, unstacking and/or painting an object would be ultimately unsuccessful. Put another way, there are an indefinitely large number of possible circumstances which, were they to obtain, could make stacking, unstacking and painting unsuccessful.[3]

The upshot of the above discussion is that the action axioms originally incorporated into Γ, those with only minimal conditions that need to be satisfied, while suitable to our simple toy case are entirely insufficient. In order to suitably qualify the axioms, the antecedents of each must include an exhaustive and indefinitely large listing of such conditions or qualifications that must be satisfied before it can be determined if an action will ultimately be successful. By way of example, consider the *Stack* axiom.

The original axiom

$$\forall x,y,s(\mathit{Holds}[\mathit{Clear}(x),s] \ \& \ \mathit{Holds}[\mathit{Clear}(y),s] \ \& \ {\sim}(x=y)) \rightarrow \mathit{Holds}[\mathit{On}(x,y), \mathit{Result}(\mathit{Stack}(x,y),s)]$$

A (partially) qualified axiom

$$\forall x,y,s(\mathit{Holds}[\mathit{Clear}(x),s] \ \& \ \mathit{Holds}[\mathit{Clear}(y),s] \ \& \ {\sim}(x=y)) \ \& \ \mathit{Holds}[{\sim}\mathrm{Hologram}(x), s] \ \& \ \mathit{Holds}[{\sim}\mathrm{Hologram}(y), s] \ \& \ \mathit{Holds}[{\sim}\mathrm{Fragile}(x), s] \ \& \ \mathit{Holds}[{\sim}\mathrm{Fragile}(y), s] \ \& \ \mathit{Holds}[{\sim}\mathrm{Explosive}(x), s] \ \& \ \mathit{Holds}[{\sim}\mathrm{Explosive}(y), s] \ \& \ \mathit{Holds}[{\sim}\mathrm{Glued}(x), s] \ \ldots \ \& \ \mathit{Holds}[{\sim}\mathrm{Lubricated}(x), s] \ \& \ \ldots\ ) \rightarrow \mathit{Holds}[\mathit{On}(x,y), \mathit{Result}(\mathit{Stack}(x,y),s)]$$

The difficulty becomes immediately apparent. There are, in fact, Horty notes, two related problems here that need consideration. The first is that the list of (pre)conditions or qualifications that must be included in the antecedent of any axiom is necessarily indefinitely large. The second problem arises even if a suitably semi-exhaustive listing of qualifications is incorporated into the antecedent. Specifically, the system would need to exhaustively verify or certify that each (pre)condition has been

[3] McCarthy in "Circumscription – a form of nonmonotonic reasoning" presents a case of "cannibals and missionaries" in which the goal is to strike upon a plan in which missionaries are ferried by rowboat from one bank to another without being eaten to illustrate the problem. McCarthy points out that in order for the problem to even get off the ground a set of *ceteris paribus* clauses about the normal workings of rowboats, oarlocks, rivers and the like need to be established in order to expressly avoid the apparent necessity of the exhaustive inclusion of antecedent conditions.

satisfied *before* being able to draw any conclusion. Given that the number of axioms required to solve even a trivial toy scenario grows with alarming rapidity with the introduction of fluents and actions and since each axiom requires that an indefinitely large number of preconditions be verified, the computational burden on a system attempting to contend with even the simplest of tasks is enormous.

The frame problem, in its original technical formulation, is then in fact a constellation of several interrelated sub-puzzles. The first constituent and the parent or umbrella puzzle, the *planning problem* is the problem of determining a sequence of actions α, derivable from a systems' Γ, the execution of which in some situation *s* will result in the satisfaction of some goal proposition ϕ.

There are, however, a number of ways in which a formal system can fail in contending with the planning problem. First, a system can fail at the *projection* or *ramification* task. Trivially, any system engaged in planning must be capable of effectively reasoning about the consequences of its proposed plans. The difficulty posed by the ramification problem is rather straightforward. Granting a plausible solution to the qualification problem, it is fairly simple to provide a system with a set of axioms that will allow it to reason effectively about the intended consequences of its actions. It is, however, incredibly difficult to design a system (*i.e.,* include all the necessary axioms) that can efficiently and effectively reason about and account for the unintended (*i.e.,* non-obvious) consequences of its actions. This is of significance for some of these unintended consequences might be relevant to the ultimate success of the plan in satisfying the goal proposition. In order to rectify this difficulty *frame axioms* – axioms that make explicit the unintended as well as intended consequence of actions - are incorporated. The problem with this solution is that for even the most trivial of tasks the number of frame axioms needed to account for all of the (possibly germane) consequences is vast (*i.e.,* indefinitely large) and computationally unmanageable. The problem becomes then one of how to generate successful action plans without having to contend with an unmanageably large set of frame axioms. More broadly, the problem is one of how to design a formal system that is capable of bringing to bear those consequences of its proposed plans that are relevant without engaging in the computationally intractable task of considering and assessing each and every consequence.

Second, systems can fail in contending with the planning problem by failing to adequately contend with the *persistence* problem – the problem of "efficiently determining which things remain the same in a changing world." (Morgenstern, 1996 p.99) Specifically, the persistence problem is one of how a formal system is to efficiently and effectively update the contents of its belief-set in light of the changes and non-changes that result (or that would result when considering plans hypothetically) from the adoption of a plan. Put another way, the persistence problem is one how to reason about and account for those facts and beliefs that do and do not change as the result of the execution of some plan (*e.g.,* acting or fixing a belief). Similar to the discussion above, frame axioms might be brought to bear to make explicit what facts and beliefs will not change and which will as a result of the execution of some plan, however (and as in the above) the number of such axioms that would be needed would be vast and when implemented would become computationally unfeasible. The problem then is one of how to engineer a system that can efficiently and effectively update the relevant beliefs/facts without engaging in the computationally burdensome task of exhaustively considering and updating the totality of the contents of its Γ/belief-set (including the "updating" of those things that remain unchanged) for every plan considered.[4]

Third, systems can fail in contending with the planning problem by failing to adequately contend with the *qualification* problem. This problem, roughly the flip-side of the ramification problem, is the problem of effectively reasoning about the (pre)conditions necessary for the successful execution of an action. As with the ramification problem, in order to determine whether a given action sequence will succeed, frame axioms must be incorporated which allow the system to reason about and verify that both the obvious and non-obvious preconditions necessary for the plan to succeed are satisfied. But, once again, the number of frame axioms needed (even in toy cases) is vast and quickly becomes computationally unmanageable. And so the qualification problem is the dual puzzle of how to engineer a formal system that (1) possesses suitably qualified axioms and (2) is capable of expeditiously and effectively verifying that the relevant preconditions necessary for a plan to succeed obtain.

SITUATING THE COMPLEXITY OF THE PLANNING PROBLEM: INTRACTABLE PROBLEMS

The above introduction to the planning and frame problems has purposefully left rather vague the extent of the computational burden faced by any formal system engaged in planned action. In order to get a more complete understanding of the nature of these puzzles and an idea of just how computationally burdensome they really are, it will be helpful to examine the intrinsic complexities of a number of formulations of the planning problem: the planning problem under complete information, planning under uncertainty and planning with "sensing" under uncertainty.

[4] Some, most notably Morgenstern and Janlert, and Haugeland take the persistence problem to be *the* frame problem.

Problems in the complexity class NP are those admitting of algorithmic solutions in polynomial time by means of a *nondeterministic* Turing machine that is, a Turing machine with the capacity to reference an "oracle" – a fictional but theoretically useful mechanism that makes instantaneous and invariably correct guesses (Papadimitriou, 1993 p. 46 & p.172). Problems solvable by nondeterministic machines in polynomial time are solvable by deterministic (that is, standard and non-oracular) Turing machines in super-polynomial time. In fact, to be in the class NP just is to admit of a **N**on-deterministic **P**olynomial time solution. Quite generally, problems admitting *deterministic polynomial time* (those problems in the complexity class P) algorithmic solutions exhibit reasonable run-time performance. Those problems admitting of *nondeterministic polynomial time* algorithmic solutions that is, those admitting only of *deterministic super-polynomial time* algorithmic solution (the problems in complexity class NP) exhibit, in general, unreasonable run-time performances. Harel explains,

> As far as the algorithmic problem is concerned, a problem that admits a reasonable or polynomial time solution is said to be tractable, whereas a problem that admits only unreasonable or exponential-time solutions is termed intractable. In general, intractable problems require impractically large amounts of time on relatively small inputs, whereas tractable problems admit algorithms that are practical for reasonably size inputs. (Harel, 1996 p.166)

We have now a more precise means by which to determine whether or not a problem is an "intractable" one. Any problems admitting of *nondeterministic polynomial time* algorithmic solution – that is, *deterministic super-polynomial time* solution - are intractable and are practically unsolvable for even very small inputs.

With some of the basic terminology in place, let us consider first an instance of the planning problem in which the system has complete information. Specifically, let us assume that the system has an exhaustive account of the values of *all* fluents in the initial situation S_0 and a sufficiently stocked Γ such that the values of all fluents in subsequent situations can also known. Put another way, such a system is engaging in planning under the condition of complete certainty of both the way things are and the way things would be were some action undertaken. It has been established that the planning problem under complete information is NP-Complete meaning that this problem is in the complexity class NP and that there exists a polynomial time reduction of this problem to SAT (the known NP problem of determining the satisfiability of sentences in proposition logic). Baral, Kreinovich & Trejo (2000) offer the following proof. Establishing that planning under certainty is in the class NP is fairly straightforward for one need only establish that the activity of verifying that a solution to the problem will succeed is complete-able in deterministic polynomial time (*i.e.,* by a standard Turing machine). In Baral's terms, for a given situation ω ascertaining whether or not a satisfactory plan exists requires that the validity of the formula $\exists\upsilon P(\upsilon,\omega)$ (where $P(\upsilon,\omega)$ is interpreted as "plan υ succeeds for situation ω" be determined. To establish that this problem is in the class NP it must be shown that 1) the quantifier runs only over plans of finite and tractable length, and 2) the property $P(\upsilon,\omega)$ is *certifiable* in *deterministic polynomial time*. The first condition Baral notes follows immediately for only plans of finite and polynomial (tractable) length may be considered. Plans of infinite length are disallowed as no evaluation could possibly be made with respect to their success and extraordinarily lengthy plans are proscribed because these are practically impossible to evaluate. As to the second condition, that a plan υ is successful in ω can easily be *certified* in polynomial time.

Specifically, since the initial state $\omega/(s_0)$ is completely known, one need only inspect the state, call it s_1, that results when the first action of the sequence is executed. Next, the state s_2 that is the result of the execution of the second action in the plan (applied to state s_1) need be inspected. Eventually, since the plans under consideration are of tractable lengths, at some situation s_n one need only determine whether in this final state of affairs the goal proposition is satisfied. Baral explains, "on each step of this construction, the application of an action to a state requires linear time; there are a polynomial number of steps in this construction. Therefore, this checking indeed requires polynomial time." (Baral, Kreinovish & Trejo, 2000 p. 255) Since certification is in P, the problem is also solvable by a *nondeterministic* Turing machine in polynomial time.[5] As such, the problem is solvable by a *deterministic* Turing machine in super-polynomial time. And so, the planning problem under complete information, Baral concludes is in the class NP.[6] And so, the planning problem even under conditions of complete information is an *intractable* puzzle.

While the planning problem under complete knowledge is difficult enough (*i.e.,* NP-complete and therefore intractable) very few plans are actually formulated under the conditions of complete knowledge. Rather, in most non-artificial environments, the values of at least some of the fluent will be unknown. Planning under incomplete knowledge is, as a member of class of even more intrinsically difficult puzzles, a significantly more computationally burdensome task.

[5] The idea here is fairly straightforward. Consider SAT - the problem of satisfying some sentence of proposition logic (*i.e.,* finding which assignments of T/F to predicates will result in the sentence being true). While solving this puzzle is quite difficult – the brute force naïve algorithm of checking 2^n possible assignments is the most efficient algorithm known – the issue of verifying or certifying a proposed solution once found is quite trivial. One need merely insert the values and assess the sentence and determine if the sentence is true under this assignment. Certification then takes place in polynomial time (in this case, linear time). If certification is in P time, then a nondeterministic Turing machine (with access to an oracle) could solve the problem in P time – for it would always make correct assignments. That is the problem has a nondeterministic polynomial time solution. And so, SAT is NP.

[6] In fact, the planning problem under complete information has been shown to be NP-complete, which for present purposes means that it is a member of the class NP and is polynomial-time reducible to the satisfiability problem of proposition logic (SAT.)

Since the formulation of plans under uncertainty entails that the values of some fluents are unknown, the planning problem becomes one of determining whether an action-sequence υ_1 exists that will result in success for every set of values υ_2 of unknown fluents. Given some situation ω, ascertaining whether a satisfactory plan υ_1 exists entails, Baral *et al.*, (2000) explains, that the validity of the formula $\exists\upsilon_1\forall\upsilon_2 P(\upsilon_1, \upsilon_2, \omega)$ be determined.[7] This problem has been established to belong to the complexity class Σ_2P-complete (alternatively NP^{NP}) for it is of iterated complexity. Planning under incomplete information is of iterated complexity for the problem of determining the values of the unknown fluents can be accomplished in polynomial time by a nondeterministic Turing machine (since such values are certifiable by a deterministic machine in P time). Once the values of the unknown fluents are determined, the planning problem reduces to the "simple" problem of planning under complete knowledge.

The values of unknown fluents, as determinable in nondeterministic polynomial time are determinable in deterministic super-polynomial time and as such this problem in isolation falls into the class NP. Once these values are known the problem reduces to one of planning under complete knowledge which is a known NP-complete problem. And so, the planning problem under incomplete information is, as an NP-complete problem nested within an NP-complete problem, decidedly intractable.

Suppose now that we consider the problem faced by a system that is capable of making conditional plans under uncertainty. That is, suppose that we consider a system that is capable "sensing" or checking the values of various fluents after the execution of

[7] $P(\upsilon_1, \upsilon_2, \omega)$ is interpreted as: plan υ_1 succeeds for the values υ_2 of initially unknown fluents in situation ω. Baral, Kreinovich & Trejo, 2000 p. 257.

an action, and that is able to modify its conditional plans in light of what is finds. Baral explains,

> In the presence of sensing, an action sequence may no longer be a predetermined sequence of actions: if one of these actions is sensing, then the next action may depend on the result of the sensing. In general, the choice of a next action may depend on the results of all previous sensing actions. Such an action is called a conditional plan. (Baral, Kreinovich & Trejo, 2000 p. 245)

As this version of the planning puzzle tracks most closely earlier discussion it of interest to note that this problem has been shown to a member of the class of problems PSPACE-Complete – a level of complexity above Σ_2P-complete in the complexity hierarchy.

The problem of planning under uncertainty with "sensing" may be reformulated as one of determining whether or not an action υ_1 exists such that for every possible result of "sensing" υ_2 there exists a second action υ_3 such that for every possible result of "sensing" υ_4 there exists a third action υ_5 and so on, such that at the end of the process the goal situation obtains. More formally, the problem may be restated as one of determining the validity of "$\exists\upsilon_1\forall\upsilon_2\exists\upsilon_3\forall\upsilon_{4\ldots}\forall\upsilon_\kappa P(\upsilon_1\ldots\upsilon_\kappa, \omega)$ in which $\upsilon_1\ldots\upsilon_{\kappa-1}$ represent actions and the results of sensing actions and υ_κ runs over all possible values of un-sensed (unknown) fluents." (Baral *et al.,* 2000 p.259). For simplicity I will omit the particulars of the proof that this problem is in the class PSPACE-Complete as such detail would be counterproductive.

It is, however, sufficient to note of this finding that, "a property whose existence question is PSPACE-complete probably cannot be *verified* in polynomial time using a polynomial-length 'guess.'" (Garey & Johnson, 1979 p.172). That is, problems in the complexity class are approaching an extraordinarily high level of complexity – for it appears that since their solutions (once found) cannot even be verified in polynomial time by a nondeterministic (*i.e.,* oracular) machine – the *certification* problem is itself one that is at least as complex as those in the class NP. Putting this another way, such problems are so hopelessly complex that the task of merely verifying just one proposed solution is itself intractable.

The above is presented so that we have an unambiguous understanding of the nature of intractable puzzles, a further understanding of the nature of the planning problem, and to make quite clear that any formal system contending algorithmically with the planning problem (*i.e.,* of providing complete and completely correct solutions) is faced with an inordinately complex and computationally unmanageable task. Specifically, the task is so computationally unmanageable as to be practically incomplete-able by even the most highly idealized of systems.

While the planning/frame problem, as original presented, is a formal and somewhat technical problem or problem-set, the underlying puzzle is not a particularly technical one nor is the puzzle limited to the particular formalism of the situation calculus. And so, having sketched the technical puzzle I will present what is sometimes termed the "philosopher's frame problem" with the aim of now discussing the problem in a more abstract, generalized and accessible way. To this end, I will consider the versions of the puzzle as discussed by Dennett, Haugeland, and Fodor.

DENNETT'S FRAME PROBLEM

Dennett offers a readily accessible set of scenarios that provide a less formal introduction to the puzzles underlying the frame problem. He begins by asking us to consider a situation in which there is a room containing a wagon, a battery and a ticking time bomb. Both the bomb and the battery are on the wagon. We are next asked to consider three robots, each of which is set the goal of retrieving the battery (its power source) from the room.

The first robot, R1, having hypothesized that pulling the wagon (atop which the battery rests) will result in the battery being removed from the room, enacts this plan. However, in executing this plan R1 is destroyed. In pulling the wagon out of the room (atop which the battery rests) the bomb is also removed. The robot R1, Dennett explains, "*knew* that the bomb was on the wagon in the room, but didn't realize that pulling the wagon would bring the bomb out along with the battery. Poor R1 had missed that obvious implication of its planned act." (Dennett, 1990 p. 147)

R1 then failed to account for the *ramifications* of the execution of the proposed action sequence. R1, in short, failed to see the very relevance of the fact (a rather obvious unintended side-effect) that since the bomb too was on the wagon it would come along for the ride.

Since R1 was unable to enact a successful action sequence due to its inability to adequately project the ramifications of its plans, a second robot is created. This new robot R1D1 is programmed to recognize not only the intended implications and consequences of its plans, but also the implications of any and all side effects of these plans. In more technical terms, R1D1 has been provided with an exhaustive set of axioms designed to eliminate R1's reasoning gaffs with respect to the ramification problem. However, this attempted correction of R1's deficiencies has its own shortcomings. R1D1, being programmed to account for the non-obvious ramifications of its plans becomes mired in the computationally intensive process of generating all of the possible implications of its proposed action sequence. Even assuming the set of implications to be finite in size, the amount of time required to complete these calculations, as an instance of the previously discussed planning problem under incomplete knowledge, is excessive. Dennett explains, noting that R1D1,

> Having just deduc[ed] that pulling the wagon out of the room would not change the color of the room's walls, [while just] embarking on a proof of the further implication that pulling the wagon out of the room would cause its wheels to turn more revolutions than there are wheels on the wagon [was destroyed by the bomb]. (Dennett, 1990 p. 147)

And so while R1D1 is equipped in principle to generate all of the implications of its proposed action-plan – and thus evade the problem that stymied R1 by generating all of the potentially relevant ramifications of its plans - it is still doomed, precisely because it has no means by which to effectively limit the inferences that it draws.

Finally, in order to contend with the difficulties plaguing the first two robots, a third R2D1 is developed. Unlike the previous models, R2D1 is imbued with the ability to distinguish between the relevant and irrelevant implications of its plans. That is to say, R2D1 has been provided with some mechanism (*i.e.,* a relevance metric) by which to categorize the implications of its plans into (1) those that are relevant to the satisfaction of its goals and (2) those that are not. However, while R2D1 is much more richly programmed, it too fails miserably in contending with the planning problem. Specifically, though R2D1 is capable of inferring the implications of its proposed actions and it is capable of assessing which consequences are relevant (and which are not) it remains motionless, effectively cognitively/computationally paralyzed. Specifically, since R2D1 is provided no means by which to determine which consequences are relevant (and thus which inferences it should bother making) beforehand, it is compelled to generate and consider every implication of its plans. This is so because R2D1 can only make determinations of relevance *a posteriori* – that is after it has generated each ramification of its plan. Dennett explains,

> "Do something!" [the programmers] yelled at it. "I am" it retorted. "I'm busily ignoring some thousands of implications I have determined to be irrelevant. Just as soon as I find an irrelevant implication, I put it on the list of those I must ignore …" and the bomb went off. (Dennett, 1990 p. 148)

And so, while R2D1 has been provided with both an exhaustive set of axioms describing the effects (both changes and non-changes) that result from its plan and a relevance metric by which to discern which implications and effects are relevant and which are not, R2D1 troubles are by no means eased. Rather, R2D1 is doubly troubled for not only must/can it generate an exhaustive set of the implications of its plans it must then assess each implication for its relevance. Since the task of generating the set of implications is what caused R1D1's demise, and R2D1 must do this and then proceed to assess the relevance of each implication, it would appear that R2D1 has no hope of effectively and efficiently contending with the planning problem.

Dennett takes the frame problem to be "a new, deep epistemological problem – accessible in principle but unnoticed by generations of philosophers – brought to light by the novel methods of AI, and still far from being solved."(Dennett, 1990 p.142) Given the claim that all three robots suffer from the frame problem it would appear that Dennett understands the problem to be a constellation of related puzzles. Taken together, then, the frame problem for Dennett is the puzzle of designing a system that is capable of expeditiously generating a successful sequence of actions given some problem posed. However, any system capable of successfully contending with the planning problem must be one that is capable of contending with the problem of expeditious (*i.e.*, non-exhaustive) and reliably correct relevance determination.

HAUGELAND'S FRAME PROBLEM

Haugeland suggests that the frame problem is the problem of "keeping temporal knowledge up to date, when there are side effects."(Haugeland, 1987 p. 82) In keeping with earlier discussion of the parent puzzle - the planning problem - Haugeland notes that any plan enacted in or on some original situation S_o will result in a consequent situation S_c. Immediately, however, a problem arises for any formal system engaged in planned action – the *persistence* problem. Most facts will remain unchanged in the resultant situation while others will change. Consequently, Haugeland notes, any formal system engaged in planned action must be able to update a variety of its situational beliefs. (Haugeland, 1987 p. 82) However, assuming the system's belief-set is of a non-trivial size, the task of exhaustively considering each belief and assessing whether each is to be updated is an intractable one. The frame problem for Haugeland then

> Concerns how a system can ignore most conceivable updating questions and confront only 'realistic' possibilities. The issue is how to 'home in on' relevant considerations, without wasting time on everything else the system knows. Thus, the challenge is not how to decide for each fact whether it matters, but how to avoid that decision for almost every bit of knowledge. (Haugeland, 1987 p. 82)

In this respect, Haugeland's version of the frame problem echoes Dennett's discussion insofar as both claim the problem to be one of how to contend with the planning problem without exhaustively considering the system's belief-set (and the set of inferences from this) and without ignoring those beliefs/inferences that are relevant. Put another way, the challenge for both is how to engineer a system that successfully contends with the planning problem while avoiding the computational paralysis associated with exhaustive consideration.

FODOR'S FRAME PROBLEM

While considerable attention will be devoted to both Fodor's version of the frame problem as applied to belief-fixation and his critique of computational psychology, I will provide here only a brief introduction to his version of the problem. The frame problem, Fodor suggests presents a "deep problem of rationality" that he terms *Hamlet's problem* - the puzzle of determining "when the evidence you have looked at is enough."(Fodor, 1987 p. 140) Given that any flexibly intelligent system will have to contend with the planning problem, the system must, if it is to generate reliably correct plans (*i.e.*, satisfactorily contend with the problem), consider a "non-arbitrary sample of the available evidence."(Fodor, 1987 p. 140) Having to consider a non-arbitrary sample of evidence, Fodor argues, is tantamount to having to consider the totality of the system's belief set. Since exhaustive consideration is untenable (as an intensively computationally demanding task) and considering only a sub-set of the belief-set, Fodor claims, is not a reliable means by which to plan and fix-beliefs, any flexibly intelligent computational system is faced with the problem of how to expeditiously and reliably arrive at correct conclusions (action-plans, belief-fixing). Fodor explains,

> The frame problem is just Hamlet's problem viewed from an engineer's perspective. You want to make a device that is rational in the sense that its mechanisms of belief fixation are unencapsulated. But you also want the device you make to actually succeed in fixing a belief or two from time to time; you don't want it to hang up the way Hamlet did. So, on the one hand, you don't want to delimit its evidence searches arbitrarily (as in encapsulated systems); and, on the other, you want these searches to come, somehow, to an end. How is this to be arranged? What is a nonarbitrary strategy for delimiting the evidence that should be searched in rational belief fixation? I don't know how to answer this question. If I did, I'd have solved the frame problem.(Fodor, 1987 p. 140)

That is, Fodor takes the frame problem not to be the problem of generating the exhaustive set of all possible implications of a proposed plan. Rather, like Dennett, he takes the problem to be one of how to engineer a system that is capable of generating reliably successful plans (fix reliably correct beliefs) *without* having to exhaustively search and evaluate all of the implications of each proposed plan/commitment and *without* having to limit or truncate search arbitrarily.

THE FRAME PROBLEM AS A GENERIC PUZZLE OF EXPEDITIOUS RELEVANCE DETERMINATION

While there is a multitude of examples and discussions as to the scope and nature of the frame problem, ranging from those that take the problem to be only the persistence problem to those that take the problem to be a constellation of related problems, there is an underlying puzzle shared by all accounts. Each discussion, while differing in emphasis and focus, presents the frame problem as one of how to engineer a system that is capable of both expeditiously arriving at reliably correct conclusions about which "things" (*i.e.,* plans, facts, beliefs, inferences and implications) are *relevant*. In support of the claim that the frame problem should be understood both as a rather generic puzzle and as a puzzle composed of a constellation of sub-problems (or problem kind) each of which instantiates the central problem of relevance determination, Glymour writes,

> The 'frame problem' is not one problem, but an endless hodgepodge of problems concerned with how to characterize what is *relevant* in knowledge, action, planning, etc. Instances of the frame problem are all of the form: Given an enormous amount of stuff, and some task to be done using some of the stuff, what is the *relevant stuff* for the task? (Glymour, 1987 p.65)

RELEVANCE DETERMINATION AND A SET OF FRAME PROBLEM INSTANCES

While the frame problem is characterizable as the problem of expeditious relevance determination, there are, as Glymour suggests, numerous instances of the problem (*i.e.,* frame-ish puzzles) that need to contended with by any system engaged in planned action, decision-making and belief-fixation. While likely not an exhaustive list, the following presents a fair sample.

THE PROBLEM-SETTING OR META-PLANNING PROBLEM

Although each of the accounts outlined begins with a system that is previously set a task with which it must contend, (*e.g.,* Dennett's robots are set the problem of getting the battery) the frame problem as the problem of relevance determination also arises in the realm of problem-setting. Any flexibly intelligent system must be capable of guiding the kinds of ends and lines of inquiry that it will pursue. That is, any intelligent system must be capable of determining which situations pose "problems" for it and which do not without exhaustively considering them all. Put another way, the system must contend with the problem of determining what it should "think about" now without having to first "think about" everything. Some situations – the relevant ones - need to be contended immediately with while others – the less pressing ones – need to be contended with eventually, while still others – the irrelevant ones – are to be ignored. Without contending with this instance of the puzzle the system would be paralyzed before it even began being paralyzed by the problems raised by Dennett, Haugeland and Fodor. And so, it would appear that any flexibly intelligent system is faced with the problem of expeditiously determining which lines of inquiry it will pursue (*i.e.,* which situations it should treat as posing a problem and begin formulating a plan) and which it will not (*i.e.,* which situations it should ignore).

THE ATTENTIONAL-DIRECTION PROBLEM

There is, in any situation that a "sensing" system might find itself, some sub-set of all the "things" that might be attended are things that should be attended (*i.e.,* that are relevant).[8] That is, any system that is capable of directing its attention is faced with the

[8] By "sensing" system I mean here any system that when engaged in a (conditional) planning task under uncertainty is capable of attending to (*i.e.,* verifying) the values of fluents and adjusting its

frame problem instance (qua relevance problem instance) of determining which "things" it is to attend to and which it is to ignore. Arriving at an all-things-considered conclusion about which "things" the system will attend is clearly computationally burdensome for so doing would require first that each "thing" be attended and second that a determination be made as to whether (or not) that "thing"' is worthy of attention. And so, the problem of attentional direction, as an instance of the frame problem, is one of how to design a system that will expeditiously attend to those "things" in its environment that are relevant (to either the problem of problem-setting or planning) while efficiently ignoring (*i.e.,* not attending to) those that are not.

THE MEMORY-ENCODING PROBLEM

Any finite system that is capable of encoding material to memory is faced immediately with the problem of determining which things are to be remembered (*i.e.,* incorporated into its Γ) and which are to be ignored. Assuming that memory is finite, it follows that not everything can be remembered. If so, then some limiting mechanisms must be at play by which to direct which "things" are to be stored and which not. Managing the contents of memory in an all-thing-considered manner would require that a conclusion be reached with respect to the relevance of each "thing" considered. The system would be faced with the computationally burdensome task of considering each "thing" and arriving at a conclusion as to whether or not it is relevant enough to be remembered. The problem then is how to design a system that can manage the contents of memory by expeditiously determining which things – the relevant things – will be encode.

conditional plans accordingly. C.f. Baral, Kreinovich & Trejo (2000) for discussion of the computational complexity of the planning and conditional planning problem.

Put another way, these two instances of the problem when taken together might form what might be call the extraction problem. Given the contents of a system's environment and a finite memory store, how is the system to extract from its environment (attend to and encode) only those facts that are relevant to the successful determination of an action plan that satisfies a goal proposition.

THE SEARCH PROBLEM

Any system engaged in planning is faced with the problem of finding those beliefs/facts stored in memory that are relevant to contending with the problem at hand. Given some problem posed some facts – the relevant ones - need to be brought to bear while most need not be. Arriving at an all-things-considered conclusion about whether (or not) each member of the belief-set should be brought to bear (*i.e.,* is relevant) is a clearly computationally burdensome task. This is so, for assuming a non-trivially sized belief-set/memory store, the system would need to first access each member of its belief-set and then arrive at a conclusion with respect to whether (or not) it is relevant to contending with the problem posed. And so, any flexible system engaged in planned action is faced with the frame problem instance of how it is to expeditiously search memory (*i.e.,* the belief-set), access and bring to bear those (and only those) facts that are relevant to contending with the problem posed.

Put another way the retrieval, access or search problem is the following: Given all that a systems knows, believes or otherwise has stored (*i.e.,* the contents of its Γ), how is the system to retrieve only those facts (*i.e.,* some subset of Γ) that are relevant to the successful determination of an action sequence and thus satisfaction of some goal proposition, given some problem to be solved or task to be completed.

THE INFERENCE PROBLEM:

Most in line with earlier discussion of the frame problem and a variant of the search problem, any system capable of planned action is faced with the problem of expeditiously making and bringing to bear the relevant inferences from both any plan under consideration and its belief set. Doing so presents a problem, for any system that attempts to arrive at all-things-considered conclusions about which inferences it should make is compelled to first exhaustively generate and second assess the complete set of producible inferences - a clearly computationally intensive task - as Dennett's examples illustrate. And so, any flexibly intelligent system is faced with the problem of how it is to expeditiously make those (and only those) inferences that are relevant to contending with the problem posed.

Put another way, the inference or projection problem is the following: Given all of the inferences that a system could produce from the contents of its Γ, how is the system to produce only those inferences that are relevant to the successful determination of an action plan that satisfies the goal proposition.

As previous discussion outlined, the inference problem is further decomposable into the ramification, qualification and persistence/updating sub-problems discusses above. Specifically, any flexibly intelligent system must be able to: (a) expeditiously and reliably infer the relevant ramifications of its plans without considering those that are (obviously) irrelevant; (b) expeditiously and reliably determine which preconditions need to be satisfied before a plan is enacted (*i.e.,* the relevant ones) without exhaustively generating and verifying each qualification; and (c) expeditiously and reliably determine which beliefs need to be updated (*i.e.,* the relevant ones) and which persist without exhaustively evaluating (*i.e.,* updating) the totality of the belief-set.

THE DECISION PROBLEM

Any system engaged in planning is faced with the task of ultimately deciding upon/enacting a course of action. That is, any flexible system must determine when it is to stop thinking about a problem and/or its plans for contending with it and act/decide. Arriving at an all-things-considered conclusion with respect to this instance of the problem is untenable since this will first require the system to first generate an indefinitely large set of plans, second to then evaluate the relevance of or assess the likely success of each before any plan is adopted and third to ultimately adopt/enact one plan. And so, any flexibly intelligent system is faced with the problem of when it is to "stop thinking" about a problem and when it is to, in effect, stop planning and act.

In the next chapter, I will focus on Fodor's version of the problem with respect to the activity of belief-fixation. So doing will serve to both situate previous discussion and bring this puzzle to bear on the computational theory of mind.

CHAPTER 2: FODOR'S CHALLENGE, FRAME PROBLEMS AND COMPUTATIONAL PSYCHOLOGY

Having introduced the frame problem and suggested both that it is reducible to a generalized puzzle of relevance determination and that the problem is best understood to be a constellation of related instances of a problem kind, I will next consider Fodor's argument that doxastic processes cannot be modeled in computationally feasible terms. Since Fodor's argument explicitly extends the frame problem to computational psychology, examining his arguments will serve the purpose of providing further insight into both the nature of the frame problem as well as the constraints that must be satisfied by any proposed solution. Since I will in the next chapter argue that Fodor's pessimism with respect to the model-ability of these processes is unwarranted, his argument will need to be unpacked.

Methodologically, while Fodor's discussion of the frame problem and his argument for the pessimistic conclusion with respect to the future of cognitive science might be presented individually, I see little benefit in taking this approach. Since Fodor's argument is broadly reducible to the following:

1. We (at least sometimes) fix beliefs in a normatively rational manner.
2. Doing so requires that we contend successfully with the frame problem – *qua* the problem of expeditious relevance determination.
3. Any system that satisfactorily contends with this problem engages in tasks that are not amenable to modeling in computationally feasible terms.
4. Therefore, our doxastic processes are unmodel-able.

I believe there to be every reason for presenting these discussions simultaneously.

Fodor (1983, 1987, 2000) argues for a rather pessimistic conclusion with respect to the future of cognitive science. This pessimism most succinctly stated by Fodor's *First law of the nonexistence of cognitive science* (Fodor, 1983 p. 107) ultimately reduces to the claim that while our cognitive processes are computationally realized, the "interesting activities of mind" (*i.e.,* our doxastic processes) are not amenable to modeling in computationally feasible terms. Fodor aims to establish that i) because we have no notion of how to even go about modeling doxastic processes in computationally feasible terms, and ii) since model-ability appears to be the only means by which we could ever come understand the workings of mind, and iii) since cognitive science has taken as its principle aim the task of explaining these processes in computational terms, we should be wholly pessimistic with respect to the future of cognitive science.

In support of the pessimistic conclusion that "(it appears) that Classical computations have no way to model [cognition]" (Fodor, 2000 p. 77) Fodor offers two distinct, though inter-related argument strains. These are what I will call the *argument from psychological isotropy* and the *argument from psychological (Quinean) holism*.

In order to situate the discussion, it will be helpful if we have a brief overview of Fodor's argument. At base, the argument proceeds from the following claims.

- Arriving at conclusions rationally is the only means by which one may reliably arrive at correct conclusion (*e.g.,* generate successful plans, fix correct beliefs).
- Arriving at conclusions rationally requires that *all* of the evidence that is both *available* (to the system) and *relevant* (to the problem) be considered.

Fodor's aim is to establish that the second of these leads to the pessimistic conclusion. Once his arguments are unpacked and two principal strains distinguished, I will suggest that the *argument from psychological isotropy* is aimed at establishing that *all* beliefs in an agent's belief-set are available to him or her during the course of deliberation over what else he or she will come to believe. The *argument from psychological (Quinean) holism*, I suggest, aims to secure that claim that relevance can be determined neither locally nor *a priori* but only globally and *a posteriori*. Put another way, this argument aims at establishing that any system capable of making determinations of relevance (at all) must be one that is capable of considering the totality of its belief-set.

Broadly, then, if Fodor can establish that:

1. Our doxastic processes are *isotropic* - that *all* of the beliefs in an agent's (or system's) belief-set are available to it in the course of its deliberations and,
2. Our doxastic processes are *Quinean* – that *all* of these available beliefs are relevant (*i.e.,* since relevance can only be determined *a posteriori,* each belief is *a priori* as relevant as any other) and,
3. We *at least sometimes* arrive at conclusions rationally.

Then we would have reason for thinking Fodor's pessimism with respect to the future of cognitive science to be warranted. It is to the task of unpacking these two argument strains that I will now turn.

THE ARGUMENT FROM PSYCHOLOGICAL ISOTROPY

Fodor's argument from psychological isotropy aims to establish that any particular belief and thus all beliefs in a system's (or agent's) belief-set are *in principle* available to it during the course of arriving at conclusions about what else it will come to believe. By way of a general overview, the argument proceeds initially from the claim that our doxastic processes (*i.e.,* the operations of belief-fixation) are computationally realized. Next Fodor claims that *psychological isotropy* is a property of our doxastic processes. This descriptive claim about our psychological processes serves as a premise in Fodor's argument that our cognitive processes must be architecturally *informationally unencapsulated*. Taken in conjunction, these claims lead Fodor to conclude, *via* a *reductio* style argument, that the workings of our doxastic processes cannot be computationally modeled. Since these operations are not apparently amenable to algorithmic description there can be, Fodor concludes, no meaningful science of mind.

The argument from isotropy proceeds in two stages. The first aims at establishing that our mechanisms of belief-fixation possess the property of psychological isotropy and as such must be informationally unencapsulated. The second stage of the arguments aims to establish that if our doxastic processes are mediated by informationally unencapsulated mechanisms then there exists no viable strategy by which to model these processes in computationally feasible terms. And so, if psychological isotropy is a property of our doxastic processes and if such processes, while computationally realized, are not model-able, since cognitive science has taken as it aim explaining cognition in computational terms, then, Fodor concludes, the workings of our doxastic processes and cognition quite generally will remain a mystery for the foreseeable future.

FROM PSYCHOLOGICAL ISOTROPY TO INFORMATIONAL UNENCAPSULATION

Before proceeding, a few concepts and some terminology must be presented. Epistemologically, *isotropy*, Fodor explains, "is the principle that *any* fact may turn out to be (ir)relevant to the confirmation of any other." (Fodor, 1983 p. 105) Fodor continues, noting,

> By saying that confirmation is isotropic, I mean that the facts relevant to the confirmation of a scientific hypothesis may be drawn from anywhere in the field of previously established empirical (or, of course, demonstrative) truths. Crudely: everything that the scientist knows is, in principle, relevant to determining what else he ought to believe. In principle, our botany constrains our astronomy, if only we could think of ways to make them connect. (Fodor, 1983 p. 105)

As presented, isotropy may be construed either as a descriptive claim about the manner in which confirmation actually occurs, or as a normative claim about the manner in which hypotheses ought to be confirmed. I will suggest that Fodor offers two arguments aimed at establishing the descriptive claim that our doxastic processes are what I will term *psychologically isotropic*. Broadly, this amounts to the descriptive assertion that, in the course of arriving at conclusions, there are no beliefs possessed by agents (or systems) that are not available to them in the course of their deliberations.

As there is little in the way of direct and explicit argument in support of this claim, I will attempt to reconstruct two lines of argument (both of which find support in Fodor's work) for the claim that isotropy is a property of our doxastic processes. I will consider a third argument in support of this claim in a later section once the argument from *psychological (Quinean) holism* has been presented.

THE INTUITIVE ARGUMENT

Psychological isotropy, understood as a descriptive claim, suggests that any piece of information (any belief in the belief-set) possessed by the agent may *at least in principle* be brought to bear in the course of deliberation. Underlying this claim is the intuition that it really does seem as if we are capable *at least in principle* of accessing *any* belief that we happen to hold, at *any* time, given *any* proposition (hypothesis or belief) under consideration. Furthermore, it really does seem that we are capable of entertaining simultaneously *any* set of thoughts – even those drawn from apparently radically disparate content domains – at *any* time, and given *any* proposition under consideration. Prima facie it really does appear, for example, that we are quite capable of entertaining thoughts about AAA and thoughts about *zymurgy* (and, of course *any* random assortment of beliefs from A to Z) at *any* time no matter what else we might be thinking about. We can, further, seemingly entertain thoughts about AAA and *zymurgy* simultaneously thus enabling us to in the course of our deliberations consider what (if any) bearing our beliefs about these matters may have on each other and on the issue under consideration. We can, so it seems, *at least in principle* consider, in the course of deciding what else to believe, *anything* else that we believe. Or, the other way round, it really does seem that there are no beliefs that we possess that are *in principle* unavailable to us in the course of our deliberations. While such a claim seems intuitively plausible, given that we have reason for thinking our introspective intuitions about the working of our own mental processes to be at least somewhat suspect and given the radical implications of the conclusion suggested by Fodor, perhaps more in the way of argument is needed.

THE ARGUMENT FROM THE HISTORY OF SCIENCE

Perhaps the intuitive argument is not quite what Fodor is offering. Perhaps he is making the following empirical claim. Suppose that we assume that psychological isotropy is not a property of our doxastic processes (*i.e.,* that information is somehow partitioned). If so, then it would follow that in the course of arriving at conclusions there must be some particular pieces of information that we possess that, as partitioned, absolutely cannot be brought to bear on some hypothesis or problem. Fodor appears to be suggesting that for any manner in which information could be partitioned there can be found an actual instance in the history of science in which an agent, in arriving at a conclusion, brought to bear information that he or she (architecturally) should not have been able to access.

For example, suppose that my doxastic processes are not isotropic, that I have beliefs about AAA and zymurgy and that the non-isotropic (*i.e.,* encapsulated) model proposed for describing my doxastic processes asserts that I should not even in principle be able to entertain thoughts about AAA and zymurgy simultaneously. Fodor's argument appears to amount to the claim that, were we to examine the history of science, someone, in arriving at some conclusion, actually did simultaneously bring to bear thoughts about AAA and zymurgy. Seemingly, making the case is, in fact, even easier for Fodor, for even if no one has yet simultaneously brought to bear his or her beliefs about AAA and zymurgy, at least one person has now.

And so, Fodor appears to suggest, since no model of mind in which sets of information are partitioned from others exists without empirical counterexample, there can be no model of the workings of our doxastic processes that denies the property of psychological isotropy.

FROM PSYCHOLOGICAL ISOTROPY TO NON-MODEL-ABILITY

Let us assume that Fodor has established that our doxastic processes are psychologically isotropic. Fodor aims next to secure the claim needed to support the pessimistic conclusion, that "isotropic systems are *ipso facto* unencapsulated." (Fodor, 1983 p. 106) While psychological isotropy is a property of minds and informational encapsulation is a property of processes, once some terminology has been explained, the move from psychological isotropy to informational unencapsulation becomes a trivial one. To be a "module," Fodor explains, is to be an

> Informationally encapsulated computational system – an inference-making mechanism whose access to background information is constrained by general features of cognitive architecture. One can conceptualize a module as a special purpose computer with a proprietary database, under the conditions that: (a) the operations it performs have access only to the information in its database … and (b) at least some information that is available to at least some cognitive processes is not available to the module. (Fodor, 1985 p.3)

Modular mechanisms then are computational systems that arrive at conclusions by considering definitionally only a subset of the total information available to the system at large, for this subset is all that is available to them. Granting that Fodor has established that psychological isotropy is a property of our doxastic processes the argument for the informational unencapsulation of these processes follows rather quickly. The *reductio* argument proceeds along the following lines. Suppose that the mechanisms realizing our doxastic processes are informationally encapsulated. By definition, such mechanisms would have at their disposal only the information present in their proprietary databases. Furthermore, there must, by definition, be information present elsewhere in the system (*i.e.,* available to some other module) that is not available to the module under consideration. As such, there must be at least one piece of information (*i.e.,* belief) that, while available to at least one other module, cannot be available to the module under consideration. In more psychological terms, there must be at least one belief that, while held by the agent (or system), is unavailable to him or her during deliberation. If psychological isotropy is a property of our doxastic processes then the mechanisms mediating these processes cannot be informationally encapsulated. This follows for psychology isotropy holds that that *any* belief the system/agent holds can at least in principle be brought to bear during deliberation and information encapsulation demands that at least one belief (held elsewhere) is unavailable during deliberation. Therefore, if psychological isotropy is a property of our doxastic processes and assuming these processes are computationally realized, then the mechanisms mediating these operations must be informationally unencapsulated – for encapsulation violates isotropy.

The argument from psychological isotropy, then, attempts to secure that claim that our cognitive system must be organized so as to allow us access to any belief that we possess in the course of arriving at a conclusion about what else to believe. Since any particular belief is available, it follows that the totality of the belief-set is available during deliberation. Let the following then be the conclusion of the argument from psychological isotropy and a statement of what I will term Fodor's availability assertion.

- Any particular belief possessed by an agent (*i.e.,* any member of a system's belief-set) is available to him/her during the course of deliberation about what else he/she will come to believe.

When taken in conjunction with Fodor's discussion of modularity, the following architectural conclusion may be drawn.

- The mechanisms mediating our doxastic processes *must* be informationally unencapsulated.

These provide Fodor with a basis for presenting a set of arguments aimed at establishing the pessimistic conclusion. In keeping with the strategy of attempting to disentangle the two argument strains, I will present Fodor's argument for how this strain leads to the pessimistic conclusion.

FROM UNENCAPSULATION TO THE PESSIMISTIC CONCLUSION

The first stage of this argument aimed at establishing that the mechanisms mediating our doxastic processes must be informationally unencapsulated. The second stage aims to establish that because our doxastic processes are not informationally encapsulated, these activities cannot be modeled in computationally feasible terms. Since the aim of cognitive science is to explain the workings of mind in computational terms and our doxastic processes are in principle un-model-able, it follows that we should be pessimistic with regard to the future of a full-fledged cognitive science.[9]

Before proceeding, it might be helpful to revisit the global structure of Fodor's argument in order to better situate the following discussion. The argument begins with the assumption that (1) our doxastic processes are computationally realized. Next, let us grant for purposes of discussion Fodor's conclusion that (2) the mechanisms mediating these processes must be informationally unencapsulated. Third, let us further grant

[9] As opposed to undertaking the "cognitive" science of perceptual systems, which as modularly realized and computationally model-able, we have every reason for being optimistic.

Fodor's claim that (3) we normal human reasoners at least sometimes arriving at conclusions (about what to do and what to believe) in a normatively rational manner. To this set, I suggest that Fodor includes the following to serve ultimately as the principle assumption of a *reductio* style argument (4) that our doxastic processes are model-able.

Fodor offers a battery of arguments at establishing that if (1), (2), and (3) are maintained then there is no means by which (4) can hold. The argument against the possibility of modeling these processes proceeds by cases. When disentangled, Fodor offers rejoinders to three strategies aimed at establishing the model-ability of our doxastic processes.

STRATEGY ONE: EXHAUSTIVE SEARCH

Granting psychological isotropy holds of us and assuming that we at least sometimes actually do arrive at warranted conclusions about what to do and what to believe, it would appear that we must possess some means by which to locate the relevant information in our belief-sets. Perhaps, one might suggest, there is a simple solution to this apparent puzzle and no reason to be pessimistic whatsoever. Perhaps we merely consider exhaustively the totality of our beliefs. So doing might very well provide us with the relevant information that is needed to contend with some problem. Furthermore, there appears to be no reason in principle why such a process could not be computationally modeled – for the exhaustive search of a set is clearly a model-able operation.

Fodor does not entirely disagree for, *prima facie*, there is no reason for thinking that the strategy of exhaustive search (as amenable to algorithmic description) is in principle un-model-able. The following is a synthesis of a number of similar arguments given by Fodor. I suggest that what follows presents both a fair distillation of these arguments and a suitable reconstruction. In chapter 1, I presented some of the findings with respect to the computational complexity of the planning and frame problems. While we need not revisit this in any detail, it is quite clear that these problems are incredibly computationally burdensome. They are in effect beyond intractable. And so, while these problems are in principle solvable by the exhaustive search strategy they are, given the amount of time needed to arrive at a conclusion – even by the most highly idealized computational systems - *in practice* unsolvable. As of yet, however, since exhaustive search is amenable in principle to algorithmic solution (and thus modeling), Fodor has not secured the pessimistic conclusion. The addition, however, of the following provides Fodor what is needed.

- We normal human reasoners arrive at conclusions expeditiously.

Since any system that engages in the exhaustive search of a non-trivially sized set is committed to an incredibly computationally burdensome task requiring a vast amount of time (and/or space) to complete, and it appears that we are rather expeditious decision-makers and belief-fixers, either:

1. The mechanism mediating our doxastic processes does not engage in the process of exhaustive search and evaluation for if it did, given the intractability of the task, we would (likely) never arrive at any conclusion whatsoever since our lives are, in the grand scheme of intractable problems, quite short.
2. The mechanism mediating these processes does, in fact, engage in a process of exhaustive search, in which case, since no one has any idea whatsoever about how to model a computational system that can tract the intractable we are compelled to accept the pessimistic conclusion.

And so, from this we are left with a choice between either denying that the mechanisms mediating doxastic processes engage in the exhaustive search and evaluation of the totality of the belief-set – in which case an alternative strategy is needed if one wishes to maintain that doxastic processes are model-able - or granting Fodor his pessimistic conclusion.

STRATEGY TWO: MASSIVE MODULARITY (OF MIND)

As the issue of computational complexity (and not reliability) ultimately is responsible for the failure of the first strategy, and modular mechanisms are by definition systems that perform tractable computations, perhaps one might suggest that the mind is composed of a number of mental modules. This claim, the *massive modularity of mind* thesis, holds that our minds are composed either largely or entirely of special-purpose modules. As modules, by definition, have access in the course of their processing to only the information present in their proprietary databases, if the set of information in these databases is sufficiently limited, there appears prima facie reason for thinking that such a massively modular architecture might provide a means by which to model the workings of mind.

And so, let us suppose that our cognitive processes are mediated by a set of such mental modules each with its own proprietary database and each engineered to contend with certain kinds of problems. Furthermore, let us assume that the contents of the databases of these modules are sufficiently limited to guarantee the tractability of their operations. Fodor offers a number of arguments aimed at establishing the un-tenability (*i.e.,* "incoherence") of the massive modularity (of mind) hypothesis. All of these, I suggest, are ultimately reducible to three fairly straightforward (counter) arguments. Two will be presented now while the third, since it relies upon the success of the argument from psychological Quinean holism, will be considered in a later section.

MASSIVE MODULAR SYSTEMS NECESSARILY VIOLATE PSYCHOLOGICAL ISOTROPY

In terms of the viability of the massive modularity thesis, Fodor's argument proceeds rather quickly. If our doxastic processes are psychologically isotropic then there are no beliefs that are in principle unavailable during deliberation. Modules are informationally encapsulated systems. Therefore, any putative mental module that engages in belief-fixation must be informationally encapsulated. As such, there must be at least one belief that, while available to the system (*i.e.,* some other module), is unavailable to the particular module under consideration. More simply put, minds cannot be massively modularly realized because, if they were, psychological isotropy would fail to be a property of doxastic processes – which it is.

MODULAR REGRESS AND THE NEED FOR AN UNENCAPSULATED "EXECUTIVE"

Suppose that one denies the claim that psychological isotropy is a property of cognitive processes, opting instead for the claim that minds are massively modularly realized. That is, perhaps it is the case that for every problem domain that we might face we possess a dedicated mental module engineered for arriving at a conclusion in that domain. Granting that complexity concerns will constrain any model proposed, let us further assume that any putative module will perform only tractable computations. At least prima facie, such a model appears capable of providing an account of our doxastic processes in computationally feasible terms. The principle criticism offered by Fodor against this strategy is that the massive modularity of mind thesis either courts a regress or is viciously circular.

Specifically, Fodor concludes that the massive modularity thesis is problematic for less informationally encapsulated "executive" mechanisms need to be posited for the account to be coherent. He explains,

> I think that, assuming that modular mechanisms are ipso facto domain-specific, the idea of a really *massively* – for example, an *entirely* – modular cognitive architecture is pretty close to incoherent. Mechanisms that operate as modules *presuppose* mechanisms that don't. (Fodor 2001, p.71)

And so, suppose, Fodor continues, that our minds are composed of a set of special-purpose modules. As informationally encapsulated, each module performs tractable computations. Such a system would not be prone to the paralysis associated with the exhaustive search strategy. As a trade for tractability, however, such modules are only allowed access to a limited amount of information available to the system as a whole – *i.e.,* their own proprietary databases. Suppose, Fodor continues, that there are two modules M1 and M2 and that M1 and M2 respond to the formal properties P1 and P2 of input representations. The module M1 "turns on" when it encounters the property P1 and M2 "turns on" when it encounters the property P2. It follows, Fodor suggests, that prior to the activation of these modules the properties P1 and P2 must be assigned to the appropriate representations.(Fodor 2001, p.72) It is this requirement that forms the basis of Fodor's objection since it appears that some additional mechanism must be posited the task of which is to assign properties to representations. Specifically, Fodor asks whether "the procedure that effects this assignment [is] itself domain specific."(Fodor 2001, p.72) Putting this another way, Fodor appears to be driving at the following point: if we suppose that the mind is composed of mental modules, some additional mechanism must be posited that makes reliably correct determinations about which of the available modules is to be activated (*i.e.,* relayed an input and "turned on") and when. I will call this the input-routing/module-activation problem.

There are, Fodor suggests, two options for contending with the input-routing/module-activation problem. Either we need to posit a single mechanism that makes such determinations, or a set of mechanisms (*i.e.,* modules) that conjunctively perform this task. The first option, Fodor suggests, must be rejected as the mechanism proposed must be substantially less modular than either M1 or M2. Put another way, the adoption of this option requires the positing of a mechanism that is capable of making reliably correct domain-general inferences which, Fodor concludes, undermines "the thesis that the mind is *massively* modular, that is, that it consists of nothing but systems that are, more or less, all equally domain-specific."(Fodor, 2001 pp. 72-3) The second option raises concerns of a regress. If the determination of which modules are to be activated (*i.e.,* "turned on") is made by a set of modules then there must be a third mechanism posited that arrives at conclusions about which of these (meta)-modules is to be activated. And so, Fodor concludes we are returned to the original puzzle for a third mechanism must once again for posited. Either a single mechanism must be posited that makes such determinations or a set of modules must be posited that performs the (meta-)module activation task. The first horn commits the massive modularist to cede that at least some of the minds activities (and the most "interesting" parts in Fodor's estimation) cannot be modularly realized (and thus not modeled) while the second compels them to explain how the module-activation problem could be adequately contended with by a modularly realized mechanism. Since the massive modularity hypothesis is offered in support of the second option (by denying the first), Fodor concludes that we are returned to the pessimistic conclusion.

STRATEGY THREE: THE HEURISTIC APPROACH

Perhaps, we are possessed of a set of heuristic processes designed to engage in a selective and limited process of search and evaluation. Prima facie, such a strategy for modeling the working of mind seems promising for it would 1) make purchase on tractability by relying upon computationally economical search, stopping and decision rules, 2) arrive at reliability correct (or at least reliably approximately correct) conclusions by heuristically bringing to bear approximately relevant material, while 3) allowing for psychological isotropy to be maintained as a property of doxastic processes. That is, let us suppose that psychological isotropy is a property of doxastic processes and, as such, the interesting activities of mind are unencapsulated. However, let us also assume that we are possessed of a battery of heuristic strategies that expedite the processes of search and decision-making. As heuristics are both computationally economical and, when properly exploited, can be relied upon to produce reliably correct determinations, a massively heuristic model of mind seems quite promising.

For example, barring for now concerns raised by the argument from Quinean holism (to be discussed shortly) such heuristics could be employed to, in effect, locate and bring to bear the relevant material without having to consider each and every belief in the set and without having to posit the existence of partitioned sets of material (*i.e.*, that information is encapsulated as in modular systems). Such a proposal, Fodor notes, "is entirely compatible with the idea that cognition is computational, so long as the course of presumed heuristic calculations is itself locally syntactically determined."(Fodor 2001, p. 42) It is, of course, this caveat that lays the groundwork for Fodor's objection to the heuristic strategy.

Fodor offers a regress argument (reminiscent of that raised against the massive modularity account) against the heuristic strategy. While particular heuristics, he suggests, might very well provide a means by which to explain how particular inferences could be both reliably correct and expeditious, the heuristics strategy as a model of mind can be successful only if an account can be provided of how and when particular heuristics are to be applied. Fodor explains, "if there are to be heuristic solutions to problems about what to do or believe, there must be something that decides which heuristics to use in solving them."(Fodor 2001 p. 41) Clearly, Fodor is here raising a variant of the circularity/regress argument leveled against the massive modularity approach.

And so, along these same lines, for the heuristics approach to be viable, Fodor contends, requires that some mechanism be posited that is capable of contending with what I will term the *heuristic selection problem* – the problem of determining which heuristic from the set of all those available should be employed given the problem posed. The problem here, Fodor contends, is that "the inferences that are required to figure out *which* local heuristic to employ are themselves often abductive."(Fodor 2001 p. 42) This raises a number of concerns.

First, given some belief to be fixed (hypothesis to be confirmed) there are, under the heuristics approach, a battery of potential heuristics that might be employed. In order to determine which heuristic ought to be employed requires that two problems be solved. First, we must have some idea of what would serve as a solution (or approximate solution) to the problem posed. Second, some mechanism must select, from the set of those available, those heuristics that are likely to result in the correct conclusion. Since this requires that a reliably correct and expeditious belief be fixed (an hypothesis confirmed that, for example, heuristic *h* should be employed), we are, Fodor concludes, returned to the initial puzzle. Specifically, relying upon the criticism of the massive modularity hypothesis (as viciously circular or courting a regress), it follows, Fodor concludes, that there can be no (meta)-heuristic solution to the heuristic selection problem. And so, echoing the concerns raised with respect to the coherence of the massive modularity approach as requiring the positing of an unencapsulated "executive," there appears to be, he suggests, no viable means – other than the positing of an unencapsulated and computationally mysterious central processor – to anchor the regress raised by the heuristic selection problem.

And so, Fodor concludes, unless (1) the heuristics approach can adequately heuristically contend with the heuristic selection problem, or (2) the massive modularists can modularly anchor the module-selection problem, or (3) it can be demonstrated that P=NP (and thus that exhaustive search is in practice tractable), there appears to be no viable means by which to model the activities of mind in computationally feasible terms. Since these options, Fodor suggests, exhaust the field and none are capable in principle of providing a framework by which to make comprehensible the operations of belief-fixation, we should, he concludes, be highly pessimistic with respect to the future of cognitive science. While evaluation of this conclusion will be undertaken in the next chapter, before we proceed, a second argument strain in support of the pessimistic conclusion must be presented.

THE ARGUMENT FROM *PSYCHOLOGICAL (QUINEAN) HOLISM*

The argument from *psychological isotropy* aimed at establishing that all beliefs in the belief set are available, thus raising the question of how a process of reliably correct belief-fixation could be made tractable. The argument from *psychological holism*, I suggest, aims at establishing that all beliefs are simultaneously epistemically relevant to determining what else the agent believes. The argument from psychological/Quinean holism, if successful, would establish that since relevance cannot be known *a priori,* it can only be determined by a process of rational consideration and evaluation that is necessarily global in character. Put somewhat differently, the relevance of any given belief to the confirmation of any given hypothesis can only be determined *a posteriori*. If successful, the argument from psychological holism would effectively render any strategy that did not rely upon exhaustive search and rational consideration inconsequential, non-responsive and irrelevant. Specifically, Fodor concludes, since we have no idea how to model something that engages in expeditious exhaustive search and since exhaustive search and rational consideration appears to be the only means by which relevance could possibly be determined, the pessimistic conclusions follows straightaway.

There appear to be two versions of the argument from psychological (Quinean) holism. The first suggests that *some* of the conclusions that we reach require the exhaustive consideration of the belief-set, while the second appears to make the significantly stronger claim that the fixation of *all* warranted beliefs requires exhaustive consideration. I will consider each in turn.

Before considering either, however, I must present some background material. Specifically, we need some idea of what Fodor means by the (normative/descriptive) claim that confirmation is Quinean. By Quinean, Fodor means "the degree of confirmation assigned to any given hypothesis is sensitive to properties of the entire belief system; as it were, the shape of our whole science bears on the epistemic status of each scientific hypothesis." (Fodor 1983 p.107) He continues,

> The point about being Quinean is that we might have two astrophysical theories, both of which make the same predictions about algae and about everything else that we can think of to test, but such that one of the theories is better confirmed than the other – *e.g.,* on the grounds of such considerations as simplicity, plausibility, and conservatism. The point is that simplicity, plausibility, and conservatism are properties that theories have in virtue of their relation to the whole structure of scientific beliefs *taken collectively*. (Fodor, 1983 p.108)

Argument One: Simplicity, Plausibility and Conservatism

> It's part of rationality to prefer the simpler of two competing beliefs, ceteris paribus; and likewise, it's part of practical intelligence to prefer the simpler of two competing plans for achieving a goal. That appeals to simplicity are ineliminable in scientific reasoning is practically axiomatic. But it would seem equally clear that comparing the relative simplicity of candidate beliefs, or of candidate plans of action, is routinely a part of reasoning in quotidian decisions about what one ought to think or do. (Fodor, 2001 p. 25)

Quite broadly, this argument is aimed at establishing that the various regress concerns discussed above can never in principle be suitably anchored. This (weaker) version of the argument from Quinean psychological holism proceeds in the following manner.

1. Arriving at conclusions in a normatively rational manner *often* depends upon our identifying properties of simplicity, plausibility and conservatism.

2. Determinations of simplicity, plausibility and conservatism cannot be made locally and depend upon the totality of one's beliefs.

3. We, at least sometimes, arrive at warranted conclusions about what else we should come to believe or do.

4. Therefore, we at least sometimes make rational determinations of simplicity, plausibility and conservatism.

5. We must then possess doxastic mechanisms whose operations do, at least sometimes, depend upon (*i.e.,* access or consider) the totality of our beliefs.

6. Thus, we do, at least sometimes, in the course of determining what we should believe, consider the totality of our beliefs.

And so, if the mechanisms mediating doxastic processes do at least sometimes arrive at conclusions that require the consideration of the totality of the belief-set, then, Fodor concludes, we are led straightaway to the pessimistic conclusion for neither the massive modularity approach nor heuristics approach will be of any assistance. Modules, by definition, preclude the consideration of all of an agent's beliefs, while heuristics preferentially access particular beliefs by not accessing them all. The only remaining strategy is that of exhaustive search. Since this is a decidedly computationally burdensome task and no one has any idea how to model a system that can "tract" an intractable task (as expeditious exhaustive search would require), we are led to the pessimistic conclusion.

ARGUMENT TWO: THE HOLISTIC NATURE OF RELEVANCE DETERMINATION

While the above, if successful, would establish that at least sometimes we must be considering the totality of our belief-sets in arriving at a conclusion, I suggest that Fodor's argument in fact aims to show that we must be considering the totality of our belief-sets *whenever* we arrive at a warranted conclusion. Earlier I noted that Fodor relies upon the following rationality claim:

- (It appears that) arriving at conclusions rationally requires that all of the evidence that is both relevant and available be considered.

The argument from psychological isotropy aimed at establishing that the entire contents of our belief sets are available to us in the course of deliberation. This stronger version of the argument from Quinean holism aims at securing the claim that *all* beliefs in this set are epistemically relevant. If Fodor can establish this, the pessimistic conclusion would follow immediately for, once again, no one has any idea how to model the working of a formal system that expeditiously undertakes exhaustive search.

Before setting out the argument from psychological Quinean holism, I will present a quotation from Quine's "Two Dogmas of Empiricism." So doing will provide some needed background about just what it is that Quinean holism entails. Quite writes,

> The totality of our so-called knowledge or beliefs, from the most casual matters of geography and history to the profoundest laws of atomic physics or even of pure mathematics and logic, is a man-made fabric which impinges on experience only along the edges. Or, to change the figure, total science is like a field of force whose boundary conditions are experience. A conflict with experience at the periphery occasions readjustment in the interior of the field. Truth values have to be redistributed over some of our statements. Reevaluation of some

> statements entails reevaluation of others, because of their logical interconnections – the logical laws being in turn simply certain further statements of the system, certain further elements of the field. … But the total field is so underdetermined by its boundary conditions, experience, that there is much latitude of choice as to what statements to reevaluate in the light of any single contrary experience. No particular experiences are linked with any particular statements in the interior of the field, except indirectly through considerations of equilibrium affecting the field as a whole. (Quine, 1980 pp. 42-3)

And so, the "web of belief" model proposed by Quine suggests that the relevance of a belief to some proposition under consideration can be neither determined in isolation nor *a priori*. Since there is no means by which to determine relevance in isolation that is, removed from the epistemic fabric of which it is part (*i.e.,* locally), and since relevance cannot be known beforehand, the entire belief-set must be evaluated for the relevance of even one belief (to some hypothesis under consideration) to be determined. Fodor notes,

> The Duhem/Quine point about globality of relevance has to do with [this]: You can't decide a priori which of your beliefs bear on the assessment of which of the others because what's relevant to what depends on how things are contingently *in the world* which in turn depends on how God put the world together. (Fodor, 2001 p. 32)

Given this global or holistic structure and the commitments that this places upon confirmation and relevance determination, Fodor continues, "there is typically no way to delimit a priori the considerations that may be relevant."(Fodor, 2001 p. 37) If so, Fodor concludes, the pessimistic conclusion follows straightaway for neither the massive modularity nor heuristics approaches will be of any assistance. Specifically, neither of these approaches will be of aid because both rely upon determinations of relevance being able to be made locally and prior to exhaustive consideration. Put somewhat differently, both approaches rely for their successes upon the assumption that what is relevant (to some hypothesis/task/problem) can be determined a priori and/or locally – that is without the need to consider exhaustively the contents of the belief-set.

As these strategies are in principle unfeasible, Fodor concludes, we are returned immediately to the original puzzle since it would appear that only the exhaustive search strategy could possibly satisfy the conditions sets by the Fodor/Duhem/Quine model of relevance determination. That is, if relevance cannot be determined *a priori* and nor can it be determined locally, then only a mechanism that *actually* engages in global exhaustive search and evaluation of the belief-set would suffice.

> There's a familiar dilemma: Reliable abduction may require, in the limit, that the whole background of epistemic commitments be somehow brought to bear in planning and belief-fixation. But feasible abduction requires, in practice, that not more than a small subset of even the relevant background beliefs is actually considered. How to make abductive inferences that are both reliable and feasible is what they call in AI the frame problem. [Furthermore] classical architectures know of no reliable way to recognize [relevance] short of exhaustive searches of the background of epistemic commitments. (Fodor, 2001 pp. 37-8)

This (stronger) variant of the argument from psychological/Quinean holism concludes the following: Since neither the massive modularity of mind thesis nor the heuristics approach are responsive, given the holistic nature of relevance determination *only* a mechanism of belief-fixation that actually engages in exhaustive search could possibly arrive at reliably correct determinations of relevance. Any system that engages in exhaustive consideration, however, faces a computationally troublesome task. Specifically, any system that expeditiously makes reliably correct determinations of relevance must be possessed of some way to make tractable an intractable problem. Since we have no idea how to model such a system and we are just such a system, we should be pessimistic with respect to the future of cognitive science.

REINTEGRATING THE TWO STRAINS

The argument from psychological isotropy aimed at establishing that, even if relevance could be readily determined, the massive modularity and heuristics approaches are problematic, for both must anchor a regress in order to be tenable. The argument from psychological Quinean holism aims at establishing that relevance can be determined neither locally nor *a priori* and as such, it follows that the regress problems faced by the massive modularity and heuristics approaches can never be anchored by anything other than a domain-general and global/holistic central system type mechanism. Since this kind of mechanism is in principle un-model-able, we should, Fodor concludes, be pessimistic about the future of cognitive science.

The aim of this chapter has been primarily exegetical. I will use this exegesis as the basis of my discussion, in Chapter Three, of some arguments and positions that might occur to readers of Fodor who understand his pessimistic conclusion in the same way as I.

CHAPTER 3: RATIONAL SYSTEMS AND FODOR'S PESSIMISM

Fodor (1983, 1987, 2000) argues for a rather pessimistic conclusion with respect to the future of cognitive science. This pessimism ultimately reduces to the claim that, though our doxastic processes are computationally realized, the interesting activities of mind (*i.e.,* belief-fixation, decision-making, problem-solving and higher-level cognition in general) are not amenable to classical computational modeling. And so, because modeling appears to be the only means by which we may come to understand the workings of our doxastic processes, we should be entirely pessimistic regarding the future of cognitive science. Fodor's argument for this conclusion is fairly straightforward and may be reduced to the following:

1. We, at least sometimes, arrive at conclusions rationally.
2. Arriving at conclusions rationally requires that all of the evidence that is both relevant (to some hypothesis) and available (to the system) be considered.
3. Any complex system capable of considering all the evidence that is both relevant and available to it cannot be modeled in computational terms.
4. Therefore, our cognitive processes cannot be modeled in computational terms.

I will present an analysis of this argument and conclude that Fodor's pessimism is unwarranted. This follows because, although it is true that any system capable of arriving at conclusions rationally cannot be modeled in computational terms, this is so for the rather interesting reason that it is impossible to model the operations of a physical impossibility. To this end, I will aim to establish two conclusions.

First, I will argue that underlying Fodor's arguments in support of Premise 2, is an impossibly demanding normative rationality principle. I next argue that this normative rationality condition cannot be satisfied by anything that is finite and physically realized. Since it is reasonable to think that our cognitive processes are finite and physically realized, it follows, by application of *ought implies can* reasoning, that this normative rationality principle is in need of weakening. (Cherniak, 1984 pp. 20, 106, 110, 113) This is significant for Fodor's principal objection to both the massive modularity of mind and heuristics approaches reduces to the claim that such processes are "irrational" and thus cannot be relied upon to model the workings of (a rational) mind. And so, by weakening the demands of the rationality principle, these objections – and thus the argument in support of the pessimistic conclusion - are undermined.

Second, I'll argue that while finite and physically realized rational systems cannot be modeled in computational terms, because they cannot physically exist, we have no reason for thinking *our* cognitive processes to be rational in that sense. That is, I will argue that the descriptive claim made by Fodor in Premise 1 above, as to the nature of our cognitive processes, is false. Given Fodor's construal of the demands of rationality, it follows that we never arrive at conclusions rationally. As such, whatever problems there are with modeling the operations of rational, finite and physically realized systems, these problems need not be and are not ours. If not, then, there are no compelling reasons of principle for thinking *our* cognitive processes to be unmodelable.

AN OUTLINE OF FODOR'S (NORMATIVE) ARGUMENT

Reliably arriving at correct conclusions requires that determinations of relevance be made. Drawing upon the Quinean model of confirmation, the relevance of a piece of evidence, Fodor contends, cannot rationally be determined a priori, nor can it rationally be determined locally; that is, in isolation from the theoretical web or fabric of the system's prior epistemic commitments. As confirmation is Quinean, the relevance of any piece of evidence (given some hypothesis under consideration) can only be rationally determined both intra-theoretically (*i.e.,* globally) and a posteriori.

As relevance can only be determined globally and a posteriori, any system that arrives at conclusions by considering less than all of the available evidence necessarily considers an "arbitrary" sample – that is, less than all – of the available evidence and, as such, is "ipso facto irrational." (Fodor, 1987 p.139) As rational systems should consider all of the evidence and as any subset of the evidence considered will necessarily be arbitrary, only a process that considers all of the available evidence could be a rational one. Underlying Fodor's argument, then, is the following normative rationality principle:

- A conclusion is reached rationally if and only if all of the evidence that is both relevant (to the hypothesis under consideration) and available (to the system) is considered.

Having set out the rationality condition underlying Fodor's account, let us next consider a psychologically isotropic cognitive system. Such systems are those in which the totality of the contents of the system's belief-set is available to it in the course of its deliberations. In psychological isotropic models there are no beliefs/evidence that are architecturally encapsulated or otherwise partitioned from and thus unavailable to the system in the course of deliberation.[10] As i) all of the material available to the system (*i.e.,* its entire belief-set) could be relevant, and ii) relevance, Fodor contends, can neither be determined a priori nor locally (*i.e.,* non-globally), any psychologically isotropic system must consider the totality of its belief-set for a determination of relevance to be made rationally. We may now restate Fodor's normative rationality condition as:

- A conclusion is reached rationally if and only if all of the evidence available to the system is considered.

[10] Isotropy, Fodor holds, is a property of "rational" confirmation and thus of any rational system capable of engaging in such activities. By "psychological isotropy" I mean only that this property holds descriptively of some individual cognitive process (*i.e.,* with respect to access to the contents of some system's belief-set).

From this, it follows that:

- A system is rational if and only if it is capable of, in the course of arriving at conclusions, considering all of the evidence available to it.[11]

As discussed in the previous chapter, Fodor's argument for the pessimistic conclusion proceeds by cases. While we need not revisit these arguments in detail, Fodor rejects the heuristics and massive modularity of mind proposals *tout court* because both are "irrational" strategies. Modules, by definition, consider only a sub-set of the evidence available to the system and thus necessarily arrive at conclusions by considering an "arbitrary" sample of the evidence. Since modules, as encapsulated, can only ever consider an arbitrary sample of the evidence available, they are *ipso facto* "irrational."(Fodor 1983, 1987, 2000). And so, because modules cannot satisfy the normative rationality condition they cannot, Fodor suggests, be relied upon to model the operations of mind. Likewise, heuristics (as reliant upon search, stopping and decision rules) bring only a subset of the available evidence to bear without considering the totality of the evidence. Since any sub-set brought to bear is "arbitrary," (with respect to the totality of the available evidence) heuristics too are "irrational" processes. And so, because heuristics cannot satisfy the normative rationality condition they cannot, Fodor suggests, be relied upon to model the operations of mind.

[11] There is an external/wide and internal/narrow reading of "availability." The external construal suggests that as the system is capable of seeking out evidence (*e.g.,* looking in the library) -- evidence that is, while not directly held by the system, broadly "available" to it. The internal construal suggests that the system need only consider the current contents of its belief-set. As the external construal would make Fodor's rationality principle impossibly demanding – as there is little that is not "available" to us in this manner – and as such the argument for the pessimistic conclusion would thereby become a strawman, I'll assume that rationality as construed by Fodor requires "only" that the contents of the system's belief-set be considered. So doing, however, does raise the concern, as Cherniak (1986) rightly notes, that such rationality conditions are in some respect too weak to really qualify, as such principles do not require the system to seek out evidence that, by external standards, it "should have known."

Underlying Fodor's arguments, then, is a particular normative rationality condition. Specifically, these proposed strategies for modeling the operations of mind are rejected precisely because they cannot arrive at conclusions rationally - in Fodor's sense. More precisely, these approaches are dismissed because Fodor holds that since we *at least sometimes* do, in fact, satisfy the normative rationality condition (*i.e.,* we at least sometimes do arrive at conclusions "rationally" – in Fodor's sense) and these approaches are "irrational" (in Fodor's sense), they cannot be relied upon to model the workings of our minds.

Given the role of the normative rationality condition in Fodor's account, it is necessary to first explore what must hold of any system capable of satisfying its demands. To this end, I will offer a set of arguments aimed at establishing two points. First, I will argue that a truly rational (in Fodor's sense) system cannot arrive at conclusions of any kind. Second, I will argue that if there are systems that can, in fact, arrive at conclusions rationally (in Fodor's sense), then these cannot be finite and physically instantiated.

CONTENDING WITH THE IN PRINCIPLE CHALLENGE

In this section I will aim at establishing the following:

- If a system is a rational one, that is, if it goes about attempting to arrive at conclusions rationally (in Fodor's sense), then either it will not be able to arrive at conclusions at all, or, if it can, then such a system must not be finite and physically realized.

Since we normal humans do arrive at conclusions, it would follow that either our cognitive and doxastic processes are not rational (*i.e.,* they are "irrational" in Fodor's sense) or that they are not finite and physically realized. In either case, I suggest, we have reason for rejecting Fodor's descriptive claim - that we, at least sometimes, arrive at conclusions rationally and thus instantiate (at least sometimes) rational systems. Since

Fodor's argument relies upon this descriptive claim, if we have reason for thinking it false, it follows that his pessimism with respect to the model-ability of *our* cognitive processes is unwarranted.

THE DEMANDS ON RATIONAL SYSTEMS

> "Rationality is a normative property; that is, it's one that a mental process ought to have." (Fodor, 2000 p. 19)

I will present two related arguments aimed at establishing that while the third premise of Fodor's argument (i.e, that any complex system capable of considering all the evidence that is both relevant and available to it cannot be modeled in computational terms) is true, it is so for a reason other than that offered by Fodor. Fodor claims that the un-model-ability of rational systems is due to concerns of computational complexity and thus to the *practical* (as opposed to in principle) impossibility of modeling the activities of such systems.[12] I will argue that, given what must hold of anything capable of satisfying Fodor's normative principle, rational systems cannot be modeled because either they cannot arrive at conclusions at all, or, if they can, then they cannot be finite and physically instantiated. Put another way, I will suggest that the activities of rational systems are *in principle* (as opposed to "merely" *practically*) un-model-able. That Fodor's rationality principle is either unsatisfiable or satisfiable only by something that

[12] Fodor suggests that the problem of rational exhaustive search is in principle computable, but in practice intractable (*i.e.,* at least in the complexity class NP.) Intractable tasks are those that are amenable to algorithmic solution and as such, they are amenable to modeling in computational terms. Relying upon an exhaustive search algorithm, however, to arrive at a conclusion when faced with particular kinds of problems (and thus particular problem-spaces) will require an inordinate amount of time. And so, such algorithms can in principle be relied upon to arrive at correct conclusions but only if they are allowed an indefinitely large amount of time to do so. As the amount of time necessary for a conclusion to be reached rationally may be excessive (*i.e.,* exponential-time or worse and may require more than the amount of time remaining until the heat-death of the universe) while an exhaustive search algorithm can in principle be relied upon to arrive at correct conclusions, it cannot in practice be relied upon to do so.

is infinite and physically unrealizable, presents a dilemma, the implications of which I will explore.

CONFIRMATIONAL REGRESS AND THE PARALYSIS OF RATIONAL SYSTEMS:

Fodor claims that for a system to arrive at conclusions rationally, all of the evidence that is both relevant (to the confirmation of some hypothesis) and available (to the system) must be considered. (Fodor, 1983, 1987, 2000) Any system capable of arriving at conclusions rationally must be capable of making determinations of relevance. As with any determination, conclusions with respect to the relevance of a piece of evidence can be reached in one of two ways – "rationally" or "irrationally." Relevance, Fodor claims, cannot be determined *a priori* nor (since confirmation is Quinean) can it be determined locally – that is, in isolation from the system's background set of epistemic commitments (*i.e.,* its belief-set). As confirmation is Quinean there is no non-arbitrary means by which to arrive at reliably correct conclusions about relevance. And so, Fodor concludes, only a rational process could arrive at reliably correct determinations of relevance.[13]

Arriving at conclusions or determinations of relevance rationally, however, requires the consideration of *all* of the relevant and available evidence. Since *all* the available material is relevant material, the system must consider the totality of its belief-set before rationally making a determination of relevance.[14]

Given the Quinean model of confirmation relied upon by Fodor and Fodor's normative rationality principle, it follows that any system that has actually arrived at a conclusion rationally (in Fodor's sense) must have already succeeded in rationally

[13] This follows, for no relevance confirmation metric can rationally be fixed *a priori* or locally.

[14] I take Fodor's (Quinean) point here to be the following: since relevance can only be determined *a posteriori*, everything is *a priori* relevant.

arriving at an indefinitely large number of conclusions with respect to the relevance of each of its beliefs to the hypothesis under consideration. However, each of these individual conclusions, if reached rationally, requires that the system has already succeeded in rationally arriving at an indefinitely large number of conclusions with respect to the relevance of each of its beliefs to this hypothesis. However, arriving at a conclusion rationally with respect to the relevance of each belief to each of these hypotheses requires that an indefinitely large number of relevance hypotheses be considered. And so on.

Specifically, then, I am suggesting that Fodor's rationality principle is untenable for it requires that a rational system contend (rationally) with a regressive and non-terminating series of rational relevance determinations. By way of example, consider some hypothesis H_1 under consideration by a rational system. For the system to arrive at a conclusion rationally it must consider all of the evidence that is both i) relevant to the confirmation of H_1 and ii) available to it.

To determine what such "consideration" might entail, given the nature of rational systems, take any piece of evidence B_1 available to the system. One of two things is true of B_1: Either i) B_1 is relevant to the confirmation of H_1 or ii) it is not. Since Fodor claims that relevance cannot be determined *a priori* nor locally (that is, beforehand and in isolation from the system's entire web, fabric, or set of background epistemic commitments) it follows that "consideration" of B_1 can be given either rationally or "irrationally." For consideration to be given rationally (*i.e.,* for a determination of relevance to be made) it must not be given arbitrarily or "on the cheap." Therefore, a relevance hypothesis RH_1 – with regard to the relevance of B_1 to H_1 – must itself be considered and intra-theoretically confirmed.[15] Before a conclusion can be reached rationally with respect to H_1, the relevance of B_1 must be determined, which requires that a relevance hypothesis be confirmed intra-theoretically with respect to the system's entire set of prior background epistemic commitments (*i.e.,* its belief-set.) And so, to arrive at a conclusion rationally with respect to H_1, the system must first arrive at a conclusion rationally with respect to RH_1.

Now take any piece of evidence B_2 that is available to the system. Either B_2 is relevant to the confirmation of RH_1 or it is not. Without restating the above, it follows that for the relevance of B_2 (to RH_1) to be determined rationally the system must "consider" B_2 with respect to RH_1 and do so intra-theoretically, that is, from within and with respect to the system's set of background epistemic commitments. If this conclusion (*i.e.,* the first relevance hypothesis) is to be reached rationally, a second relevance hypothesis RH_2 – with respect to the relevance of B_2 to RH_1 – must be "considered" and intra-theoretically confirmed. And so on….[16]

Given that the system under consideration is a rational one, it is faced with a paradoxical and impossible task. Specifically, it must contend with (*i.e.,* rationally anchor) a regressive chain of rational hypothesis confirmations that cannot, in principle, be rationally anchored. And so, any system that attempts to arrive at conclusions rationally will be caught in this ensuing regress and unable to arrive at conclusions at

[15] The point here is that "consideration" in a rational system cannot simply be given to a piece of evidence "on the cheap" by the system by merely "accessing" the piece of evidence nor can "consideration" entail that the piece of evidence be placed near the hypothesis for its relevance to become apparent. Rather, the system must arrive at a conclusion rationally (in Fodor's sense) with respect to the relevance of some piece of evidence. It is because a conclusion must be reached – and reached rationally - if the system is to be a rational one – that the regress is able to take hold.

[16] The argument offered here can be run in either direction, for by assuming a conclusion to have in fact been reached rationally, it follows that an infinite number of relevance hypotheses must have been rationally confirmed.

all. As I shall show in more detail later, any system that attempts to arrive at conclusions rationally (*i.e.,* by satisfying Fodor's normative rationality condition) will be hopelessly cognitively paralyzed – unable to anchor/halt the endless regress of relevance hypothesis confirmations.

Since we do, in fact, arrive at conclusions, it follows that we cannot – even sometimes - be arriving at these conclusions rationally (*i.e.,* we cannot satisfying Fodor's normative rationality principle). Since we cannot - even sometimes - arrive at conclusions rationally, we cannot therefore be instantiations of a Fodor-style rational system.

One may object that I have overstated Fodor's position in claiming that he is committed to the view that *all* material is relevant material. Rather, the objection continues, Fodor's claim is only that any particular belief *could* be relevant. I will consider two responses here.

If Fodor is making the claim that all material *could* be relevant and that, for example, the possibility is real that one's botanical beliefs are not, in fact, confirmationally relevant to a particular astronomical one, a number of points bear mentioning. First, there would be no way of knowing *a priori* which beliefs are relevant and which are not. Therefore, since any particular belief *could* be confirmationally relevant, for one to be rational (in Fodor's sense) there would be no option but to actually determine whether or not it, in fact, is relevant. This, of course, requires that the system "consider" and consider rationally this bit of information. Second, since this would hold for each belief in the set, one must, in order to be rational, consider them all. And so, even if Fodor is claiming only that some material is, in fact, never confirmationally relevant to any hypothesis under consideration, any rational system must still consider and consider rationally the totality of its belief set.

Understood in this manner, all available material is relevant material because the irrelevance of a piece of information to some particular hypothesis under consideration would be, given both the system's aim of determining what else it *should* believe and its inability to know *a priori* which things confirmationally bear on other things, quite relevant. For example, that one should come to determine that one's botanical beliefs are, in fact, irrelevant to one's astronomical ones, would be, as knowable only *a posteriori*, quite relevant to the task of determining what else one *should* believe for one now knows that what one believes about matters botanical is in this instance confirmationally irrelevant.

However, there is reason for thinking that Fodor's reliance upon the Quinean model may commit him to the stronger claim that, in fact, *all* available material is relevant. The very structure of the Quinean web of belief model, whereby belief-fixation is confirmationally dependent upon the holistic equilibration of the entirety of the web – and thus upon the totality of the beliefs in the set – suggests that *all* of a system's beliefs are relevant to determining what else it should come to believe. That is, given that belief sets are both isotropic and exhibit the web structure proposed by Quine, it would follow that each belief in the set is either *directly* or *indirectly* confirmationally linked to all other beliefs in the set. Put another way, any particular belief in the set bears confirmationally on every other belief in the set.[17]

Discussion has been so far somewhat abstract and perhaps an example will better serve the point. Drawing upon this model then, let us consider some botanical

[17] See Chapter 2, pp. 60-61 for direct quotation from Quine regarding this point.

belief (B) that is both located somewhere at the periphery of the field/web of the system's belief set. This suggests that this particular belief serves to confirm those beliefs "below" it or those beliefs that are more "central" than it, as well. (*e.g.,* if B violated these central beliefs than either it would not have been accepted or the very structure of the web itself would have been readjusted to accommodate it). Assuming the system to be an isotropic one, that is assuming that there are no beliefs or sets of beliefs that are partitioned from other sets, given the Quinean model, then, this particular botanical belief B would serve to, at least in part, *directly* confirm a number of the system's others beliefs. In so doing it would serve also to confirm directly or indirectly those beliefs most "centrally" located. *Indirectly,* then, or transitively *via* the more central beliefs, B would lend confirmational support to all other beliefs in the set (since the belief set is both Quinean and isotropic).

As isotropic and Quinean, the same would hold for every belief in the set, including any of those at the periphery. And so, on the Quinean model, since one's botanical beliefs at least indirectly confirm one's astronomical beliefs and one's astronomical beliefs at least indirectly confirm botanical ones, each is, in fact, relevant to the other. Since this would hold of *any* two beliefs in the set, that is since any two beliefs, as part of the same confirmational "fabric" or web, are mutually, though perhaps indirectly, confirmationally dependent, it would appear that *all* available material (*i.e.,* the entirety of the belief-set) is relevant material.

Furthermore, it would be rather puzzling on the Quinean model if everything the system believed was not relevant, since this would suggest that portions of a system's belief set would be isolated, partitioned or otherwise fractioned off from the web itself.

This in turn would suggest that confirmation/belief-fixation is, at least in some cases, not a *global* activity. It also suggests a model that is non-isotropic.

For example, suppose that botany really is entirely irrelevant to astronomy. For one's botanical beliefs to be entirely irrelevant to one's astronomical ones, it would need to be the case that no matter what we came to believe about botany and no matter what reevaluation and readjustment of the web these might occasion, this would have no bearing whatsoever on what one should believe about matters astronomical. That some finding in botany might result in the reevaluation and readjustment of the web, including perhaps some radical revision of the "central" beliefs, suggests that confirmation cannot be partitioned in this way under the Quinean account, since what we should believe about astronomy is certainly confirmationally dependent upon these central beliefs.

And so, in any particular instance, since *all* beliefs in the set are confirmationally inter-dependent, as the Quinean model suggests, one's botanical beliefs would in fact be relevant to the credence that one should give an astronomical one. From this it follows that, in order to be rational, one must "consider" the totality of the contents of the belief set (*i.e., all* that is available). Put another way, since everything we believe is confirmationally relevant to everything else we believe and all that we should come to believe, rational confirmation would require that everything be considered and considered rationally. Interestingly, the task for the rational system under this interpretation is no longer one limited to determining whether or not some particular belief is relevant, but rather, given that everything the system believes is relevant, the task becomes one of determining *how* or *in what way*, each belief in the set bears on the particular hypothesis under consideration.

THE PHYSICAL UN-REALIZABILITY OF RATIONAL SYSTEMS

The above discussion suggests that any rational system is compelled to consider a series of relevance hypotheses intra-theoretically. Since, Fodor claims, there is no non-arbitrary means by which to fix relevance[18] other than by doing so rationally, any rational system *en route* to arriving at any conclusion must first arrive at an indefinitely large and non-terminating number of intermediate conclusions. In this respect, Fodor's normative rationality standard is highly idealized for it requires the rational "consideration" of (*i.e.,* that conclusions be reached rationally with respect to) an indefinitely large and non-terminating set.

Cherniak (1986) offers a number of arguments aimed at establishing that no finite system could satisfy such idealized rationality conditions. While the particular idealization underlying Fodor's account is slightly weaker than that directly considered by Cherniak – for Fodor's account is idealized "only" insofar as the system must consider a regressive and non-terminating (*i.e.,* infinite) set – Cherniak's arguments and conclusions remain applicable. Since any system that is able to arrive at conclusions rationally must be capable of rationally considering a non-terminating (*i.e.,* infinite) set, it follows that such a system must be capable of undertaking an infinite number of relevance determinations in a finite amount of time. This follows for one of two reasons. Either (i) if a rational system could succeed in eventually ending the regress of rational consideration (*i.e.,* confirming a relevance hypothesis) then there must have been at least

[18] Given Fodor's (1983, 1987, 2000 pp. 63-4) discussion of *Hamlet's* Problem – the problem of non-arbitrarily (*i.e.,* rationally) determining when the evidence considered is enough or the problem of non-arbitrarily (*i.e.,* rationally) determining when to stop thinking about something – he appears to, in some way, acknowledge the potential for such a regress. Though, as noted earlier, Fodor takes this to be a practical engineering problem and not a problem *in principle* for rational systems.

one additional belief (*e.g.,* that a is relevant to b) added to the belief set. This would hold for each relevance hypothesis confirmed. And so, returning to the regress set out above, things look even worse for the rational system, since the set of beliefs that would need to be considered would be ever-increasing. That is to say, each relevance hypothesis confirmed "along the way" would yield another belief that would need to be considered and considered rationally. Or (ii) if each newly confirmed relevance hypothesis is *not* added to the belief set, then the system would be compelled to endlessly reconfirm each relevance hypothesis. And so, in the first case, the number of relevance hypotheses that would need to be confirmed would be ever-increasing (*i.e.,* as new beliefs are added to the set). In the second case, the number of relevance hypotheses that would need to be confirmed would also be ever increasing (*i.e.,* because the system must endlessly reconfirm each relevance hypotheses).

By way of example, considering even a very small set of only 2 beliefs (a,b) and an hypothesis under consideration (c) provides a suitable toy example. First, relevance hypotheses would have to be initially considered and confirmed for the product of a,b and c. Assuming that the conformational regress set out could be anchored, each of these "considerations" would result in the creation of a new belief (*e.g.,* that a is relevant to c, and so on). Running through the entire set once would *at least* double the number of beliefs in the set. In order to arrive at any conclusion rationally, the system would now need to consider the relevance of each of these and so on. Alternatively, if these are not appended to the systems belief set, then, to arrive at any conclusion rationally it would need to endlessly (re)consider and (re)confirm each relevance hypothesis. Either way, the rational system is faced with considering a non-terminating (*i.e.,* infinite) set. And so, in keeping with Cherniak's conclusion, it would appear that any system that is actually capable of arriving at conclusions rationally (in Fodor's sense) cannot be a finite one.

That rational systems cannot be finite, while of interest and potentially troublesome, is not yet problematic. But, when taken in conjunction with the claim that anything that is physically realized (instantiated) must necessarily be limited and finite, the conclusion poses a problem. Specifically, it would seem to follow that any system capable of arriving at conclusions rationally (in Fodor's sense) cannot be physically realized. To put this point another way, nothing physically realized could instantiate a (Fodor-style) rational system. To the extent that the operations of a rational system cannot be physically instantiated (*i.e.,* they are physically unrealizable) it follows straightaway that they cannot be computationally modeled either.[19]

A DILEMMA

The above discussion suggests one of two conclusions: i) given the apparent regress with which it must be contend, no rational system could arrive at conclusions at all, or, ii) if we suppose there to be rational systems that arrive at conclusions, then such systems can be neither finite nor physically realized/realizable. The normative rationality principle, when taken in conjunction with Fodor's descriptive claim (that we at least sometimes arrive at conclusions rationally) presents the following dilemma or paradox.

- Fodor's normative rationality condition is maintained and the descriptive claim – that we do at least sometimes arrive at conclusions rationally – is also maintained, in which case it would follow that we at least sometimes satisfy Fodor's normative rationality condition. Given that the normative condition cannot be satisfied by anything that is finite and physically realized, it follows that at least sometimes our cognitive processes are neither finite nor physically realized.

[19] That such operations are unmodelable follows by way of the *Church-Turing* thesis since there are no conclusions that could be reached by a rational system in a finite number of steps or operations. (*C.f.,* Cherniak, 1984 pp. 15, 78)

- Fodor’s normative rationality condition is maintained and physicalism is assumed true, in which case it would follow that Fodor’s descriptive claim is false as neither we nor any other finite and physically realized system could ever arrive at conclusions rationally.

Adopting the first horn of the dilemma appears to come at the price of a commitment to mystery interactive dualism for it follows from this that if we (at least sometimes) arrive at conclusions rationally, then our minds must (at least sometimes) be infinite and physically unrealizable.[20] Assuming that some brand of physicalism is more plausible than mystery interactive dualism, there appears little reason to maintain an unrealizable and unsatisfiable normative rationality condition.

Adopting the second horn suggests that, whatever may be true normatively about the way we should go about arriving at conclusions, if our cognitive processes are physically realized then our cognitive systems are not and cannot be rational (in Fodor’s sense). Because we have good reason to think that our cognitive processes are physically realized, we have reason for rejecting the normative rationality principle as unsatisfiable.

There is a third option here that I have purposefully left unconsidered. Connectionism, quite interestingly, seems best to model Fodor’s beliefs about the holistic character of cognition. However, while radical connectionist networks, Fodor convincingly argues, provide a means by which to accommodate the properties of Quineanism and isotropy, it is “simply hopeless” (Fodor, 2000 p. 47) as a model of the workings of mind and should be rejected on the grounds that it cannot account for

[20] As interactive dualists generally maintain that human (as opposed to divine) minds are finite, Carruthers raises the interesting point that adopting this horn means accepting that we are (in this one respect) divine. *C.f.* Cherniak (1994) p. 94 for initial discussion of the claim that “we do not have God’s brain.”

"patent truths about the productivity, systematicity, and compositionality of typical cognitive systems." (Fodor, 2000 p. 50) He continues,

> Here, as so often elsewhere, networks contrive to make the worst of both worlds. They notoriously can't do what Turing architectures can, namely, provide a plausible account of the causal consequences of logical form. But they also can't do what Turing architectures can't, namely, provide a plausible account of abductive inference. It must be the sheer magnitude of their incompetence that makes them so popular. (Fodor, 2000 p. 47)

Since Fodor clearly finds connectionism anathema on a number of other grounds, and we do appear to have good reason for thinking such proposals problematic for a number of the reasons he has set out, I have not included this as a viable third option.

ON FODOR'S DESCRIPTIVE CLAIM

The previous discussion suggested that either we do not arrive at conclusions rationally (and thus are not instantiations of rational systems) or that we are such systems (in which case our cognitive processes must be infinite and physically unrealizable). I will suggest that we have little reason for accepting Fodor's descriptive claim - that we ever arrive at conclusions rationally - and thus little reason for believing our cognitive processes to be infinite and physically unrealizable.

In support of this, I will present an argument similar in form to Fodor's empirically based argument from the history of science outlined in the previous chapter.[21] Specifically, I will suggest that a similar empirical claim may be made with respect to the un-tenability of Fodor's rationality condition - read *descriptively*.

[21] Very briefly, Fodor's argument suggests that any massively modular model of mind proposed must posit the encapsulation/partitioning of discrete sets of information. A commitment of this, he continues, is that there must be beliefs broadly available to the system as a whole that cannot, in principle, be simultaneously entertained – because they are partitioned from each other. Any such model is implausible, he concludes, because a counterexample (from the history of science) can be found in which someone, somewhere at some time actually (in fact) did succeed in simultaneously entertaining the two bits of information that it should not – under some particular modular model proposed – be able to simultaneously consider.

Consider any belief that you hold. I suggest that for any such belief, there is at least one ancillary belief, piece of evidence or bit of information that, while broadly available to you, was not considered in the deliberative process that led to the fixing of that particular belief. For example, I believe that "Manhattan is an island." I have just found at random in my dictionary the word "grommet." Until this moment, while I have had thoughts about Manhattan, about islands and about grommets, I have never brought to bear my beliefs about grommets on the hypothesis that Manhattan is an island. And so, while psychological isotropy demands that any belief is available to be brought to bear, it does not follow from this that all of my beliefs are in fact ever brought to bear.

I take my belief that Manhattan is an island to be warranted. However, in arriving at my conclusions about Manhattan's island status, I failed to bring to bear my beliefs about grommets. I thus failed *ipso facto* to arrive at *that* conclusion rationally (in Fodor's sense). As an empirical matter, I suggest that for *any* belief that I hold and for *any* conclusion that I have reached or may reach, there are/will be at least one belief i) that is available to me and 2) that I failed to bring to bear/consider.

To put the point slightly differently, I am suggesting that, as an empirical matter, for any conclusion reached, there is at least one piece of evidence that is both available to me and that would, because I failed to consider it, "surprise" me were it brought to my attention by some external source (*e.g.,* by means of random dictionary searches or other person). The fact that I can be surprised by one of my very own beliefs in the course of arriving at some conclusion/fixing some belief suggests that I could not have considered it in the course of my deliberation. If so, then I could not have arrived at *that* conclusion rationally (in Fodor's sense).

Because I can be surprised by one of my very own beliefs given *any* hypothesis under consideration (*i.e., any* conclusion reached or belief fixed), I cannot have arrived at *any* of my conclusions rationally (in Fodor's sense). And so, it would appear that Fodor's descriptive claim – that we *at least sometimes* arrive at conclusions rationally - is false.[22] As such, it is also false that we ever we *even sometimes* satisfy Fodor's normative rationality condition.

With respect to Fodor's challenge, then, whatever the problems inherent in modeling the operations of rational (in Fodor's sense) systems - and there appear to be many - to the extent that we are not and (if our cognitive processes are finite and physically realized) cannot be instantiations of rational systems, these problems need not be ours.

NORMATIVE RATIONALITY COME WHAT MAY

It is the adherence to both the normative rationality principle and the descriptive claim (*i.e.,* that we, in fact, *at least sometimes* satisfy the normative principle) that undergird both the pessimistic conclusion and Fodor's prescription that cognitive science[23] must wait on the development of a suitably mysterious Quinean theory of computation. (Fodor, 1983, 2000) And so, it appears that Fodor is advocating the adoption of the third horn in the following trilemma.

- We maintain our commitment to computationalism and our commitment to the normative rationality principle, in which case we are compelled to revise our commitment to physicalism.

[22] At the very least, we have reason for thinking the descriptive claim false and absolutely no reason for thinking it true.

[23] Fodor suggests that while we should continue to work on attempting to understand and model perceptual/input processes, as these are amenable to modular description and thus computational modeling, we shouldn't waste any more time trying to understand the working of our doxastic processes. So doing, he claims, would require that we have a suitably Quinean theory of computation and we neither have one of these nor do we have any idea how we might go about getting one.

- We maintain our commitment to computationalism and our commitment to physicalism, in which case we are compelled to revise our commitment to the normative rationality condition.

- As "rationality is a normative property; that is, it's one that mental processes ought to have," we maintain the commitment to the normative rationality condition come what may. And so, if we maintain commitments to both physicalism and to the descriptive claim (that we do at least sometimes satisfy the normative rationality principle), then we are compelled, given the unsatisfiability of this normative rationality principle by any finite, physically realized and computationally modelable system, to revise our commitment to computationalism.

I have argued that we have reason to think both that no finite and physically realized system could in principle satisfy Fodor's rationality condition and that Fodor's descriptive claim is false. These conclusions, when taken in conjunction with Rey's (1997) arguments that computationalism is the most plausible (and perhaps only) hope for our understanding the workings of mind, provide us with good reason for rejecting the third horn of the trilemma. In light of this, and assuming that physicalism is more plausible than mystery dualism, it would appear that we have far more compelling reasons for thinking the second horn the most plausible option.

The aim of this section has been narrow: to suggest that underlying and motivating Fodor's pessimism is an adherence to an unsatisfiable normative principle of rationality. I have suggested that adherence to this principle compels Fodor, in light of its unsatisfiability and paradoxical implications, to ignore other options and call for the rejection[24] of computationalism so that this normative principle can be made satisfiable.

[24] Fodor, at times, (1983, p. 126; p. 129 and 2000 pp. 44-5, p. 71, pp. 77-8, p. 105, p. 108, pp. 111-2, p. 114) does appear to be suggesting that, in light of the paradox caused by the unsatisfiability of the rationality condition given the nature of our current theory of computation, a new theory of computation is needed. Such a theory would be quite unlike anything currently available and would at least prima facie appear to require radical revision of some of our central logical notions. That is, such a new Quinean theory of computation would seemingly be able to explain not only such practical concerns as how intractable operations could be "tracted" but also rather troubling ones like how finite systems could complete an infinite number of operations in a finite amount of time.

Given the options of either i) weakening the demands of the normative rationality principle or ii) come what may revising our commitments to computationalism or physicalism - or both - in order to make satisfiable the normative condition, we have little reason for thinking Fodor to have made a compelling case for pessimism.

The fact that a normative principle cannot be satisfied by any finite and physically realized system militates against our thinking it tenable. The fact that computationalism fails to provide a model of the activities of an infinite and physically unrealizable system should not be a strike against it. Since there is much to recommend in computationalism, and because we have reason for thinking the normative rationality principle undergirding Fodor's argument to be unsatisfiable and thus in need of weakening, we should reject the conclusion that the workings of the "interesting" activities of mind are in principle un-model-able. This follows for once the rationality principle is weakened to one that is actually satisfiable, the *a priori* arguments for rejecting *tout court* the heuristics and massive modularity approaches are no longer supported. Rather, since it would appear that *our* minds are instantiations of an "irrational" (in Fodor's sense) system – for we both arrive at conclusions and fail to do so rationally (in Fodor's sense) – and since the heuristics and massive modularity of mind programs proceed quite generally under the assumption that cognitive process are "irrational" (in Fodor's sense), these approaches appear promising.

ALTERNATIVE MODELS

I argued in the previous section that Fodor's pessimism relies upon both an untenably demanding principle of normative rationality and an unwarranted descriptive claim that we ever arrive at conclusions rationally (in Fodor's sense). From these I concluded that we have no reason for thinking the workings of mind to be on principle un-model-able. Arguing for this point of principle, however, goes only so far in contending with the challenge raised by Fodor. What remains, of course, is for a positive account to be provided of how these processes might be modeled in computationally feasible terms. In response to this rather practical "engineering" challenge of showing how a computational system could undertake the kinds of operations that we know the human mind to be capable (*e.g.*, bringing disparate material to bear, exhibit flexibility and creativity but also systematicity of thought), I will outline some alternative approaches. My aim here is not to provide a complete model of the workings of mind, but rather to show that there is a plausible and computationally feasible alternative overlooked by Fodor.

Before proceeding, however, it will be beneficial to very briefly review Fodor's two principle objections to the massive modularity of mind (and the heuristics approach) hypothesis outlined in the previous chapter.

Fodor's first objection reduces to the claim that minds cannot be massively modularly realized because, if they were, psychological isotropy would fail to be a property of mind – which it is. That is, since our doxastic processes are isotropic, and modules, as informationally encapsulated devices, necessarily violate isotropy, the mind cannot be massively modularly realized.

Fodor's second objection reduces to the claim that no massively modular system could possibly contend with the input-routing or module activation problem (*i.e.,* the problem of how a modular system could possibly make determinations about which other modules are to be routed which information and thus activated or "turned on.") Contending with this puzzle (*i.e.,* making a reliably correct determination about how and where information is to be routed and which modules are to activated), Fodor argues, requires a system that is capable of undertaking holistic/global processes. Since modules are, by definition, informationally encapsulated devices, they are in principle incapable of undertaking operations that consider the totality of the system's belief-set. This is the principal result of what I termed the argument from Quinean holism outlined in the previous chapter. Since, Fodor maintains, only a mechanism capable of global/holistic processing could adequately contend with the input-routing or module activation problem, and modules are incapable of undertaking global operations, the mind cannot, he concludes, be massively modularly realized.

> Here's the moral: Really massively modularity is a coherent account of cognitive architecture only if the input problem for each module (the problem of identifying representations in its proprietary domain) can be solved by inferences that aren't abductive (or otherwise holistic); that is, by domain-specific mechanisms. There isn't, however, any reason to think that it can. (Fodor, 2001 p. 78)

Given this, the challenge for the massive modularity theorist (and, as the objections are conceptually identical, the heuristics-approach theorist) is to provide a model of a computational system that is capable of modularly (or heuristically) contending with the input-routing or module activation puzzle. It is to the task of outlining such a model that I will now turn.

Carruthers (2006) argues for a reinterpretation of the term "encapsulation" – the principal feature of modular systems – that will provide a helpful and accessible starting point for the discussion. Generally, an informationally encapsulated system is one "whose internal operations can't be affected by most or all of the information held elsewhere in the mind."(Carruthers, 2006 p. 58)[25] There is, Carruthers notes, a scope ambiguity resulting in an equivocation with respect to the demands placed on

[25] As Fodor notes, "A module *sans phrase* is an informationally encapsulated cognitive mechanism."(Fodor, 2001 p.58)

encapsulated systems. The modal operator, he explains, can take either a "wide" or a "narrow" scope with respect to the quantifier. Taken "narrowly," informational encapsulation demands that, "concerning most of the information held in the mind, the system in question can't be affected by that information in the course of its processing."(Carruthers, 2006 p.58) This conception of encapsulation – that adopted by Fodor – suggests that there are discrete sets of information that are permanently (*i.e.,* physically) partitioned from each other. Each module, then, has its own proprietary database that is physically partitioned from information held elsewhere in the system. Or, the other way around, there is information elsewhere in the system that is in principle inaccessible and thus unavailable to the module in the course of its processing.

Taken "widely," encapsulation demands that a system be "such that it can't be affected by most of the information held in the mind in the course of its processing."(Carruthers, 2006 p.58) Or, put somewhat differently, taken "widely" encapsulation demands that the module will actually consider only a small sub-set of the information that is globally available to it. Carruthers explains,

> It can be true that the operations of a module can't be affected by most of the information in the mind, without there being some determinate sub-division between the information that can affect the system and the information that can't. For, it can be the case that the system's algorithms are set up such that only a limited amount of information is ever consulted before the task is completed or aborted. Put in this way: a module can be a system that must consider only a sub-set of the information available. Whether it does this via encapsulation as traditionally understood (the narrow scope variety) or via heuristics and stopping rules (wide-scope encapsulation) is inessential. (Carruthers, 2006 pp. 58-9)

Carruthers' point then is that "narrow" scope (*i.e.,* traditional Fodor-style) encapsulation is only one manner in which an operation might be modularly realized. While encapsulation may be undertaken by physically partitioning classes of information (*i.e.,* by physically fixing a module with a limited and proprietary database) it may also be undertaken by processes that heuristically limit the material that a module will actually operate upon. Such "process encapsulation" then serves to limit the set of information considered by the module, not by physically partitioning the set, but rather by limiting the information actually brought to bear for processing. Since it is processes that limit the material actually operated upon by a wide-scope encapsulated device, conceptually, such modules could have a proprietary database that is (with respect to the narrow scope interpretation) identical to the set of information that is globally available to the system as a whole.

Narrow scope encapsulated systems then will definitionally violate the isotropy condition for there is, by stipulation, information that such devices cannot in principle access. For precisely this reason, narrow-scope encapsulated systems cannot undertake global computations. As such, narrow scope encapsulated systems also necessarily violate the Quinean holism condition, for the reasons outlined by Fodor. In contrast, wide-scope encapsulated devices need not violate the isotropy condition for there are no bits of information that are in principle unavailable to them in the course of their operations. By design heuristics (by means of the exploitation of search and stopping rules) bring to bear only some sub-set of the available information. Different heuristics, when applied to the same set of available information, bring different sub-sets of material to bear. And so, depending on the nature of the heuristic processes employed, a wide-scope encapsulated device (unlike a Fodor-style module) could, in principle, be provided with and thus consider during the course of its processing, any particular bit of information held by the system at large – though it will actually consider only some sub-set of this totality (*i.e.,* whatever the particular heuristic exploited actually brings to bear). Put another way, in wide-scope encapsulated systems, while only some sub-set of the available information will actually be considered, there are no bits of information that are, in principle, unavailable. And so, with respect to Fodor's first objection, it need not follow that just because an operation is modularly realized that it must also violate the condition of isotropy.

While narrow-scope encapsulated systems are, for the reasons outlined by Fodor, incapable of arriving at reliably correct conclusions about which of the available material is to be brought to bear, wide-scope encapsulated systems need not be. With respect to the issue of reliability, we have no reason for thinking that we arrive at conclusions rationally (in Fodor's sense) and thus no reason for thinking that we are, in this respect, cognitive *optimizers*. Rather, in light of the unsatisfiability of Fodor's normative rationality condition, I argued that we have reason to weaken this rationality condition to one that is satisfiable by the likes of us. Since we cannot be cognitive "optimizers" in this sense (*i.e.,* we cannot arrive at conclusions rationally in Fodor's sense), it follows that we must be "satisficers" (*i.e.,* we must consider only *some* of the available evidence). That we do in fact deploy processes that are *satisficing* (*i.e.,* that are "good enough" without being optimal) is well accepted among cognitive scientists. With respect then to the heuristic search, stopping and decision-making rules relied upon by wide-scope encapsulated systems, it is reasonable to suggest that evolutionary pressures would have favored those search heuristics (*i.e.,* those heuristics that bring to bear sub-sets of the available information without consider the set exhaustively) which, though irrational (in Fodor's sense), are "good enough."[26] While such evolutionary claims are at least

[26] That such heuristic processes are not optimal – that they exhibit characteristic breakdowns - is also well known. C.f. Kahnman 1982.

prima facie plausible, the issue of the accuracy of such heuristics is, at base, an empirical matter.

With respect to this, Gigerenzer *et al.,* (1999) investigated the reliability, tractability and robustness of a number of search, stopping and decision-making heuristics comparing them to more "rational" (*i.e.,* informationally and computationally intensive) strategies. Both very simple one-cue heuristic stopping/decision rules and complex and computationally intensive rational "benchmark" strategies were considered. Specifically, Gigerenzer investigated how accurate very simple heuristics, that both fail to incorporate all of the available evidence and which rely upon only one cue, could be. Quite surprisingly they found that, with respect to accuracy, the three simple "fast and frugal" heuristics tested outperformed or tied those computationally intensive "rational" strategies that considered all (or most of) the available evidence. (*e.g.,* Dawes' rule, Franklin's rule and multiple linear regression). (Gigerenzer *et.al.*, 1999 p.87) Furthermore, by dramatically reducing the impact of data over-fitting (*i.e.,* over-generalization) which reduces the accuracy of rational strategies, these simple heuristics were found to be quite robust, as well. (Gigerenzer *et. al.,* 1999, pp. 109-110, pp. 127-136) And so, contrary to the common lay assumption that reliability is proportional to the amount of information considered (*i.e.,* that "more is better") we have reason to think that very simple heuristics can be surprising effective.

Underlying Fodor's conclusion is the assumption that only a system capable of arriving at conclusions rationally (in Fodor's sense) could arrive at reliably correct conclusions. But, Gigerenzer *et al.,* findings suggest that heuristics are reliable processes for arriving at correct (*i.e.,* satisficing) conclusions. While irrational processes (in Fodor's sense) they are "good enough" and "good enough" is good enough given that we have every reason for weakening Fodor's impossibly demanding rationality principle.

That heuristic processes are expeditious and robust (*i.e.,* they generalize without over-fitting) and are nearly, as, or more reliable than "rational" strategies, is of interest. However, in order for the heuristics approach to be a viable alternative, it must contend with the heuristic variant of the input-routing problem – what I have termed the heuristic-selection problem (*i.e.,* the problem of when to employ which heuristic).

With respect to the proposed meta-heuristic regress that Fodor claims any massively heuristic system must contend (*i.e.,* the heuristic-selection problem discussed in Chapter 2), there is no reason to think that heuristics and thus meta-heuristics could not be organized such that particular search, stopping, and decision meta-heuristics are exploited to determine which (1st order) heuristics are to be employed. It is likely, however, given the very nature of heuristics, that as one goes "up" the meta-heuristic chain that there will be progressively fewer and fewer meta-heuristics per order. In other words, one would expect there to be far fewer meta-heuristics (2nd order) than heuristics (1st order) and far fewer still meta-meta-heuristics (3rd order) than meta-heuristics, and so on. If so, then there would be no reason, in principle, why such a regress must be viciously circular or un-anchorable as Fodor suggests.[27]

While the heuristics approach has available to it an internal means by which to contend with the heuristic-selection problem, there are other alternatives by which heuristics might be selected - without the need to posit a central executive. I will return to this following the presentation of Barrett's enzymatic model.

[27] Carruthers (2006) p. 358 offers a similar rejoinder.

With Carruthers' distinction marked and the conception of a wide-scope encapsulated device outlined, I will next consider an alternative model of how a modular system may contend with the input-routing or module selection problem overlooked by Fodor. I will begin by setting out Baars (1988) and Shanahan & Baars' (2005) global workspace model which provides a general account of how a massively modular system might be relied upon to contend with the input-routing problem. Next, I consider how the global workspace architecture (and similar "bulletin board" models of mind) might be realized by examining Barrett's model in which modules are metaphorically likened to enzymatic systems.

BAARS & SHANAHAN'S GLOBAL WORKSPACE ARCHITECTURE

Shanahan & Baars (2005) argue that Fodor's version of the frame problem relies upon a commitment to a particular architectural model that "betrays the assumption that it is the responsibility of the cognitive process itself to make the selection of relevant information."(Shanahan & Baars, 2005 p.12) Specifically, they note that while Fodor offers little detail as to the particular computational model he has in mind when making the claim that unencapsulated processes cannot be modeled, there are, they suggest,

> Strong hints of a commitment to a centralized, serial process that somehow has all of the requisite information at its disposal, and that has the responsibility of choosing what information to access and when to access it. Although parallel peripheral processes are part of the picture, they are passive sources of information that wait to be called upon before delivering their goods. (Shanahan & Baars, 2005 p.16)

In particular, they suggest that Fodor is led astray by Dennett's discussion of the frame problem (discussed in chapter 1) in which robots engage in a serial process of exhaustive consideration – generating and assessing each alternative (ramification) one-at-a-time. They note, however, that "the design of Dennett's robot is absurd," for each robot is required to carry out a "serial computation that exhaustively works through a long list of alternatives one-by-one before it terminates."(Shanahan & Baars, 2005 p.11) And so, under this serial and exhaustive model, Fodor takes the question to be one of how possibly a computational system could expeditiously bring relevant material to bear. Shanahan & Baars, however, suggest that Fodor's version of the puzzle should be recast as one of how relevant information might actively be made salient and available by modular processes themselves. In response to this recast puzzle, they provide a different architectural model, the global workspace model (GWM), in which no central executive is posited and in which the "responsibility for selecting relevant information [is distributed] among multiple parallel processes."(Shanahan & Baars, 2005 p.12) They explain,

> The essence of the global workspace architecture is a model of combined serial and parallel information flow. Multiple parallel specialist processes compete and co-operate for access to a global workspace. A specialist process can be responsible for some aspect of perception, long-term planning, problem solving, language understanding, language production, action selection, or indeed any posited cognitive process. If granted access to the global workspace, the information a process has to offer is "broadcast" back to the entire set of specialists. The means by which access is granted to the global workspace can be likened to an attention mechanism.
>
> The contents of the global workspace unfolds in a serial manner. But it is the product of massively parallel processing. A sequence of moment-to-moment snapshots of the global workspace would reveal a meaningful progression, and each state would typically be related to its predecessor in a coherent way. Yet the state-to-state relation itself is highly complex, and could not be obtained by a single computational step on a conventional serial machine. Rather, it is the outcome of a selection procedure that has at its disposal the results of numerous separate computations, each of which might have something of value to contribute to the ongoing procession of thoughts. … Unconscious information processing is carried out by the parallel specialist processes. Only information that is broadcast via the global workspace is consciously processed. (Shanahan & Baars, 2005 pp.12-13)

And so, Shanahan & Baars are offering the global workspace model as a direct counter to Fodor's claim that a central "executive" must be posited in order to contend with the various selection problems outlined in the previous chapter (*i.e.,* input-routing, module-activation, heuristic-selection problems).

Instead of positing an "executive" that serially and exhaustively queries each available module and makes a determination about which of these is to be activated (*i.e.,* routed representations) and when, the global workspace model suggests that this responsibility is undertaken actively by the specialist inferential processes (*i.e.,* modules) themselves working both in parallel and competitively. (Shanahan & Baars, 2005 p.16)

By way of example, Shanahan & Baars consider the processes underlying perception and object recognition – specifically those involved in recognizing a Rorschach inkblot image as an elephant. Fodor's account, they suggest, is committed to a model of the central processor that "poses a series of questions one-at-a-time – is it a face? Is it a butterfly? Is it a vulva? And so on – until it finally arrives at the idea of an elephant." (Shanahan & Baars, 2005 p.16) The global workspace model, in contrast, posits a set of specialist processes working cooperatively and competitively in parallel – one of which is "always on the lookout for elephantine shapes." (Shanahan & Baars, 2005 p.16) Once this particular device is activated (*i.e.,* when an elephant(ish) image is perceived) "the information that it has to offer makes its way to the global workspace, and is thereby broadcast back to all the other specialist processes." (Shanahan & Baars, 2005 p.17)

While global workspace theory provides a plausible account of perceptual processing, there is no reason to suppose that a host of other cognitive processes could not be undertaken by means of the same processes, Shanahan & Baars suggest. And so, once the elephant shape detector is activated and its output is relayed to the global workspace, this information is broadcast to all of the system's other specialists to process. Following further parallel and competitive processing by modular specialist processes, further elephant-related information would be brought to bear in the global workspace.

With respect to Fodor's challenge – specifically his claim that only an un-modelable central executive could possibly contend with the input-routing (module/heuristic activation) problem - Shanahan and Baars conclude that the global workspace model provides an account of how the operations of an unencapsulated system could be massively modularly realized. Specifically, they suggest that global workspace theory provides a computationally feasible model that "explains how an informationally unencapsulated process can draw on just the information that is relevant to the ongoing situation without being swamped by irrelevant rubbish" (Shanahan & Baars, 2005 p. 25) by relegating the task of relevance determination to the specialist processes themselves. Such parallelism, cooperation and competition for access to the global workspace, they continue, "confers great computational advantage without compromising the serial flow of conscious thought, which corresponds to the sequential contents of the limited capacity global workspace." (Shanahan & Baars, 2005 p. 25)

With respect to the relevance of global workspace theory to the (philosopher's) frame problem in particular, Shanahan notes,

> The particular blend of serial and parallel computation favoured by global workspace theory suggests a way to address the frame problem in the philosopher's sense of that term … In particular, in the context of so-called informationally encapsulated cognitive processes, it allows relevant information to be sifted from the irrelevant without incurring an impossible computational burden. More generally, broadcast interleaved with competition facilitates the integration of the activities of a large number of specialist processes working separately. (Shanahan, 2006 p. 438)

Furthermore, Shanahan & Baars note,

> If the frame problem is a genuine puzzle, the human brain incorporates a solution to it. In global workspace theory, we find clues to how this solution might work. Global workspace theory posits a functional role for consciousness, which is to facilitate information exchange among multiple, special-purpose, unconscious brain processes. These compete for access to a global workspace, which allows selected information to be broadcast back to the whole system. Such an architecture accommodates high-speed, domain-specific processes (or "modules") while facilitating just the sort of crossing domain boundaries required to address the philosopher's frame problem. (Shanahan & Baars, 2005 p. 2)

While global workspace theory is quite promising it is also, as presented, somewhat suggestive. What remains is for an account to be provided of how such a computational system might be instantiated. Particularly, some account of how modular devices themselves could contend with the input-routing/module-activation problem is needed. To this end, I will turn to Barrett's enzymatic account of modularity to help flesh-out Baars' proposal.

BARRETT'S (METAPHORICAL) ENZYMATIC ACCOUNT

Like Carruthers and Shanahan & Baars, Barrett (2005) argues that Fodor's pessimism relies upon a commitment to both a very specific account of modularity and the particular systems-level model that derives from this. In response, Barrett proposes a way of thinking about the operations of modular processes drawing upon the operations of enzymatic systems that provides a *metaphor* for how a "bulletin board" architecture (*e.g.*, global workspace model) might be realized.

Enzymes catalyze reactions, systematically combining and transforming substrates to generate new molecular products in a manner that "can be regarded as a kind of computation."(Barrett, 2005 p.268) Glossing over the biochemical details, enzymes function quite generally in the following manner. Cells contain a number of different enzymes each of which has its own characteristic "shape." The particular shape of an enzyme determines both the kinds of substrates with which it may interact and the kinds of operations or transformations that it serves to catalyze. Put more directly, enzymes are specified both by the kinds of substrates that can adhere to them and by the types of reactions that they undertake (catalyze) on these substrates. Cells also contain many proteins, each of which with its own particular characteristic "shape." Enzymes adhere to any substrate molecule (protein) of the proper geometry (*i.e.,* if a substrate's "key" fits an enzyme's geometric "lock," a reaction is catalyzed.) In keeping with the analogy, enzymes may have multiple "locks," pockets, active sites, or input nodes, each of which accepts a specifically shaped "key." And so, enzymes may be adhered to by a number of substrates simultaneously. As such, the particular "shape" of the output molecule of any enzymatic reaction will depend on both the substrates operated upon and the particular kind of transform that is undertaken by the enzyme. Once transformed, these newly minted molecules are outputted and released back into the cellular central pool where they can be "picked up" by any enzyme capable of adhering to it (*i.e.,* by any enzyme with the right shaped "lock") thus potentially catalyzing a subsequent reaction with another substrate or substrates. In this manner, cells can come to contain many proteins of varying geometries, some of these will be, somewhat conceptually, raw molecular "primitives" (*i.e.,* untransformed or minimally processed substrates), while others, as the result of cascades of enzymatic transforms, are highly processed and complex.

Enzymes are of interest, Barrett suggests, for they exhibit the three principal properties of modular computation systems. The specificity with which enzymes and substrates interact suggests that enzymes accept as input only particular kinds of substrates (*i.e.,* those that are specifically "shaped"). Second, enzymes perform specific operations, systematically combining or otherwise transforming substrates and generating new molecules in a rule-governed manner. Finally, the output of these enzyme-catalyzed transformations (*i.e.,* a newly minted molecule) are of a form that is acceptable as input to other processes (*i.e.,* other enzymes can accept and operate upon these outputs). These properties are significant, Barrett explains, for "enzymes are computational devices that solve problems that cognitive computational systems also face: they achieve functional specificity in an "open" system."(Barrett, 2005 p.269)

With respect to this, there are a number of features of enzymatic computational systems that are applicable to the challenges set out by Fodor. First, enzymatic systems provide a means by which to explain how a massively modularly realized cognitive system might contend with the input-routing (module activation) problem in a way that does not require the positing of a central "executive" mechanism. Echoing Carruthers' discussion of wide-scope encapsulation, Barrett marks a distinction between the "access specificity" and "processing specificity" of devices. Access specificity is "the breadth of information that a device has access," while processing specificity is the "breadth of information that a device actually processes."(Barrett, 2005 p.274) Enzymatic devices, Barrett argues, are "access-general," for all of the substrates contained in a cell's central pool are non-partitioned and thus available to any enzyme, and "processing-specific," for each enzyme undertakes a particular rule-governed transformation on only those substrates that are properly templated. And so, he continues,

> One can have an enzymatic system in which all of the enzymes in the system have access to all of the substrates, and in which only the 'correct' reactions are catalyzed. This is because of the lock-and-key nature of molecular recognition processes, which depend on the diffusion of information for their proper functioning. (Barrett, 2005 p.270)

And so, in enzymatic systems there is no need for a central "executive" that decides how information is (representations are) to be routed. Rather, input specificity is achieved by template matching – (*i.e.,* a module becomes active only when a representation maps to its particular recognitional front-end). Substrates (in biological systems) and, by analogy, representations (in cognitive systems) that are present in the system's central informational pool (*e.g.*, the cellular pool in biological systems and the "bulletin board" or global workspace in cognitive systems) are diffused or "globally broadcast" to each enzymatic specialist device in the system. Those molecules/representations that are properly templated (*i.e.,* a match to a module's particular recognitional front-end) are "picked up," operated upon and the resultant transform returned to the central pool.[28] Since these specialist devices (like enzymes) operate in parallel, and only those that are capable of accepting as input the representation broadcast will become active, no "executive" decision needs to be made about which modules are to be activated and when they are to be activated.

The enzymatic model can also account for how both complex and novel representations might be modularly generated. Since enzymes are capable of accepting a number of substrates they are thus "multi-dimensional" with respect to the kinds of substrate-representations that they may operate upon and the kinds of operations that they can perform. (Barrett, 2005 p.270) And so, while some enzymes will have (conceptually) rather simple recognitional front ends, others may be structured such that they can interact only with highly processed substrates (*i.e.,* substrates/representations

[28] Conceptually, one may envision this as a process whereby either representations in the central pool are actively globally broadcast (*i.e.,* relayed) to all specialists processes simultaneously or one in which representations stay put and each specialist is constantly and actively scanning the contents of the central pool "looking" for any representations that are the proper shape.

that are the result of cascade of prior transforms) thus allowing for a high degree of input specificity. At the same time, as enzymes accept as input any substrate with the properly shaped key or keys, and molecules may have many keys, there is the potential for complex molecules to form a partial or incomplete "fit" with an enzyme (*i.e.,* some but not all of the enzyme's active sites are filled by the substrate). And so, Barrett explains, while each recognitional event is Boolean (*i.e.,* logic-gated), "the binding procedures of enzymes have analog properties [because] many individual chemical bonding events contribute to the recognition process. The sum of these determine recognition; there can be better or worse degrees of fit."(Barrett, 2005 p.270)

Along similar lines, enzymes can also engage in "byproduct processing" whereby they accept as input substrates that mimic portions of the protein that the enzyme was "designed" to recognize. Many synthesized drugs, Barrett explains, function by mimicking the shape of proteins that are in the "proper domain" of an enzyme, thus forging an "artificial" fit by exploiting the active sites of enzymes. It is this capacity, Barrett continues, "of low level byproduct processing by enzyme-like cognitive devices [that] may be what permits novel combinations and processing of representations beyond, in some sense, what the mechanism was designed to do."(Barrett, 2005 p.270)

Furthermore, enzymes need not only be such that they adhere to just one particular substrate. Rather, they can be capable of recognizing and thus operating upon substrates of particular classes or kinds. That is, since substrates may possess multiple sub-unit "keys" and since some of these "keys" might be shared by multiple substrates, those molecules with the same sub-unit "key" in common form a class or kind, although their geometries might otherwise vary. In this manner, properly templated enzymes might come to pick up and transform substrates that, while otherwise geometrically/representationally quite different, share a sub-unit in common. With respect to this, Barrett explains that in some instances those sub-units not recognized might be unmodified, preserved or "carried through" in the output molecule. In other reactions these sub-units could be transformed by the operations of the enzyme and thus not be preserved in the output molecule/representation. This property of "carry through" (*i.e.,* the non-modification of portions of the substrate) is computationally significant for two reasons. First, it provides a means by which to explain how particular properties (and likewise truths) can be preserved through a cascade of enzymatic transformations. Second, Barrett explains, the capacity of enzymatic systems to undertake operations on only a portion of a representation is significant for parallel distributed systems notoriously have great difficulty in undertaking truth-preserving operations on minimally altered representations.

In the enzyme model, class-level processing is achieved by means of the transformational application of a "tag" to a molecule. Once appended, such "tags" may then be picked up by another enzyme, thereby influencing both the kinds of other substrates/representations the enzyme may adhere to (accept as input) and how the recipient enzyme processes or transforms these substrates/representations. In this manner then the application of a "tag" to a substrate/representation can influence a cascade of enzymatic transformations. At the same time, carried-through tags also provide a means by which to mark a substrate's/representation's membership in a class – irrespective of how many other transformations are undertaken on it. So doing provides an account of how information about some particular property of an object (and thus some predicate or "truth" in the cognitive domain) might be both preserved from one operation to the next and also be brought to bear on subsequent enzymatic computations/transformations.

Not only does enzymatic "tagging" allow for class-level processing and informational carry-through, but it also, Barrett explains, "allows for horizontal and 'top-down' control in which the output of devices can influence other devices at the same horizontal level, a kind of feedback which is not typical of Fodorean modular systems."(Barrett, 2005 p.271) In addition to the tagging of substrate molecules, enzymes can output unattached tags. These free-floating tag-substrates, when recognized by the front-ends of particular enzymes can have the effect of turning these "off" or "on" (by altering the geometry of an enzyme's front-end and thus affecting the kinds of substrates to which they may bind) and/or of modulating the transformations that these will undertake on recognized molecules in the manner outlined previously.[29]

With respect to Fodor's objection to the massive modularity of mind approach to modeling doxastic processes, the enzyme model, Barrett explains, "is important because it points to potential computational solutions to problems that standard versions of modular architectures are said to face."(Barrett, 2005 p.272)

Given that enzymes have access to any and all substrates present in the cellular pool and since they undertake specific transformations on only those substrates that are "recognized" it is clear that they are, in Barrett's terms, access-general and processing-specific devices. With respect to the challenge posed by Fodor, the enzyme account

[29] Given the model outlined, such a process does not turn a module "on" or "off" in the manner suggested by Fodor. Rather substrate modulation functions to alter the recognitional front-ends of particular enzymes making it such that they can either accept or not accept the substrates currently floating about in the cellular central pool. Over time, as the contents of the central pool are transformed, a module that was once inactive or "off" might very well become "on" and vice versa.

provides a model of a modularly realizable system that achieves processing specificity even under a condition of access generality. (Barrett, 2005 p.275) Or, put another way, the enzymatic account provides a model of a modular system that is both process-encapsulated and isotropic.

Furthermore, as enzymatic devices are subject to feed-forward, horizontal and top-down[30] influence by means of the tagging operation outlined above, the computations undertaken by an enzyme can be modified in light of and are thus sensitive to any number of factors. That is, unlike Fodor-style modules, enzymatic devices are capable of performing operations that are both content and context sensitive. This is of course significant for each enzymatic device itself undertakes only a (syntactically) local computation. And so, the enzyme model provides an account of how a local computational device (a module) can process and generate content and context sensitive outputs - something that on Fodor's model isn't possible.

Enzymatic tagging allows for information, once obtained, to be retained between operations/transformations. Rather, when "carried through" in the manner outlined, such tags allow for particular information to be preserved and, given the input specificity of enzymatic devices, to be processed by only the "right" kinds of devices (*i.e.,* those with the right kinds of recognitional front-ends). In this way, particular tags once added may be carried through and influence all subsequent operations undertaken on the representation. The tagging of substrates/representations, Barrett explains,

[30] There should be no confusion here about the nature of such top-down influence. The idea here is that if particular representations are maintained in the global workspace then they will be globally broadcast to all specialist systems. In this way then a representation might be relayed to enzymatic devices that were part of the transformational cascade that resulted in itself. By modulating the front-ends of these devices, and relaying the representation back to them, the same device may transform further (though differently) the representation.

> Can provide a kind of functional input restriction, even in the absence of hard-wired piping. Like a child safety lock, a single tag can render a representation 'out of bounds' for a host of computational processes without establishing a wall or physical partition between the representation and those processes. …Tag-mediated input restriction suggests that the input pool of a device could in principle be changed on the fly, granting or denying access to a database simply by changing the set of representations to which a particular tag is attached, without altering the design of the device in any way. (Barrett, 2005 p.278)

Turning now to cognitive processes more explicitly, Barrett suggests that higher-level semantic categories can be enzymatically computed by means of the operations of a set of processes, each of which appends a semantic tag to a representation. Returning to the bulletin board or global workspace model, each tag-appending device will output a transformed substrate to the shared pool, board or workspace. Via carry-through, the representation will come to be multiply semantically tagged. In this manner, quite complex and higher-order semantic categories could be computable. The following example may help to clarify,

> Consider, for example, how the semantic property PREDATOR might be computed from a perceptual input, using an imaginary example. We could envision a three-step process. First, information from a lion passes through the object parsing system, and is deposited in the central pool. Next, this representation is matched with a perceptual template that adds a LION tag and returns it to the public representation pool. Finally, a third mechanism takes as input the representation with the LION tag – perhaps using something akin to a lookup table of animal tags that satisfy its input criteria - and adds a PREDATOR tag. … Once the PREDATOR tag is added to the representation, this tag can then be used to admit this information to various predator-specific computational procedures. A 'higher-order' or 'abstract' semantic category has been computed from raw sensory inputs. In principle there is no reason a particular representation could not carry many, many tags. (Barrett, 2005 p.279)

This example is relevant in light of Fodor's argument that such things as "cheater detection" cannot be modularly realized. This follows, he suggests, because determining what constitutes a cheating situation or a social exchange cannot be based upon the perception of "Very Subtle Clues" alone. (Fodor, 2001 p.76) Rather, Fodor concludes, determining if a situation is a social exchange requires "thinking" –

something that modules just cannot do, for the reasons discussed in the previous chapter.(Fodor, 2001 pp.76-77) Barrett's example, however, suggests that enzymatic processes need not accept only perceptual cues. Rather, via the tagging operation, both perceptual, contextual and content specific information may be brought to bear in the local computational processes. With respect to cheater detection, then, via transformational cascade, many tags may be appended to some (initial) representation. Some of these tags may be the result of an enzymatic transform that accepts a "raw" perceptual input (*e.g.,* the addition of the LION tag in the above example). Others would be the result of an enzymatic transform that accepts only particular highly processed and multiply tagged representations (*e.g.,* the addition of the PREDATOR tag in the above example). If a putative cheater detection module has a particular template front-end, then only those representations that are properly multiply-tagged (semantic and otherwise) would be "picked up," operated upon and its output returned to the central pool. In this manner, the highly complex determination that "X is a cheat" or "this is a social exchange" becomes amenable to modular description.

And so, the enzymatic model provides a means by which to explain how modularly realized processes can contend with the input-routing (module-activation) puzzle without having to posit a central executive. Furthermore, by means of a process of template matching, the enzymatic model attains functional specificity without violating the isotropy condition – as traditional Fodor-style modules must. So doing alleviates the need for a central executive to be posited while also explaining how the "right" representation/information ends up being processed by the "right" operation at the "right" time.

The enzymatic model also provides a response to Fodor's argument from (Quinean) holism. Specifically, as an "access general" system, the enzymatic model allows for any piece of information to "propagate through all relevant inference devices. [By means of this,] all inferences that the system is capable of generating, given its current set of rules and its complete knowledge database, will be generated."(Barrett, 2005 p.284) With respect to the argument from (Quinean) holism, enzymatic modular systems possess many features that are applicable to contending with the problems raised by abduction and belief-revision. Unlike Fodor's model, the enzymatic metaphor account allows for representations to be cognitively "percolated." Specifically, the model allows for the same representation to be operated upon by many processes (in parallel), thus allowing for the representation to be simultaneously compared to many beliefs (other representations) held by the system. (Barrett, 2005 p.283) Furthermore, unlike Fodor's feed-forward model, the enzymatic metaphor account allows for horizontal control and feedback (both positive and negative) in the manner outlined. So doing, Barrett explains, provides a means by which to explain how hypotheses consistent with the available evidence might be reinforced. (Barrett, 2005 p.283) Since the tagging of a representation by one device can affect how it is processed by other devices, the reprocessing or rehearsal of a representation accounts for how, "the same piece of information can lead to different inferences depending on how it has been previously tagged (which in turn can depend on the context in which it is presented.)" (Barrett, 2005 pp. 283-4) Finally, with respect to the argument from (Quinean) holism, Barrett notes,

> In enzymatic computational systems every device in the system has access, in principle, to every representation in the system, and therefore, can in principle leverage the inferential power of every other device in the system. … One might

> think of this kind of global process – representations seeping through the system by diffusion and generating all possible inferences as they go – as the system going to catalytic equilibrium. (Barrett, 2005 p.284)

Abduction, he continues, is the reverse of this process. By starting with a diverse set of facts, the system can, by application of a carried-through suppositional or pretense tag, "back-generate the single fact which, when passed through many inference systems, would produce these diverse inferences."(Barrett, 2005 p.284) At catalytic equilibrium, he proposes, that representational substrate that is capable of accounting for the available evidence will be discerned. As suppositionally tagged, the system may do so without also tainting the system's belief-set along the way.

Returning to earlier discussion, we have from Barrett's metaphor a plausible means by which to contend with the heuristic selection problem. Specifically, there is no reason to think that enzymatic processes must be limited to the transformation of substrates. Rather, just as Barrett suggests that tags (*e.g.,* LION) are retrieved and appended to substrates, it is reasonable to think that the activation of an enzyme might result in the employment of some particular heuristic. This is even more plausible given that the particular front ends of enzymatic devices themselves are rather heuristic in design – for they only select (*i.e.,* adhere to) particular substrates while they have access to them all. And so, once activated, an enzymatic device might output or invoke a directed search of memory (preferentially exploiting some particular tag or tags to expedite search) thus preferentially retrieving particular (kinds) of material for other specialists to accept (or not) depending on the current "shape" of their front-ends. Likewise, the employment of particular heuristic stopping and decision rules could be controlled enzymatically in much the same manner. This is relevant for the global workspace model suggests that the competition (for access to the global workspace) between or among specialist devices is resolved heuristically. While the details are in need of fleshing out, the point is that the enzymatic model provides a viable control structure for contending with what I have termed the *heuristics selection problem* and one that does not require the positing of a domain-general central executive mechanism.[31]

Before proceeding, a potential objection to Barrett's account needs to be considered. Were Barrett offering his account as a complete computational model of the operations of mind, we would have reason for thinking the proposal problematic. That is, were we to think that Barrett is suggesting that cognition is, in fact, undertaken enzymatically, there is reason to think this implausible. And so, one might object, for example, noting that any fully enzymatic system must rely on passive diffusion to move substrates around and thus to move processes and cascades along. That is, the modular sub-systems/operations of an enzymatic system must passively "wait" for a substrate of the proper shape to "float by." The smaller the size of the "pool," the faster substrates can be diffused, picked up, transformed and released back. Therefore, given the physical facts about diffusion gradients, as the size of the pool increases, so too does processing time. And so, the concern continues, any large-scale enzymatic system would necessarily be very slow. Since the pool size of any enzymatic system that attempted to contend with the kinds of problems currently under consideration would have to be quite large, it is, the concern continues, an implausible suggestion.[32]

[31] While I argued previously that the heuristics approach has an internal means by which to contend in principle with the heuristic selection problem (*i.e.,* that there are likely to be fewer and fewer heuristics as one progresses "up" the meta-heuristic chain) it seems (at least prima facie) unlikely that heuristics are selected ultimately by one n^{th} order meta-heuristic.

[32] There is, as Cherniak rightly notes (personal communication, 2008), given the very real physical constraints on systems relying on diffusion, a reason why cells are so very small. *C.f.,* Adelman (1994), Benenson *et. al.,* (2001) and MacDonald, Stefanovic, & Stojanovic (2008) for discussion of the promise and limits of biomolecular/DNA computing.

There are, I think, two responses to this worry. First, if we suppose Barrett to be in fact offering his account as a model of mind, (i.e,. that minds are enzymatic systems) the structure of the global workspace framework should go some distance in responding to the concern. Specifically, global workspace theory, unlike a truly enzymatic system, need not rely on diffusion gradients to move representations about. Rather, Baars' model suggests that either (a) consumer subsystems are actively scanning the workspace for representations of the right "shape" or (b) each representational substrate is globally broadcast (*i.e.,* actively relayed) to every consumer sub-system directly – whereby only those of the right "shape" are processed. And so, under the global workspace framework (unlike that of truly enzymatic systems) the very real processing-time concern raised about any computational system that relies upon diffusion to move representations/substrates, need not apply.

However, given Barrett's discussion, it does not appear that he is offering this account as a model of mind. Rather, he is offering a metaphor, an heuristic for thinking about modular processes on analogy with enzymatic ones. And so, while modules are not enzymes and thinking is not enzyme catalyzed substrate transformation, we can, by thinking in these terms, gain some insight into the puzzle of how the input routing problem might be contended with locally. This is turn helps us to better understand how the modular specialist systems posited by global workspace theory could be, without the need for a generalist input routing device, selective with respect to both the kinds of representations that they will operate upon and the kinds of transformations that they will effect.

The aim, then, of this section is not to provide a complete model of the workings of mind but rather to show that there are viable and computationally feasible alternatives that meet the practical computational engineering challenges posed by Fodor.

Very briefly, however, Carruthers (2006) provides a detailed account of how doxastic processes (including abduction and the creativity and flexibility of thought) can be accommodated by a variant of this model.[33] Expanding on the general framework of the global workspace model, Carruthers suggests a role for natural language in integrating the outputs of the various specialist inferential mechanisms.

> The quasi-perceptual imagistic representation is globally broadcast and made available inter alia to the language comprehension sub-system, which attaches a content to it and makes that content (as expressed in some sort of Mentalese representation, perhaps in the form of a mental model) available to a suite of central/conceptual modules. … When 'P' [a sentence in English, say] is mentally rehearsed the comprehension sub-system extracts form it the mentalese representation |Q|, which can then (especially if it takes the form of a mental model) be globally broadcast for the different central modules to get to work upon, providing that any aspect of it meets their input conditions. (Carruthers, 2006 pp. 264-5)

[33] As an additional example, borrowing from Franklin & Graesser's (1999) implementation of the global workspace model, Shanahan & Baars consider the activity of analogical reasoning since this is precisely the kind of unencapsulated process that Fodor claims is un-model-able. (Fodor, 1983 p. 105) Following a review of the current computational models of analogical reasoning (*e.g.*, ARCS, MAC/FAC, SME, IAM, and LISA), Shanahan & Baars concludes that Hummel & Holyoaks's (1997) LISA model (learning and inference with schemas and analogies) is "the most accurate and psychological plausible of the current computational models of analogical reasoning." (Shanahan & Baars, 2005 p. 20) The LISA model is of interest for it relies upon a serial mapping operation while undertaking retrieval (of representations in memory) by means of parallel processes. This combination of serial and parallel processes is in direct alignment with the architecture proposed by the global workspace model. Specifically, Shanahan & Baars explain,

> The currently active propositions in LISA's working memory correspond to the current contents of the global workspace model. That parallel activation of propositions in LISA's long-term memory corresponds to the broadcast of the contents of the global workspace. And the distributed propositions in LISA's long-term memory correspond to the parallel, unconscious processes in global workspace theory. But most importantly, the necessary serial presentation of propositions in LISA's limited capacity working memory matches the necessarily serial presentation of coherent material in the limited capacity of the global workspace. (Shanahan & Baars, 2005 p. 23)

Put another way, the most plausible model of analogical reasoning on offer is "also the one that maps most cleanly and elegantly onto a global workspace architecture."(Shanahan & Baars, 2005 p. 20)

And so, building upon Baars' general framework, Carruthers suggests that quasi-perceptual image-models are constructed of rehearsed sentences in mentalese which are then globally broadcast to all specialist devices for further processing by those consumer specialists the "front-ends" of which are properly templated. The language production sub-system, Carruthers continues, engages, in effect, in a process of linguistic (re)description of the various (transformed) images returned by these specialist devices. Carruthers' proposal, then, is that the inner rehearsal of sentences, results via operations of image re-construction (undertaken by means of back-projections to and processing by various perceptual sub-systems), in the generation of image-models of the contents of these sentences which are then globally broadcast to the set of specialists for further processing. Through cycles of rehearsal in inner speech and global broadcast of the integrated image-models generated to the set of manifold specialists, the language faculty functions to both integrate and consolidate the information relayed from potentially quite disparate specialists into a coherent statement in mentalese. In performing this function, the language faculty facilitates creativity and flexibility of thought by enabling information from disparate content domains to be brought to bear and integrated into an image-model.

CONCLUSION

I have argued in this chapter that the normative rationality principle underlying and motivating Fodor's pessimistic conclusion is, as unsatisfiable by any finite and physically realized system, impossibly demanding. We have then, good reason for rejecting Fodor's *normative* rationality condition. I next argued that it is false that we *ever* arrive at conclusions rationally (in Fodor's sense). We also, then, have reason for rejecting Fodor's *descriptive* claim. In light of some *ought implies can* reasoning, this normative principle must be weakened to one that that can be satisfied by the likes of us (or, at the very least, to one that can be satisfied by a finite and physically realized system). (Cherniak, 1984 pp. 20, 106, 110, 113) While I have made no attempt at setting out a suitably weakened rationality condition, as this was not my aim, any such revised principle must, so it would appear, take the very general form:

- Arriving at a conclusion in a rationally warranted manner requires that *some* of the available and relevant information be considered.

Weakening the rationality condition, even in this very general way, however, has significant implications for Fodor's pessimistic conclusion. Specifically, since Fodor's rejection of both the massive modularity of mind and heuristics approaches reduce, at base, to the claim that these are "irrational" strategies, and we have reason to reject Fodor's rationality condition, then it follows that we have then no (*a priori)* reason to dismiss these strategies *tout court* in the manner suggested by Fodor. Specifically, since heuristics and modular systems can, in principle, satisfy a weakened version of the rationality condition, we have no reason for rejecting these strategies on principle. Since we have no reason to think that we ever arrive at conclusions rationally (in Fodor's sense), then we have no reason for thinking *our* cognitive processes un-model-able as a matter of principle.

Establishing this point of principle, however, responds only to Fodor's *a priori* arguments, for it still might be the case that, as a practical matter (as opposed to one of principle), our doxastic processes are not computationally model-able. Specifically, what remains is for a plausible and computationally feasible model to be provided that contends with the particular engineering challenges posed by Fodor (*e.g.,* isotropy, holism, the input-routing, module-activation, and heuristic-selection problems). In response to these rather practical (as opposed to in principle) engineering challenges, drawing upon Gigerenzer, Carruthers, Baars & Shanahan and Barrett's discussions, I outlined an alternative model. The aim of this section has not been to provide a complete model of the workings of mind, for this is far too involved a task, but rather to show that there is a plausible and computationally feasible alternative overlooked by Fodor that provides a viable framework from which to proceed.

CHAPTER 4: ON (THE VERY IDEA OF) BRINGING EMOTIONAL AFFECT TO BEAR

In the remaining two chapters, I will argue for the following claim: Given what (at least some of what) the "emotions" are and the role occupied by affect in attentional direction, motivation, meta-planning, and decision-making, bringing emotion to bear might provide a promising approach for explaining how we might contend with some of the frame problems that arise in practical reasoning and decision-making. In what follows, I will not be providing a complete account of what the emotions are, nor will I be providing a comprehensive solution to the frame problem. I will also not be undertaking an exhaustive discussion of all of the roles that emotion might occupy in practical reason, nor will I consider all of the roles that it might play in helping us contend with various instances of the frame problem. Rather, I will be arguing that the consideration of emotion might provide a promising way of approaching some of these problems. Suggesting this requires that two principal challenges be addressed.

First, it must be shown that emotion or affect occupies the right kind of relationship with/to the operations of reason generally and practical reason and decision-making in particular. Clearly, in order even to suggest that bringing emotional affect to bear on instances of the frame problem holds any promise whatsoever, some account both of what the emotions are and of their relationship to/with reason must be set out. As a first step, I will begin by presenting a review of some representative theories of what emotions are and a discussion of the relationship between reason and emotion proposed by these accounts. Following this discussion, I will argue that we have no reason – in advance of our considering specific proposals - to think that bringing emotional affect to bear on these puzzles should be irrelevant nor have we any reason for thinking such a proposal to be an in principle "non-starter." Having argued that the suggestion is not worthy of rejection *tout court*, I will suggest that, given what the emotions are and their relationship to/with cognition generally, we have reason for thinking that the emotions, at least on principle, may provide a promising approach for contending with some of the frame problems that arise in practical reason and decision-making.

Second, given the dual aspects (*i.e.,* the two horns of the dilemma of speed and accuracy) of the frame problem, if consideration of emotion is to help us in understanding how we might contend with some of the instances of the problem that arise in practical reason and decision-making, then it must be shown both to expedite (i.e, help in "tracting") and to increase the accuracy of these operations. Clearly, were emotion to either (a) invariably slow, hinder or otherwise make more computationally burdensome or (b) decrease the accuracy of the operations of practical reason and decision-making, then we would have no reason for thinking that bringing emotion to bear on this puzzle should hold any promise. Similarly, if the influences of emotion on practical reason are, with respect to either the tractability/complexity or the accuracy conditions, arbitrary (*i.e.,* if emotion is as likely to expedite and increase the accuracy of these processes as it is to hinder them), then emotion would be entirely "orthogonal" to the puzzle. If emotion is orthogonal with respect to issues of tractability and accuracy of the operations of practical reason and decision-making, then bringing them to bear should hold no promise either.

Taken together, if emotion assists in both expediting and increasing the correctness of practical reason and decision-making and if it occupies the right kind of relationship to/with those of cognition, then consideration of emotion might provide a promising approach for helping to understand how we might contend with some of the frame problems that arise in practical reason and decision-making.

The first challenge will be the focus of the remainder of this chapter. The remaining challenges will be considered in the next. Before even beginning to set out a case in support of the claim that consideration of emotion might provide a promising approach for understanding how we might contend with a number of the frame problems that arise in practical reason and decision-making, I need first to consider the following *prima facie* irrelevance objection: Given what the emotions appear to be (and what is involved in having them), the emotions are entirely irrelevant to the frame problem. As irrelevant, bringing emotion to bear on the frame problem – in any way – holds no promise for helping us to understand (and much less to model) the operations of practical reason and decision-making.

The irrelevance objection, I suggest, stems from a particular and pervasive philosophical view about both what the emotions are and of what their relationship to/with reason must be. Specifically, this challenge reduces to the claim that if emotions are always dependent upon belief, then they too will be infected with or by all of the frame problems that arise in belief-fixation. If so, the concern continues, bringing emotion to bear on those frame problems that arise in practical reason would serve only to import those belief-fixation frame problems into this other domain. If emotions are always belief-dependent, then bringing them to bear on those frame problems that arise in practical reason and decision-making would be an obvious non-starter. That is, if emotions are always belief-dependent and if the fixation of belief is itself infected with frame problems, then bringing emotion to bear on the question of how we might contend with some of the frame problems that arise in practical reason would succeed only in bringing a frame problem tainted operation to bear on the question of how we might contend with some of the frame problems that arise in another domain. So doing, the concern continues, might very well make the situation worse but it certainly would not make it any better. At best, emotion would be irrelevant.

In order to meet this challenge – establishing the very modest point that the emotions are not irrelevant to the frame problem – I will begin by setting out some representative and rather traditional philosophical *propositional attitude* accounts of the emotions and their *cognitive appraisal theory* counterparts in psychology. Next, I will discuss how these approaches result in the claim that bringing emotion to bear on the frame problem (in any form) would be entirely nonresponsive - a "non-starter" proposal. Following this, I will present and consider the evidence most often relied upon in support of the cognitive appraisal approach to emotion. This evidence, I suggest, is both methodologically problematic and fails to adequately establish the claim that emotions are "cognitive." Since the irrelevance objection relies upon this particular conception of what the emotions are, I suggest that - in advance of our examining particular proposals – there is little reason for thinking emotion to be irrelevant to the question of how it is that we contend with some instances of the frame problem. Next, I turn attention to an alternative account of emotion provided by *automated appraisal theory*. Following this, I will argue that the automated appraisal/basic emotion approach, while likely deficient as a *complete* account of what the emotions are, does provide a plausible partial response to this question. Putting this somewhat differently, while we know little about what the emotions *really* are, we do know that *at least some* of what they are can be explained in terms of the activities of automated, autonomous and modularly realized appraisal mechanisms and innate emotion programs.

At the outset, let me state once again that my aim in this chapter is a modest one. I will not be providing a complete answer to the vexed and confused question of "what the emotions are." Since the evidence bearing on this question is so grossly insufficient, it is unlikely that a complete account could, at this time, be provided. For the present aim (*i.e.,* of suggesting that bringing emotion to bear might help us to understand how we contend with some instances of the frame problem) I need not, however, provide a complete answer this question. Rather, I need only provide reason for thinking that *at least some* of what the emotions are are activities undertaken by automated, autonomous and plausibly modularly realized processes that influence practical reason and decision-making in ways relevant to contending with the dual horns of the frame problem.

By way of alleviating a potential source of confusion, I will note at the outset that the debate between "cognitivist" (*e.g.,* Lazarus) and "automated appraisal" (*e.g.,* Zajonc) accounts of emotion, to be discussed subsequently, focuses on the question of just how "cognitive" emotions are. While terminologically unhelpful, that a process is "cognitive" in the cognitivists sense (*i.e.,* Lazarus' sense), requires that "higher-level thinking" - specifically the systematic manipulation of propositionally formulated sentences resulting in the fixation of belief - be undertaken *before* any emotion is experienced or emotional response induced. Automated appraisal theorists (*e.g.,* Zajonc) deny this claim and hold instead that "cognition" (understood in this sense) is not a precondition of emotion or of the induction of an emotional response. To make this explicit, what is at stake in this debate is *not* the issue of whether the induction of a particular emotion requires prior information processing, but rather whether it requires prior processing by those systems involved in the fixation of belief.

Put somewhat differently, both camps agree that some form of information processing is requisite for the induction of an emotion. What is at issue is the extent to which the emotions are *belief-dependent*. Lazarus and the cognitivists maintain that the kind of information processing undertaken prior to the induction of any emotion is "cognitive" insofar as all emotions are belief-dependent, while Zajonc and the automated appraisal theorists hold that these prior processes are "non-cognitive," *i.e.*, are undertaken by processes that are automated, autonomous and rather perceptual in nature. Under this latter account, emotions (at least some of them) are not belief-dependent. Any theory, then, holding that the emotions are belief-dependent (*i.e.*, not merely cognition-dependent) is a cognitive appraisal/cognitivist account. Those holding emotion (at least some of them) to be "cognition"- dependent but *not* belief-dependent are non-cognitivist accounts. Put another way, any account denying that (at least some) emotions are belief-dependent (regardless of whether or not such accounts claim emotions to be the result of "some sort of prior cognitive" operation) is not a cognitive appraisal account. If this distinction is not properly marked and maintained, that is if what is to count as "cognition" is allowed to be broadened to include *any* prior informational processing, then there would be absolutely nothing about which Lazarus/Zajonc and the cognitivist/non-cognistivists accounts disagree – since both agree that some sort of informational processing is required. It is only by marking the distinction that the debate even exists.

It is perhaps worth noting that only very strong interpretations of the cognitivist and automated appraisal accounts would necessitate the rejection of the other. That is, unless one holds that *all* emotions are belief-dependent or that they are *all* "non-cognitive," then there is no reason to think that *some* of the what the emotions are might be the result of "cognitive" process while *some* might be the result of "non-cognitive" automated appraisal processes. While I will discuss this subsequently in greater detail, it is only under the strong interpretation of the cognitivist approach that emotion should be entirely irrelevant to the frame problem. And so, if emotion is to be of help, reason needs to be provided for our thinking it both likely false that *all* emotions are belief-dependent (*i.e.,* are "cognitive") and likely true that *at least some* of what the emotions are are undertaken by "non-cognitive" operations. If at least some of what the emotions are are "non-cognitive" (*i.e.,* undertaken by automated and autonomous appraisal processes) and these operations influence and inform practical reason and decision-making in the proper manner, then bringing emotion to bear on some of the frame problems that arise in these domains might hold some promise.

PROPOSITIONAL ATTITUDE THEORIES OF EMOTION

While there are a number of propositional attitude accounts of emotion and since a complete review would be both lengthy and likely unhelpful, I will focus here on providing a brief introduction to and overview of the principal tenets of both "pure" and "hybrid" propositional attitude accounts.

Solomon (1976) provides, perhaps, the most accessible discussion of the principal tenets of the pure propositional attitude account of emotion. He explains that "an emotion is a judgment (or set of judgments) … an evaluative (or "normative") judgment, a judgment about my situation and about myself and/or about other people."(Solomon, 1976 pp. 185-6) Solomon's account is a "pure" propositional one insofar as he claims that, "My shame *is* my judgment to the effect that I am responsible for an untoward situation or incident."(Solomon, 1976 p. 186). The pure propositional attitude approach to the emotions then takes as fundamental the idea that emotion types are indentified with and distinguished from each other (*i.e.,* other emotion types) by the particular propositional attitude involved. As Solomon's example suggests, a token of some emotion is the kind of emotion that it is because of the type of belief entertained. Shame, for example, requires the belief that "I am responsible for an untoward situation" while the tokening of an instance of fear would require, on a similar analysis, that the agent hold the belief that the current situation is dangerous.

The second feature of all "pure" propositional attitude theories is the claim that the physiological alterations associated with emotional responses are entirely (or largely) irrelevant to what an emotion is. Solomon explains, "That anger has a biological backing and includes sensations is inessential to understanding the emotion, though no doubt significant in certain measurements, which only *contingently* correlate with the intensity of the emotion or its significance."(Solomon, 1984 p. 249) That is, according to the propositional attitude approach there is a necessary conceptual connection between particular emotional states and particular beliefs while the connection between emotional states and particular physiological states is merely contingent. And so, according to the propositional attitude account, in order for one to be afraid for example, one *must* hold the belief that some aspect of their current situation is dangerous. One need not in order to be in some particular emotional state (fear, for example) exhibit any of the "contingent" physiological manifestations commonly associated with that particular emotional response. Put another way, since for an agent to be in an emotional state just is to say that the agent holds some particular belief, there is no contradiction in one being in a particular emotional state and exhibiting none of the physiological responses typically associated with such states. Since emotional states are propositionally defined, (*i.e.,* belief-identical) were one to exhibit the physiological responses commonly associated with an emotional response while not holding the right type of belief, one would be merely in an "aroused" but not an emotional state.[34]

Modifying Solomon's approach, Lyons (1980) outlines a "hybrid" account of what the emotions are that incorporates physiological states into a propositional attitude theory. Specifically, Lyons contends that both propositional attitudes and physiological manifestations (*i.e.,* states or alterations) are necessary for emotion. While both are required, Lyons posits a uni-directional causal relationship between the two aspects. Specifically, he takes an emotion to be the holding of a belief that in turn induces a particular physiological state. Setting out his account of the emotions, Lyons explains, "X is deemed an emotional state if and only if it is a physiologically abnormal state caused by the subject of that state's evaluation of his or her situation."(Lyons, 1980 pp.57-58). It is clear from this that Lyons' account is still fundamentally very much in line with Solomon's propositional attitude approach for it is belief that is ultimately fundamental to the induction and determination of an emotional state. This is so because the physiological aspect of any emotion is, according to Lyons' account, belief-dependent that is, *caused* by (*i.e.,* induced by and thus a byproduct of) belief. This implies that the fixation of a belief must occur prior to emotion.

Lyons clarifies his position further noting, "In general a cognitivist theory of emotion is one that makes some aspect of thought, usually a belief, central to the concept of an emotion and, at least in some cognitive theories, essential to distinguishing different emotions from one another."(Lyons, 1980 p. 33) He further maintains, in keeping with the general propositional attitude approach that, "the emotions presuppose certain judgments … as to what properties a thing possesses."(Lyons, 1980

[34] Kenny (1963) provides an analogous reductive account of emotion to propositional attitudes.

p.7) As such, "emotion is based on knowledge or belief about properties."(Lyons, 1980 p.138) By way of example, Lyons continues, "Saying I am angry at X or I love X implies that at some time I have apprehended certain qualities in X."(Lyons, 1980 p.71) And so, while Lyons' account does allow for physiological response to play a part in what emotions are, since these responses are caused by prior "cognitive" appraisals, it is belief that does the work (*i.e.,* is determinative of the kind of emotion entokened).

EMPIRICAL SUPPORT: COGNITIVE APPRAISAL THEORY

At base, then, Lyons account suggests that while emotion requires both that a particular belief be held and that a particular ("abnormal") physiological state be enacted, the latter is a byproduct of the former. Under this and all propositional attitude accounts, beliefs are either identical to or a component cause of emotion.

Parallel to the propositional attitude approach to the emotions in philosophy, and often relied upon to support these claims, is *cognitive appraisal* theory in cognitive psychology. Cognitive appraisal theory emphasizes the role of "higher cognitive" processes in the formation and induction of an emotional response. Like the propositional attitude approach, cognitive appraisal theories maintain that emotional responses are induced by (*i.e.,* caused by) those cognitive processes involved in belief-fixation. Specifically, according to cognitive appraisal theory, emotional responses require that the subject fix a propositionally formulated belief about the meaning of an event and the relationship between this event and the subject's ongoing projects and goals. (Lazarus, 1982, 1984)

As discussed previously, cognitive appraisal theory holds not just the uncontroversial view that emotions are preceded by information processing operations, but rather, that the information processing undertaken is decidedly "cognitive" in character - that the same information processing operations involved in belief-fixation are also employed in emotion. (Lazarus, 1982, 1984) Simply put, cognitive appraisal theory maintains that emotions are the result of "cognitive" processes, are *always* belief-dependent and thus are *always* the result of (*i.e.,* are caused by) some prior activity of "thinking."

Perhaps the most often referenced study in support of the cognitivist/ propositional attitude approach is that of Schachter & Singer (1962). The aim of this experiment was to show that while physiological arousal might be a component of emotion (as Lyons suggests), it is not determinative of the kind of emotion experienced. Put the other way around, Schachter and Singer aim to show that the kind of emotion experienced is determined by the type of cognitive evaluations/judgments reached – *i.e.,* by what the subjects believe and think and not by the mere fact that they are biochemically aroused. And so, the very same level of physiological arousal, they conjecture, will be experienced as a different emotion depending upon what the subject believes.

Subjects were divided into four groups. Group one was given a placebo, while groups two through four were given injections of adrenaline to increase their levels of physiological arousal. Of the three groups injected with adrenaline, the first was told truthfully what the physiological effects of the adrenaline injections would be, the second was given no information and the third was deliberately misinformed about the nature of the injection and its physiological effect. Half of the members of each group were then subjected to conditions designed to anger the subjects while those remaining were subjected to a condition designed to make the subjects happy. Specifically, each test condition employed a number of confederates/sham subjects working with the researchers and/or purposefully irksome and highly personal questionnaires in order to make the subjects “happy” or “angry.” Results were obtained from both Schachter and Singer’s observations of the subjects during the trials and by means of subjects’ responses to questionnaires following the experiment.

Schachter & Singer found that all subjects in the anger and happy conditions reported being angry and happy respectively. Those subjects given a placebo exhibited and reported minimal emotional arousal (relative to the other conditions). With respect to the three groups who received the adrenaline injection, the following was found. Those who were told about the effects of the injection both exhibited and reported the lowest levels of emotional arousal. Those who were told nothing about the nature of the injection or its effects exhibited and reported an elevated level of emotional arousal. Finally, those purposefully misinformed about the effects of the injection exhibited and reported the highest level of emotional arousal. These findings cognitivist/propositional attitude theorists maintain establish that a subject’s level of emotional arousal is entirely dependent upon their appraisal of their environmental context. That is, these findings are taken to establish that emotions are differentiated and labeled by subjects based on their appraisals of (*i.e.,* beliefs about) their situations. Propositional attitude theorists maintain that these results support their claim that emotions are identified (and differentiated) not by physiological states but rather solely upon the basis of the cognitive appraisals undertaken by the agent – *i.e.,* the beliefs (and/or desires) they currently hold.

While there a number of cognitive appraisal accounts on offer[35] all share the commitment to the dual claims that an emotional response requires prior "cognitive" appraisal of the situation/stimuli (*i.e.,* that a belief be fixed) and that these prior cognitive appraisals are the cause of any physiological responses associated with an emotional response of a particular kind.[36] Analogous then to Lyons' claim, cognitivist theories hold both (a) that all emotional responses are the result of higher level "cognitive" processes of appraisal (*i.e.,* "thinking") and (b) for a subject, for example, to be afraid of X requires that one fix a belief about X – namely that X is dangerous (or at least "fear-worthy"). Whatever physiological alterations come to manifest are secondary to and dependent upon the processes of belief-fixation.

ON COGNITIVIST/PROPOSITIONAL ATTITUDE ACCOUNTS OF EMOTION AND FRAME PROBLEMS

Cognitive appraisal/propositional attitude theories of emotion, as outlined above, maintain that emotions are either identical to or are caused by processes of "cognition" – of the operations of belief-fixation. Since, as Fodor's arguments aim to establish that any system engaging in "thinking" is computationally inexplicable and un-model-able (since it must be able to solve the frame problem), quite trivially it follows that if the emotions are identical to or caused by beliefs then the emotions too would be prone to the frame problem and would also be computationally inscrutable. Clearly, under any account in which it is maintained that *all* emotions are belief-dependent, the proposal that by bringing the emotions to bear some light might be shed on the question of how it is that we contend with some of the frame problems that arise

[35] Fridja, Kuipers & TerSchure (1989), Fridja (1986), Oatley & Johnson-Laird (1987) Gehm & Scherer (1988), Roseman, Spindel & Jose (1990), Smith & Ellsworth (1985), Reisenzein & Hofman (1990), Smith & Lazarus (1993), Lazarus (1982), (1984) & (1991).

[36] That is, in those accounts that allow for physiological responses to play a role in emotion at all.

in practical reason, would be ill-received – viewed as a quite pointless proposal that, as decidedly circular, would be a "non-starter." Put somewhat differently, if the operations of belief-fixation are, as Fodor suggests, computationally un-model-able and if emotions are invariably belief-dependent, then there would be no reason whatsoever for thinking that by bringing emotion to bear we should have any more hope of understanding the workings of mind than we do by ignoring it. This is so because under the cognitivist/propositional attitude account the emotions are just another cognitive event in need of explanation and modeling. And so, bringing emotion to bear in the hopes of explaining how it might be that we contend with some of the frame problems that arise in practical reason and decision-making would fail to offer any explanation at all. Rather, so doing would serve only to chase the problem about.

SOME STANDARD PUZZLES FOR PROPOSITIONAL ATTITUDE/COGNITIVIST ACCOUNTS OF EMOTION

There are a number of rather standard objections to or puzzles raised about the propositional attitude/cognitivist account that should be briefly set out and discussed.[37] While I will discuss Griffiths' rather global critique of the propositional attitude/cognitivist approach shortly, since his assessment of the limited powers of conceptual analysis as a tool for understanding what the emotions are is compelling, I will focus on examining the empirical support for and the objections to this approach.

Propositional attitude/cognitivist accounts are often charged with being unable to account for what have come to be called in the literature "objectless emotions" - those emotional states such as generalized anxiety or depression that lack specific intentional objects. Since, at least in some cases, there is no particular thing about which one is

[37] This rather standard set of puzzles for propositional attitude/cognitivist theories outlined here is a distillation of a number of versions of these objections made by a number of different theorists. Most of these, in some form or other, can be attributed to Zajonc (1980, 1984).

anxious, there can be no explicit proposition entertained and thus no specific belief that is identical to or the cause of the emotion.[38]

Second, it is claimed that some propositional attitude approaches, particularly those like Solomon's in which physiological alterations (*i.e.,* affective responses) are taken to be entirely irrelevant to the question of what emotions are, are either (a) committed to positing far too many emotions or (b) unable to provide an account of why certain appraisal/evaluations are emotional while others are not.

With respect to the first option, the objection continues, under those accounts in which emotions are belief-identical, what reason do we have for thinking that *all* cognitions (*i.e., all* instance of belief-fixation) are not emotions? If all cognitive appraisals (*i.e.,* all instances of belief-fixation) are emotions, then there would be a vast number of emotions – one for each belief fixed in the limit.

Alternatively, if such accounts wish to maintain vernacular emotion types (*i.e.,* disallowing that every belief-fixing episode is an emotional episode) then some explanation is needed for why/how it is that some appraisal/belief-fixing episodes are emotions and others are not. Seemingly, the objection continues, something in addition to belief would be needed to determine and differentiate the emotion types. Taken together then, those approaches that maintain that emotions are belief-identical face the problem of having to explain how/why it is that some beliefs are emotional and others are not.

This concern, while raising doubt about the tenability of a "pure" propositional attitude/cognitivist approach to emotion, is not directly applicable to those accounts

[38] A similar concern is often raised with respect to the inability of propositional attitude/cognitivist approaches to offer an account of "mood" which appear to be both "emotional" in character yet objectless (*i.e.,* non-intentional).

that allow for physiological states to play some role in determining and differentiating emotions. For example, Lyons' account in which emotions are taken to be reducible to both a particular belief and the physiological change (*i.e.,* affective response) induced by this belief, is largely immune to this particular charge. Specifically, under this variant of the propositional attitude/cognitivist approach, it would follow that not all belief-fixing episodes are emotion-episodes precisely because not all beliefs result in the induction of an alteration in the subject's physiological state of the relevant sort.

Those variants (like Lyons') in which physiological states are incorporated, however, face the question of how to explain those situations in which some particular affective response (*i.e.,* physiological alteration) is induced by the "wrong" kind of belief. Consider for example a situation in which the physiological state associated with "joy" is induced in response to the belief that some situation is a highly dangerous one (*e.g.,* base-jumpers, street drag-racers, waterfall barrel-riders, and other "extreme sport" enthusiasts). If each emotion requires *both* that a particular type of belief be entertained and that a particular kind of ("abnormal") physiological state be induced, then cases such as these in which the "wrong" kind of state is induced, are at the least somewhat puzzling. Specifically, the concern continues, one wonders what emotion such folks are *really* in.

If belief is doing the work in determining the type of emotion entokened, then it would seem that one should say that these folks are *really* afraid (because they genuinely believe the situation a dangerous one). However, at least prima facie, this does not appear to be the case at all. If belief is taken to be in fact determinative of the emotion entokened (*i.e.,* if such folks *really* are afraid because of what they believe irrespective of the particular physiological state induced) then one wonders what role physiological states are actually playing in determining and differentiating emotion. If physiological states play no role, then the original criticism outlined above would remain applicable, for some account would still be needed as to why certain beliefs are emotions and others are not. If physiological states do play at least some role, then an explanation is needed of what precisely this role might be.

Alternatively, if the subject's physiological state is doing the work in determining the type of emotion entokened, then it would seem that one should say that these folks *really* are "happy/joyful" irrespective of their beliefs about the dangerousness of the situation. However, if this second option is adopted, it becomes unclear what role belief is actually playing in determining and differentiating emotion. That is, endorsing this option would appear to come at the price of the rejection of the principal claims of the cognitivist/propositional attitude approach (*i.e.,* that *all* emotions are belief-dependent).

Propositional attitude/cognitivists theories are sometimes criticized for their inability to contend with cases in which agents experience genuine emotions in response to imagined (imaged) content (Greenspan, 1988). If so, since the propositional attitude/cognitivist approach holds that all emotions are belief-dependent (*i.e.,* they are all reducible to or caused by belief) and the subjects in such imaginative cases explicitly do not believe the content of these imaginings, then it would follow that either the agent is not experiencing a genuine emotion (as some propositional attitude/cognitivist theorists maintain and Greenspan denies) or the propositional attitude/cognitivist approach to emotion is incomplete.

Furthermore, the propositional attitude/cognitivist approach is charged with being unable to provide an adequate explanation for those cases of emotions that occur in the absence of belief. There are, so it would appear, instances in which we respond emotionally (long) before we have fixed a belief. I freeze, then recoil, my gut constricts, heart rate increases and I attend immediately to the sound near me – I am afraid. Only later do I come to believe that in fact I am standing near bags of trash teeming with rats. That one can respond emotionally prior to the fixation of a belief poses a puzzle, of course, for the propositional attitude/cognitivist approach which claims that all emotions are belief-dependent.

Likewise, propositional attitude/cognitivist approaches are charged with being unable to account for those cases in which one becomes "emotional" *in spite* of one's beliefs. For example, many believe that being in a graveyard at midnight is no more dangerous than being in one at noon (*i.e.,* there are no ghouls, demons, zombies, beasties and the like), yet at least some of these folks will be genuinely afraid. This, of course, presents a puzzle for the propositional attitude/cognitivist approach for it would appear that their only recourse is to claim either (a) that these folks *really* are not afraid, or (b) that they *really* do not believe that graveyards are zombie-free at midnight.[39]

In addition to this rather standard set of objections to or puzzles about the propositional attitude/cognitivist approach briefly discussed above, Griffiths (1997) argues that the entire propositional attitude approach to emotion is methodologically suspect. Before outlining Griffiths' critique, however, I will examine briefly the empirical evidence most often relied upon in support of the cognitivist (and, by association, the propositional attitude) approach.

[39] Disgust also provides some nice examples. Normal subjects, it has been found, will invariably refuse to drink water into which sanitized rubber dog feces, fake vomit and/or plastic roaches have been briefly dipped. All subjects know full well – *i.e.,* believe – that the water is in no way contaminated, yet all are disgusted by the prospect of drinking it.

Zajonc (1980), (1984) and Berkowitz (1994) heavily criticize the research methodology relied upon in many of the studies offered in support of cognitivist/propositional attitude approach to emotion. Specifically, Zajonc and Berkowitz note that all of the experiments offered in support of the cognitive appraisal approach rely upon subjects reporting upon their own cognitive processes. Taking Schachter and Singer's study as an example, subjects are required, upon completion of the trial, to reconstruct from memory the processes undertaken by them during the trial and verbalize these to the examiners. While most now acknowledge the unreliability of evidence from introspection, with regard to the particular paradigms employed in the studies taken to support the cognitivist position, Phaf (1996) argues that the conclusions rely upon the assumption that subjects are capable of accurately reporting upon their cognitive processes when many of these are likely to be unavailable to consciousness. Similarly, Berry (1987) suggests that experts consistently report that they employ "rational" problem-solving strategies that bear little resemblance to the heuristic strategies that they actually employ when solving problems in their particular domains of expertise. And so, if evidence from introspection is suspect and if subjects are prone to thinking that they are employing one kind of process when they are in fact employing one quite different, then we have reason for thinking the empirical evidence relied upon to support the cognitivist approach to be problematic.

Furthermore, and pertinent in particular to the often cited conclusions of Schachter and Singer (1962) there are the findings of Parkinson & Manstead (1992, 1993), Nisbett and Wilson (1977), Gazzanaga & Smiley (1984). These suggest that subjects often construct, invent or *confabulate* explanations and rationales for their behavior/physiological states when the cause is uncertain.[40] And so, under conditions of uncertainty, subjects tend to explain their behavior as the result of some inner mental process though the actual cause is attributable to some other factor (*e.g.,* an environmental cue.) And so, if subjects are known to confabulate (*i.e.,* invent) explanations for their behavior/physiological states under conditions of uncertainty, and the injection of adrenaline causes physiological arousal, then these subject's self-reports would be suspect. The phenomenon of confabulation, then, provides an alternative and an at least equally plausible explanation for Schachter and Singer's (1962) results. This, when taken in conjunction with the general unreliability of evidence from introspection, suggests that the empirical support for the cognitivist/propositional attitude approach is by no means decisive.

GRIFFITH'S CRITIQUE OF THE PROPOSITIONAL ATTITUDE APPROACH

The propositional attitude approach to emotion relies, Griffiths (1997) contends, upon the analysis of emotion terms to define and differentiate emotions. This general approach, he continues, "assume[s] that a concept is entirely constituted by what is currently believed about its referent." (Griffiths, 1997 p. 24) It is then conceptual analysis, Griffiths suggests, that undergirds the philosophical support for the claim that all emotions are belief-dependent (*i.e.,* "cognitive"). Griffiths, however, argues that conceptual analysis is a deficient tool for understanding the emotions. (Griffiths, 1997 p. 35) He explains,

> Our best definition of water is the formula HOH, but this could not have been established by conceptual analysis! Ordinary English speakers in the past did not know that water was HOH, but this is still an important part of what "water" means. (Griffiths, 1997 p. 35) … Conceptual analysis can tell us only what people currently

[40] Thanks to that member of P. Greenspan's seminar (2002) on emotion - I unfortunately do not recall whom – who first pointed out the potential relevance of the phenomenon of confabulation as an alternative interpretation of Schachter & Singer's findings.

> believe about emotion. There are many puzzles about emotion that cannot be resolved by mere analysis of what people currently believe. To answer these questions it is necessary to look at the referent of the concept as well as the concept itself. It may even be necessary to revise the concept in order to better accommodate its referent.[41] (Griffiths, 1997 p. 39)

And so, given the standard set of puzzles faced by the propositional attitude/cognitivist approach outlined above and the inability, Griffiths contends, of conceptual analysis to offer insight into the nature of the referent of our emotion terms, we have reason for thinking the propositional attitude/cognivitist approach to emotion to be problematic and at the very least an incomplete account of what the emotions are.

While my goal in this section is not to provide a complete response to the question of "what the emotions are," we do have reason for thinking the account offered by the propositional attitude/cognitivist approach to be problematic. I conjectured earlier that initial resistance to the suggestion that bringing emotions to bear on the frame problems that arise in practical reason and decision-making might hold promise for explaining how we might contend with such problems, is likely rooted in a particular view about the nature of the emotions themselves – specifically in a tacit assumption that something like the propositional attitude/cognitivist approach must be true. Since, however, we have reason for thinking this initial objection (or charge of irrelevance) to be based upon theoretical foundations that are empirically unsupported and methodological problematic, we have reason for thinking the claim that *all* emotions are belief-dependent to be unsupported. And so, if at least some of what the emotions are might be undertaken by "non-cognitive" processes, then we would have no reason for thinking that bringing emotion to bear on some of the frame problems that arise in

[41] Griffiths continues, "But the proponents of the propositional attitude approach seem quite undaunted by its bleak history and are so firmly committed to the methodology of conceptual analysis that they may regard my cure as worse than the disease." (Griffiths, 1997 p.39)

practical reason and decision-making should be irrelevant. Likewise, in advance of our considering specific proposals, we also do not appear to have reason for thinking the proposal to be an obvious non-starter.

The point here is a modest one. While there are other considerations that must be addressed (*i.e.,* it remains to be shown that emotion helps to expedite and increase the accuracy of practical reason and decision-making) there appears to be no reason - in advance of our considering specific proposals - for thinking emotion to be obviously irrelevant.

AUTOMATED APPRAISAL THEORY

The second of two broad approaches to the question of what the emotions are, the *automated appraisal/"basic" emotion* approach, has received comparatively little philosophical attention and, as such, is somewhat (philosophically) undeveloped. In what follows, I will briefly outline this approach and present a representative sample of the evidence in support of the automated appraisal account of emotion. Following this, I will argue that while the automated appraisal approach is unlikely to provide a complete response to the question of what the emotions are, there is ample evidence for thinking the processes outlined by such accounts to be at least part of the picture (*i.e.,* a partial response to the question of what the emotions are). And so, I will suggest that whatever the emotions turn out to be, at least part of what they are will be explicable in terms of the operations of automated appraisal mechanisms/basic emotion programs.

ZAJONC'S AUTOMATED APPRAISAL ACCOUNT: THE EVIDENCE FROM "AFFECTIVE PRIMING"

Lazarus' claim that emotions are cognitive in the manner outlined above instigated a debate between him and Zajonc. The Lazarus (1982, 1984) and Zajonc (1980, 1984) debate concerns the extent to which cognition (in Lazarus' and Fodor's terms) is required for emotion.[42] In contrast to the cognitivist approach, Zajonc argues for a dual-systems model whereby affective/emotional responses are undertaken by autonomous and automated "non-cognitive" systems that are functionally distinct from those implicated directly in cognition (*i.e.,* belief-fixation).

That subjects generate preferences for stimuli that cannot be consciously perceived (Kunst-Wilson & Zajonc, 1980; Murphy & Zajonc, 1993) is taken by Zajonc to support the hypothesis that affective/emotional appraisals are undertaken by information processing systems that are distinct from those implicated in belief-fixation. Specifically, Zajonc's "affective priming hypothesis" maintains that affective appraisals occur without cognitive appraisal. While there is considerable evidence in support of various formulations of the affective priming hypothesis, I will, with the aim of keeping discussion short, focus on presenting only few representative results.[43]

Zajonc & Murphy (1993), for example, studied the effects of cognitive and affective priming on a judgment task involving previously neutral stimuli (novel Chinese characters). Happy or angry faces were employed as affective primes while circles and squares of different sizes and symmetrical or asymmetrical geometric shapes were employed as cognitive primes. Subjects were exposed to the primes - images of faces or geometric shapes - either briefly for 4 msec or longer for 100msec prior to the

[42] Again, while the terminology is decidedly unhelpful what is at stake in this debate is not the issue of whether or not emotion requires prior information processing but rather whether emotions require information processing by the same systems implicated in belief-fixation and planning. Both agree that some form of information processing is prerequisite to the induction of an emotion. Lazarus maintains that the kind of information processing undertaken is "higher-level" (*i.e.,* doxastic) while Zajonc holds the kind of information processing to be rather perceptual in nature.

[43] Izard, 1991, 1992, 1993; Etcoff & Magee, 1992; Bargh, Chaiken, Govender & Pratto, 1992; Greenwald, Klinger & Lui, 1989; Markus & Kitayama, 1991; Niedenthal, 1990; Martin, Williams & Clark, 1991; Neidenthal & Kitayama, 1994; Dagleish & Watts, 1990; Hunt & Ellis, 1999; Kraiger & Isen, 1989; Kunst-Wilson & Zajonc, 1980; Pratto & John, 1991; Ingram, 1984; Teasdale, 1983; Niedenthal & Showers, 1991; Niedenthal & Setterlund, 1994.

presentation of the Chinese character. It was found that the affective primes (happy/angry faces) when presented for the short 4 msec duration significantly influenced the subjects evaluations of the "goodness" or "badness" of the target characters. The cognitive primes (square, circles and symmetrical/asymmetrical geometric shapes), however, were found to influence subjects' judgments concerning whether the character represented an object that was round, square, symmetrical or asymmetrical only when presented for the longer duration (100 msec) and not for the shorter. Zajonc & Murphy explain these results noting, "when affect is elicited at levels outside of conscious awareness, it is diffuse and its origin and address are unspecified. Because of its diffuse quality, nonconscious affect can 'spill over' onto unrelated stimuli."(Zajonc & Murphy, 1993 p. 736) That affective appraisals can influence cognitive processes in this manner, Zajonc concludes, suggests both that these processes (*i.e.,* affective appraisal and cognitive appraisal) are distinct and that automated emotional appraisals occur independently of and in many cases prior to cognitive appraisals. As autonomous and rapid, the output of these automated appraisal operations inform and influence cognitive processes in the manner described.

Similarly, DeHouwer, Hermans, & Eelen (1998), in an affective variant of the Stroop (1935) task, found further support for an automated and distinct emotional processing (sub)system. Subjects were presented with a discrimination task in which they were to determine as quickly as possible whether a target word was an adjective or a noun. In response to a noun the subjects were to respond by saying aloud "positive" and "negative" to an adjective. The target words themselves were either positively, negatively or neutrally affectively valenced. Again, in keeping with the Stroop-like nature of the experiment, subjects were explicitly instructed to ignore the emotional valence of the target words themselves. Similar to the original Stroop test, subjects' response latencies (the time required to respond) were measured. It was found that the response latencies were significantly longer in cases in which the valence of the target word was in contrast with their verbal response. (*e.g.,* in cases in which the word was a noun and thus the correct response should be "positive" subjects took considerably more time to respond when the target word was itself negatively valenced.) This suggests that though subjects were given explicit instruction to ignore the emotional valence of the target words such appraisals were automatically generated and brought to bear, thus interfering with the discrimination task. DeHouwer's finding appear to support a dual-system model whereby affective appraisal occurs automatically, independently of and prior to cognitive appraisals and in so doing informs (in this case interferes with or facilitates) cognitive processes.

EKMAN'S AUTOMATED APPRAISAL ACCOUNT: THE BASIC EMOTIONS

In addition to the evidence from affective priming, anthropological support for automated appraisal theory is available as well. In response to the prevailing view at the time that emotions are "social constructions" (Harre, 1986; Averill, 1980; Oatley, 1993) – the view that emotions are culturally inculcated learned behaviors that are acquired in much the same manner as one learns any other cultural mores.[44] Underlying social constructivist accounts quite generally is the claim that one cannot experience joy, sadness, anger and so on without first having learned/been taught to do so (*e.g.,* in much the same way that one learns which fork to use for which course). Clearly, under

[44] Occasionally social constructivists liken the acquisition of emotion to the learning of language. While I think few modern social constructivists would endorse this analogy, historically the claim is understandable given the view of language acquisition predominant at that time.

the social constructivist account having an emotion would be a decidedly cognitive activity.

In response to the social constructivist's claim, Ekman, Sorenson & Friesen (1969) and Ekman & Friesen (1971) conducted a set of studies (similar to those undertaken by Darwin) on remote populations in Borneo (1969) and New Guinea (1971). Due to their isolation, members of neither group had had any interaction with Western culture (*e.g.*, none had seen western photographs or films, knew English, or had lived in any Western settlement.) As isolated from western influences, there would have been no way for members of either population to have learned any of the western emotions posited by social constructivism. Since the studies are quite similar, I will focus on briefly outlining the findings of Ekman & Friesen's (1971) study involving members of the Fore population of New Guinea.

Fore language-speaking subjects (both adults and children) were presented with a set of three photographs (each depicting the face of an American expressing an emotion) and told a short story (in Fore and by a Fore member of the research team) involving a particular emotion. Subjects were then asked to select which of the three pictures matched that described by the story. Ekman & Friesen found that the pictures intended by them (*i.e.,* Westerners) to match the story were also chosen by Fore speaking subjects. That is, the Fore selected the same facial expressions in response to emotion-involving stories as did the Western researchers. Next, Fore subjects were videotaped and photographed making the facial expression appropriate to each of the stories. With these images, the experiment was repeated on American subjects. American college students too chose those Fore facial expressions that tracked the emotion involved in each of the stories. Simply put, both American college students and members of isolated tribes in Borneo and New Guinea (neither of whom knew anything about the other's cultural or purported emotional mores) chose and expressed the same emotions in response to the same kinds of stories.

The (1971) study focused on six facial expressions – Joy/happiness, Distress/sadness, Fear, Anger, Disgust and Surprise – the so called "basic emotions" - which Ekman concludes are universal (pan-cultural), innate and undertaken by automated and autonomous (sub)systems. Additional support for the claim that these basic emotions are automated and universal comes from Ekman *et al.*, (1971) in which Japanese and American subjects were videotaped while watching both neutral/pleasant and stress-inducing film shorts. While there are marked differences in the socially acceptable "display rules" for emotional expression between the two cultures (Japanese evidently find excessive emotional displays unpalatable, even rude in some instances, while Western cultures tend to be more permissive of such displays), the question Ekman & Friesen considered was the extent to which the social constructivist approach to emotion is correct. If correct then Japanese subjects should really be emotionally unfathomable and entirely affectively outré to the Westerner for the simple reason that they are not actually experiencing the same emotions in response to stressful images as the Westerner. That is, the question under consideration is whether Japanese and American subjects actually respond to the stressful images differently because of their particular cultural inculcations?

American and Japanese subjects were observed in two conditions. First, the subjects were videotaped while watching neutral and stressful films alone. In this condition the facial expressions of both the Japanese and American subjects matched to an extremely high degree (specifically, a .96 correlation was found). In the second condition, an interviewer was present in the room while each subject viewed the film. Interestingly, in this situation Japanese and American expressions differed markedly. Notably, Japanese subjects tended to "politely smile" during the trial - overall displaying more positive expressions and fewer expressions of disgust than did their American counterparts. Of greater interest however was the finding that when the videotapes of Japanese subjects were analyzed in slow-motion, even these subjects expressed – though only briefly – identical facial expressions in response to the stressful films as did the American students. That is, both Japanese and Americans were found to automatically respond identically to the stressful film clips. It was only after the subject's cognitive processes were able to catch up to these automated processes (a few hundred milliseconds later), Ekman argues, were the Japanese subjects emotional expressions curtailed or coerced into the more culturally acceptable display of a polite smile. With respect then to Ekman's aim, it appears that Japanese subjects both experience and express the same emotions in response to the films as did westerners, though the ultimate manner in which emotion is displayed, communicated or conveyed to others (*i.e.,* the particular "display rules" employed) do exhibit cultural variation.

Support for the nativist claim underlying Ekman's account comes from Eibl-Eibesfeldt (1973). In this set of experiments, the emotional expressions of children and infants born blind and deaf were studied.[45] Such infants, it was found, make identical facial expressions as those who are unaffected. Since the blind and deaf infants could not have possibly learned to make such expressions - as the social constructivists claim -

[45] Subsequently, thalidomide-affected infants were studied to determine whether emotion expression might be learned tactilely. Here too, it was found that these infants generated identical expressions as unaffected infants.

these studies are offered as a poverty-of-the-stimulus type argument in support of the nativist claim.

The species-universality of emotional recognition and display when taken in conjunction with the nativist claim, resulted in Ekman positing the existence of 6 (or 7) innate, automated and modularly realized affective appraisal mechanisms - the "basic" emotions or emotion programs as they are sometimes referred to in the literature. Underlying the account is the claim that (at least some) emotional responses are non-cognitive, automated and complex coordinated sets of responses to both internally generated and externally perceived stimuli. The activation of each program results in specific and measurable alterations in facial and vocal expression, musculoskeletal modifications such as changes in stance, visual and auditory orienting and particular muscular responses such as flinching, constriction of the gut, increased heart rate and respiration, alteration of hormonal/endocrine system homeostasis (*i.e.,* increase or decrease in the production or release of hormones) and changes in autonomic nervous system activity. Significantly, as autonomous neural programs, the appraisals and alterations induced are undertaken and transpire without cognitive mandate. He explains,

> Since the interval between stimulus and emotional response is sometimes extraordinarily short, the appraisal mechanism must be capable of operating with great speed. Often the appraisal is not only quick but it happens without awareness, so I must postulate that the appraisal mechanism is able to operate *automatically*. It must be constructed so that it quickly attends to some stimuli, determining not only that they pertain to emotion, but to which emotion. (Ekman, 1977 pp. 58-59)

And so, once activated emotion programs automatically result in manifold and highly predictable and specifiable changes in an agents facial expression, voice, stance, endocrine system and autonomic nervous system activity. Emotions (at least some of them) then just are the manifold and coordinated set of changes induced in an agent by an automated and neurally "hard-wired" modular appraisal mechanism. To token a particular emotional response under Ekman's account just is for a particular emotion-program to be activated and the manifold and coordinated alterations to unfold.

That each emotion can be specified and differentiated by physiological manifestation, as Ekman suggests, is an empirical claim. While work is ongoing, there is significant evidence of physiological and autonomic nervous system differentiation of the basic emotions.[46] Given the available evidence, Ekman maintains, there is sufficient support for our thinking there to exist at least seven discernable and distinct human emotion/affect programs – the "basic emotions" - which roughly track the vernacular emotion terms of "disgust," "contempt," "anger," "fear," "joy," and "sadness."[47] The existence of a seventh basic emotion "surprise/startle" is currently under debate. That automated, autonomous and non-cognitive processes mediate startle is not under dispute. Rather, what is at issue is whether startle should be considered an emotion at all.

In addition to the automated appraisals undertaken by these emotion programs, cognitive processes too may be engaged in appraisal under the dual-systems account outlined. That is, in keeping with the general framework of Zajonc's account, Ekman's model is a dual-system one insofar as there are both discrete and autonomous emotional appraisal mechanisms and the processes of cognition. And so, automated appraisals

[46] Ekman, Levenson & Friesen 1983; Levenson, Ekman & Friesen, 1990; Levenson, Cartensen, Friesen & Ekman, 1991.

[47] Ekman is not relying on analysis to specify emotion kinds. Our vernacular term "fear" for example and what "fear" is according to the basic emotion program account may be very different things. This follows, for Ekman's basic emotion programs need not track *all* instances of what vernacularly we might consider fear - such programs may include more or less than our vernacular emotion categories suggest. Put another way, because the affect program account requires, as Griffiths (1997) argues, conceptual revision, there is no reason to think that these programs must or should track our existing folk emotion categories.

and cognitive appraisals, since they are undertaken in parallel, may match or they may differ, thus providing a means by which to contend with one of the puzzles confronting the cognitivist approach. Specifically, under this two-systems account there would no mystery as to how one could both believe that one has no reason to be afraid of X, while fearing it nonetheless.

While Ekman considers them distinct systems, he is, of course, not claiming that they are entirely non-interactive. Rather, given the evidence from affective priming it is clear that affective appraisals influence cognitive ones. Likewise, since it is also clear that thoughts (*i.e.,* internally generated quasi-perceptual images) can give rise to emotions, Ekman is not claiming that affect programs cannot be activated by what a subject comes to believe. And so, while the particular mechanisms of affective and cognitive appraisal are distinct (*i.e.,* mediated by distinct systems), affective appraisals can influence cognitive processes and the output of cognitive processes may be accepted as input into the automated appraisal system, thus allowing for belief, once evaluated by the automated appraisal mechanism, to result in an emotional response.

CLARIFYING THE NATIVIST CLAIM OF AUTOMATED APPRAISAL THEORY

It is worth further clarifying the nativist aspect of Ekman's claim. He is not, of course, suggesting that all people are afraid, for example, of the same things. And so, the particular stimuli that will activate an emotion program are not entirely innately fixed. It is, however, clear from Ekman's findings that the "output" of each emotion program is innate and species-universal. And so, while subjects from differing environments (including whatever culture they are apart) may not, for example, fear the same particular things, all normal people will generate the same emotion-specific characteristic responses to evolutionarily recurrent situation types.[48] Emotions, Ekman explains, "evolved for their adaptive value in dealing with fundamental life tasks."(Rosenberg & Ekman, 1994 p.216) And so, there are, he continues, commonalities in the kind of stimuli that will elicit an automated, complex and coordinated emotional response in humans. Put another way, Ekman suggests that while it need not be the case that we all fear the same particular stimuli, we do all express and experience fear and we do all express and experience fear in response to the same kinds of evolutionary recurrent situations.

The "input" aspect of the appraisal mechanism, while not entirely innately fixed (*i.e.,* we can learn to respond emotionally to new stimuli) is, however, heavily evolutionarily constrained. And so, in addition to the considerable evidence that all normal humans respond with fear, for example, to looming objects, snake-like shapes, displays of teeth, and growls, Ekman, drawing upon Seligman's (1971) learning preparedness account, suggests that our evolutionary history has (pre)disposed us to make particular associations with greater ease. That is, we are (pre)disposed to quickly learn to respond emotionally to certain kinds of stimuli – if these obtain in the environment.[49] And so, while we may not all respond with fear to same kinds of things,

[48] There is, however, a body of evidence from animal studies suggesting that some fear-inducing stimuli are innate in the "input" aspect as well. For example, looming objects, snake-like shapes, displays of teeth and growls appear to serve as universal fear-inducing stimuli. Additionally, there are a number of species-specific fear inducing stimuli – for example, mice will, though they have never seen a cat, exhibit a characteristic fear response in the presence of a feline scent. Similar results have been found with respect to disgust. Monkeys, too, appear to exhibit a fear response to snakes regardless of whether they have seen a snake before or seen a conspecific react to the presence of snake. *C.f.* Ohman, (1993, 2001)

[49] Likewise, we are also (pre)disposed to make some associations with less ease and others only after a considerable number of learning trials. Still others, it appears, we are (pre) disposed to not make at all. We are then "prepared," Seligman suggests, by our particular evolutionary history to make certain kinds of associations with greater ease than others. For example, while we quickly learn many associations (some phobias, for example, are acquired with incredible ease in

we are, as a species, disposed to learn and to respond emotionally to the same kinds of evolutionarily recurrent situations and stimuli - were these to obtain in our particular environment. As such, what causes the activation of a particular agent's automated appraisal mechanism will depend both upon our species-specific evolutionary history (*via* both certain "hardwired" stimuli and learning preparedness) and the agent's own particular interactions with his/her environment – which includes, of course, his/her culture.

Likewise, while all normal subjects will respond characteristically to the same evolutionary recurrent situation types, there is also a role for learning with respect to the particular "display rules" employed. And so, while both Americans and Japanese, for example, both exhibit the same characteristic emotional responses to the same kinds of stimuli, the manner in which these automated responses are *ultimately* "displayed" (*i.e.,* communicated to others) is, as culturally dependent, clearly learned.

Ekman's findings then provide reason for thinking that at least some emotions are innate and universal (pan-cultural). However, if the automated appraisal account is to serve as a plausible alternative to the propositional attitude/cognitivist model, we need reason for thinking there to exist the kinds of automated affective appraisal module(s) postulated by, among others, Ekman and Zajonc.

THE EVIDENCE FROM NEUROANATOMY

In outlining his "dual – pathway" model of emotion, LeDoux (1998) provides further support both for Ekman's automated appraisal approach and the dual-systems

some instances after only one learning trial), it is incredibly difficult to get children – and many adults - to be afraid of things like automobiles and handguns. *C.f.* Seligman & Hager's (1972) discussion of the "Garcia effect" (named after John Garcia) who found that rats can learn to associate taste with nausea, but not sounds or sights – thus suggesting that not all stimuli can come to serve as conditioned stimulus of an unconditioned response.

hypothesis outlined by Zajonc. LeDoux argues that emotional appraisals are undertaken by means of two anatomically distinct pathways – the thalamo-amygdala "low road" and the thalamo-cortico-amygdala "high road." I will provide a brief review of his account and findings.

THE THALAMO-AMYGDALA PATHWAY: "THE LOW ROAD"

Normal rats, when presented with a paradigm in which a footshock (US) follows the presentation of an auditory stimulus (CS), quickly learn to associate the tone (CS) and the footshock (US). Of surprise was the finding that the abalation or lesioning of the auditory cortex had absolutely no effect on the ability of the rats to develop conditioned responses. While cortical involvement is unnecessary for the development of the conditioned response, conditioning is entirely prevented when the rats' auditory thalamae were ablated. Such findings, LeDoux explains, are relevant for in order for conditioning to occur, "the audiotory stimulus has to rise through auditory pathways from the ear to the thalamus, but does not have to go the full distance to the auditory cortex." (LeDoux, 1998 p. 152) This is, of course, relevant for these findings suggest that emotional learning and appraisal neither relies upon nor requires cortical involvement. Since the role of the thalamus, all agree, is decidedly "non-cognitive" – it is often attributed the role of a processing-inert relay-station (Zigmond *et. al.,* 1999) – it would appear that the processes by which rats respond with fear to (new) stimuli is not a "cognitive" one. Rather, LeDoux suggests, this process is subserved by an autonomous and dedicated non-cognitive affective (sub)system – an automated affective learning and appraisal module.[50]

[50] Again, the term "cognitive" (about which the Lazarus-Zajonc debate revolves) is unhelpfully retained to some degree in LeDoux's discussion. The non-cognitivist claim, once again, is not

Through further ablation studies in conjunction with WGA-HRP tracer injection experiments, LeDoux (1992) discerned that by lesioning the connections between the thalamus and the amygdala, conditioning could be completely prevented. Since, as a number of studies confirm, ablation of the amygdala completely prevents conditioning, it is then the amygdala and not the thalamus that is responsible for mediating emotional conditioned learning and appraisal. This direct thalamo-amygdala pathway, which functions entirely independently of any cortical involvement (and thus "thinking") relying only upon the minimally processed information relayed to it from the thalamus, is significant for,

> The fact that emotional learning can be mediated by pathways that bypass the neocortex is intriguing for it suggests that emotional responses can occur without the involvement of the higher processing systems of the brain, systems believed to be involved in thinking, reasoning and consciousness. (LeDoux, 1998 p.161)

These findings also lend support to Ekman's claim that, while the output of the automated appraisal module is fixed, the kinds of stimuli that will trigger an emotional response need not be. *Via* the mechanism of emotional learning set out by Cahill *et. al.,* (1992, 1993, 1994, 1995, 1996) and LeDoux (1995, 1998), it is clear that subjects can come to learn to respond emotionally to a vast range of "trigger" stimuli. However, the kinds of stimuli that will serve as triggers (*i.e.,* inducers of an emotional response) will be, LeDoux claims, directly influenced by our species-specific (pre)dispositions to make particular kinds of associations – in the manner suggested by Seligman's learning preparedness model.

that affective appraisal occurs without information processing of any kind but rather that it occurs independently of those information processes that are involved in belief-fixation. The cognitivists hold the view that all emotions are either belief-identical or belief-dependent.

THE THALAMO-CORTICO-AMYGDALA PATHWAY

McCabe *et al.,* (1992) explored the role of sensory-specific cortical involvement in the development of conditioned responses. Rabbits were presented with two similar auditory tones, only one of which was paired with a foot shock. The rabbits, in order to properly respond, needed to discriminate between the tones. As expected, controls rapidly learned to distinguish between the tones, enacting a conditioned response only to the tone paired with the shock (CS+). Rabbits however with lesions of the auditory cortex were found to respond to both tones as if either was indicative of foot shock.

In addition to the direct thalamo-amygdala pathway, a second pathway exists, LeDoux explains, in which crude representations of stimuli are relayed *via* the thalamus to sensory-specific cortical areas for further (in this case auditory) processing. After further sensory-specific perceptual processing (*i.e.,* refinement of the crude auditory representation), these representations are relayed (back) to the amygdala for (re)appraisal. Importantly, however, much of the "processed" information relayed to the amygdala from the sensory-specific cortical areas is still decidedly non-cognitive (in Lazarus' sense) for the kind of processing undertaken is limited to modality-specific perceptual processes (*e.g.,* visual, auditory or tactile "image" refinement). And so, while this "high road" pathway, we will see, serves to relay cognitively generated representations (*i.e.,* quasi-perceptual images that are the result of "thinking") to the amygdala for appraisal, much of what is shunted along this route are, given LeDoux's discussion, "cleaned up" image-representations that are the result of decidedly perceptual and thus modularly explicable sensory specific operations.

THE AMYGDALA AS AN AUTOMATED AFFECTIVE APPRAISAL MODULE:

Ablation, lesioning or damage of the amygdala has long been noted to "result in selective impairments in emotional perception and appreciation, as well as emotional expression in humans and animals." (Eichenbaum, 2002 p. 280) Drawing upon a number of earlier findings (Kluver-Bucy, 1937, 1939) LeDoux suggests that the amygdala functions "like the hub of a wheel" (LeDoux, 1998 p. 168) since it receives both crude (*i.e.,* minimally processed)[51] inputs from areas of the thalamus, further perceptually processed representations from sensory-specific cortical areas, as well as inputs from the hippocampus. Through these pathways, LeDoux argues, "the amygdala is able to process the emotional significance of individual stimuli as well as complex situations. The amygdala is … involved in the appraisal of emotional meaning. It is where trigger stimuli do their triggering." (LeDoux, 1998 pp. 168-9)

The amygdala-realized affective appraisal mechanism, LeDoux argues, (though not in such terms) is modularly realized. This follows for two principal reasons. First, the amygdala alone mediates and maintains the conditioned learning of emotionally significant stimuli. The lesioning studies clearly suggest that ablation of the amygdala results in an inability of subjects to respond emotionally to previously learned stimuli and to learn to respond emotionally to new stimuli. Such patients are, in effect, emotional/affective amnesiacs. Second, it is known that emotional memory and declarative/episodic memory are double-dissociable. Lesioning studies by LeDoux, Damasio & Bechara among a number of others establish that emotional/affective responses to stimuli are retained even when subjects are incapable of remembering any

[51] For example, with respect to the visual case, the thalamus receives input directly from the optic nerve. And so, while the information relayed has been "processed," the processing is limited to that undertaken by the retina, which all agree is minimal.

of the declarative facts about the episode/event due to damage to the hippocampus. There are, LeDoux explains,

> Two different memory systems … one involved in forming memories of experiences and making these memories available for conscious recollection at some time later, and another operating outside of consciousness and controlling behavior without explicit awareness of the past learning … this system forms implicit or nondeclarative memories about dangerous or otherwise threatening situations … and are created through the mechanisms of conditioning. (LeDoux, 1996 p. 181)

Tranel (1993), undertaking a formalized version of Claparede's (1911) anecdotal findings, provides evidence for the dissociability of affective/emotional and declarative memory. In a set of controlled situations, amnesic patients interacted with three researcher-cohorts: a "good guy," who was invariably pleasant and accommodating of the subjects' requests), a "neutral guy," who engaged the subjects in tasks that were neither pleasant nor unpleasant), and a "bad guy," who was brusque, inhospitable, unaccommodating of the subjects' requests and who engaged subjects in a highly tedious and purposefully irksome task. The subjects repeatedly interacted with the three cohorts in random order but always for a controlled amount of time over the course of five days. After five days the subjects were asked to participate in two tasks.

In the first, subjects were presented with a set of four photographs of which one was one of the "guys" in the experiment. Subjects were then asked questions such as "whom would you go to if you needed help?" and "who would you think is your friend?" While the subjects had no recollection whatsoever of ever having met any of those depicted, when the "good guy" was present in the photo array, subjects chose him over 80% of the time. The "neutral guy" was chosen at chance levels and the "bad guy" was nearly never chosen.

In the second task, the subject was asked to look at an array of three photos of faces and tell the researchers what they knew about them. As expected, nothing came to mind – for they had no declarative memories of ever having met any of the confederates. However, when asked who among those depicted was a friend, the "good guy" was consistently chosen. Anecdotally, Damasio recounts that one patient David, "even upon completion of the experiment … was found to hesitate and flinch when he came upon the 'bad guy' in the hall and he was unable to explain his behavior."(Damasio, 1999 p. 46)

Bechara, Tranel, Damasio, Adolphs, Rockland & Damasio (1995) provide further evidence for both the double-dissociability of declarative and affective memory and for the localization of the automated affective appraisal mechanism in the amygdala. In this study, human patients with bilateral damage to the (1) the amygdala, (2) hippocampus and (3) both regions were habituated to a set of colored slides and pure tones, some of which would serve as conditioned stimuli (CS+) and some of which would not (CS-).[52] During the conditioning phase, CS (slides or tones) were presented in random order to the subjects. Following the presentation of a CS+ stimuli, an unconditioned stimuli (US) – a loud and obnoxious horn – was sounded.

Control subjects, as expected, exhibited rapid and robust conditioning to the CS. Patients, however, with bilateral amygdala damage failed to develop any conditioned response,[53] while those with selective hippocampus damage exhibited normal conditioning. Patients with damage to both regions were also found to be incapable of developing a conditioned response.

Following the conditioning task, subjects were asked a set of questions about the task (*e.g.*, how many colors, name the colors, how many colors were followed by a horn,

[52] Habituation is presumably necessary to remove the influence of the priming affect noted by Zajonc (1980).

[53] Both behaviorally and by *SCR* measurement.

name those colors) which aimed at assessing declarative memory performance. Both controls and amygdala damaged subjects were able to correctly respond to the questions posed while those with pure and combined hippocampal damage were incapable of so doing.

Bilateral damage to the amygdala in humans, then, completely prevented the acquisition of a conditioned emotional response but had no effect on the ability of patients to recall factual information about the trial. Bilateral hippocampal damage, in contrast, prevented the acquisition of factual information about the stimuli and pairings, but had no effect on the ability of patients to acquire conditioned emotional responses. Combined damage resulted in the inability of patients to acquire conditioned responses and the inability to acquire and recall factual information about the task. (Bechara *et.al.*, 1995 p.1117)

These findings provide clear support for a "double dissociation between emotional and declarative learning in humans" and thus for the existence of dissociable and localizable systems that mediate affective and cognitive appraisals. (Bechara *et.al.*, 1995 p.1118) We have then from these double-dissociability findings, evidence for thinking that the amygdala-realized mechanism of affective appraisal maintains and relies upon a proprietary database of emotional/affective memory in the course of its processing. Put another way, the automated affective/emotional appraisal mechanisms operations rely upon information that is encapsulated with respect to information available to other systems. That is, at least some of the information available to the amygdala-realized appraisal system is not available to other systems and at least some information that is available to other systems is not available to the emotional appraisal process. As automated, obligatory, rapid and encapsulated, the operations of the affective appraisal system should be, as plausibly modular in design, amenable to modeling.

It is once again, however, worth stressing that Bechara & Damasio and LeDoux's use of the terms "appraisal" and "evaluation" in explaining the role of the amygdala (while potentially confusing in light of earlier discussion) should not be taken in the cognitive or "rational" sense of Lazarus and the cognitivists. Rather, in keeping with Ekman's and Zajonc's hypotheses, these theorists maintain that the amygdala functions as a purely automated and autonomous processing system – an automated affective appraisal module. And so, while the input accepted may be crude thalamic relays, refined perceptual representations that are the result of sensory-specific cortical processing, or retrieved image-memories from the hippocampus, the appraisals undertaken by the amygdala itself are automated and autonomous - requiring neither cortical involvement nor "thinking" (*i.e.,* higher-level cognition) in the way claimed by the cognitivist approach.

This is not to say, however, that representations generated by "higher-level" cortical processes cannot be accepted as input to the amygdala. While LeDoux focuses on the operations by which perceptual-process generated representations are appraised, he acknowledges that the amygdala – as a rather general purpose affective appraisal mechanism – takes as input both perceptual images and cognitive images (*i.e.,* quasi-perceptual sensory image reconstructions). Put another way, while the input to, and thus the content of, each appraisal may come from a number of sources (*e.g.,* crude perceptual images from the thalamus or refined ones from modality specific cortical areas), the appraisal operation itself is decidedly modular and non-cognitive in Zajonc's sense. While the process by which cognitively generated representations (*e.g.,* beliefs or suppositions) come to be appraised is not the focus of LeDoux's account, we will see, in Damasio and Bechara's model (to be discussed in the next chapter), an account of how this might be undertaken and realized.

While the amygdala-realized appraisal device is massively connected to numerous regions of the brain[54] and, as such, exerts significant, though regrettably incompletely understood, influences on a number of systems, the output of this appraisal device is, by all accounts, "shallow." (*c.f.,* Fodor, 1983 p. 86) Since the amygdala-realized appraisal mechanism relies upon a clearly dissociable (and thus proprietary) database of affective memory, the appraisal operations undertaken are "encapsulated" with respect to information available to other systems. Furthermore, since this structure is anatomically distinct, undertakes functionally dissociable activities, is automated, obligatory in operation and the output is "shallow" we have, so it would appear, reason for thinking it to satisfy the relevant criteria of a modularly realized system. If so, then at least that aspect of emotion (*i.e.,* that portion of what the emotions are) that is explicable in terms of the operations of an automated and autonomous appraisal system should be amenable to modeling in computationally feasible terms. If the operations of this automated and modularly realized appraisal device informs and influences practical reason and decision-making in the right kinds of ways, which remains to be shown, then consideration of emotion might be of some help

[54] Amorapanth, LeDoux, & Nader, 2000; Naher, Majidishad, Amorapanth, LeDoux, 2001; Pitkanen, Savander & LeDoux, 1997; Pitkanen, Stefanacci, Farb, Go, LeDoux & Amaral, 1995; Adolphs 1999; Adolphs, Denburg & Tranel, 2000; Panksepp, 1998.

Briefly, the central nucleus of the amygdala sends output projections to brain stem, basal forebrain, locus coruleus, dorsal motor nucleus of the vagus nerve, trigeminal and facial motor nuclei and lateral hypothalamus while the lateral and basal nuclei (of the amygdala) send afferent projections to the ventral striatum, cingulate cortex and orbitofrontal cortex. Projections are also found to regions implicated in neuro-peptide and neurotransmitter manufacture and release, which in turn exert profound (*i.e.,* global) influence on manifold brain structures and processes.

in explaining how we contend with some of the frame problems that arise in those domains.

ON THE LIKELY INCOMPLETENESS OF THE "BASIC EMOTION" AUTOMATED APPRAISAL ACCOUNT

While the dual-systems model sketched by automated appraisal theory is supported (*i.e.,* we have reason for thinking that at least some of what the emotions are are the result of automated and autonomous "non-cognitive" appraisal systems), the automated appraisal approach appears to be, given the current evidence, inadequate as a complete response to the question of "what the emotions are." Since there is no empirical evidence supporting the strong claim that *all* emotions are "basic" (*e.g.,* there is no evidence supporting the claim that emotions such as *love*, *Schadenfreude*, *ennui* and the like are undertaken by discrete emotion-programs) it would appear that, at the very least, such a claim is premature. Rather, in keeping with the general intuition underlying the cognitivist account, prima facie it does appear that at least some emotions are the result of "cognitive" appraisal processes. That is, it does seem that at least some emotions might be belief-dependent (*i.e.,* require "thinking"). If so, that is, if *any* emotions are belief-dependent, then automated appraisal theory provides only a partial response to the question of what the emotions are.

Extending this line of reasoning to the limit, Griffiths argues for an eliminitivist stance toward "emotion" quite generally, suggesting that our vernacular use of the term actually refers to (at least) two entirely distinct activities. Some of what we mean by "emotion" is explicable in terms of the operations of automated affective appraisal mechanisms. Some, however, the belief-dependent emotions (*i.e.,* the "higher cognitive" emotions) are not likely candidates to be explained in terms of the operations of automated and autonomous appraisal modules.

ON THE RELEVANCE OF THE DUAL-SYSTEM MODEL PROPOSED BY AUTOMATED APPRAISAL THEORY TO THE FRAME PROBLEM

If at least some emotional appraisals are undertaken by automated and modularly realized processes that operate independently of, prior to, and in parallel with cognition, then it is reasonable to think that emotion might inform decision-making and practical reasoning (*i.e.,* cognition) in ways that are relevant to contending with some of the frame problems that arise in these domains. Since the automated and autonomous appraisals are rapid and appear to be undertaken by operations that are modular in character, they should not be themselves prone to frame problems. If so, then the emotional appraisal process should be amenable to modeling. Since all agree that emotion influences cognition,[55] if it does so in the *right way* (that is, if it helps to expedite and increase the accuracy of practical reason and decision-making) then we would have reason to think that bringing emotion (at least that aspect of what emotion is that is explicable in terms of the operations of an automated appraisal process) to bear might help explain how we contend with some of the frame problems that arise in these domains.

CONCLUSION

The aim of this chapter has been modest. I have not set out to answer the question of what the emotions are. However, in setting out the propositional attitude/cognitivist account, I have suggested that we have reason for thinking this approach to be empirically problematic and methodologically questionable. The

[55] Even propositional attitude/cognitivist theorists would agree to this, for there seems little contentious in the claim that what we believe influences what else we come to believe and what we do.

internal difficulties of this approach when taken in conjunction with the evidence for the existence of discrete and dissociable automated appraisal operations, suggests that the propositional attitude/cognitivist model is a problematic one as a *complete* account of what the emotions are. And so, while it is possible (and quite likely) that some emotions might be belief-dependent, we have little reason for thinking them *all* to be as the propositional attitude/cognitivist approach maintains.

Rather, it would appear that some of what the emotions are are undertaken by automated affective appraisal modules operating independently of the operations of cognition (*e.g.*, the "basic" emotions or affect programs). However, as discussed earlier, whatever the complete answer to the question of what the emotions are ultimately turns out to be, if any of these happen to be belief-dependent, that is if any require prior "thinking," then the automated appraisal approach would provide an incomplete response to the question of what the emotions are as well. Since the evidence does not support the strong claim that *all* emotions are "basic" and at least prima facie it does appear likely that some emotions might be belief-dependent, we have reason for thinking this account as a *complete* response to the question of what the emotion are to be rather incomplete as well.[56]

Regardless of whether or not *all* emotions are reducible to the operations of dedicated appraisal modules, it is apparent that *at least some* of what the emotions are is explicable in these terms. Since at least some emotional appraisals are undertaken by

[56] There is absolutely no agreement on what a "higher cognitive" emotion is other than being something that is vaguely "emotional," something that likely involves the fixation of a belief (and thus something that likely involves activity of the neocortex) and something that has not (yet) been shown to be undertaken by an automated appraisal process. Very loosely, the set of higher cognitive emotions is often taken to include emotions such as love, guilt, shame, embarrassment, pride, envy and jealously, though even many of these (shame and pride in particular) are posited as "basic" by some theorists.

automated, autonomous and modular processes, and emotion, all agree, informs and influences cognition, (*e.g.*, the affective priming and anatomical findings outlined) it would not be unreasonable to think that emotion might be a plausible candidate for further consideration and investigation. Put another way, if the propositional attitude/cognitivist approach is correct then emotion, as always belief-dependent, would not be, as a matter of principle, a viable candidate.[57] If, however, the dual-systems model proposed by automated appraisal theory is correct (even if only in part) and non-cognitive and cognitive processes operate independently and in parallel with the former informing and influencing the latter, then, at least in principle, emotion could be a candidate. Just how viable a candidate it is will depend upon whether it informs and influences the operations of practical reason and decision-making in the *right way* - that is, whether it helps to expedite and increase the accuracy of these operations. The last chapter will focus on this question.

[57] As discussed earlier, this would amount to a circular and/or a non-starter proposal.

CHAPTER 5: EMOTION AND SOME FRAME PROBLEM INSTANCES

The previous chapter considered a number of prima facie objections to the proposal that emotion might be brought to bear on some instances of the frame problem. I argued that, with respect to these, we have little reason for thinking that – in advance of our considering specific proposals – emotion should be irrelevant to the frame problem. Rather, since a *dual-systems* model[58] in which the activities of automated emotional appraisal and cognition are undertaken by functionally distinct processes, is supported, we have reason for thinking that bringing emotion to bear on the problem should be, at least *in principle*, a plausible approach. However, it is also clear that in order for emotion to be relevant to the frame problems that arise in practical reason and decision-making, it must be shown to interact with, inform and/or guide these operations in the right kind of way. If so, then consideration of emotion might shed some light on the question of how we might contend with some of the problems that arise in these domains. Specifically, since automated emotional appraisals are rapid and autonomous, and their operations are modularly characterizable (and thus model-able), if emotion can be shown to inform and influence practical reason and decision-making in ways relevant to helping us contend with the dual horns of the dilemma posed by the frame problem (*i.e.,* speed and reliable correctness), then we would have reason for thinking that bringing emotion to bear might hold some promise for our understanding and eventually modeling these activities.

The automated operations of emotional appraisal (*i.e.,* the "basic" emotions or affect programs) and those of cognition are, we have every reason to think, mediated by distinct processes. That the two interact is, however, also clear. Specifically, that what

[58] *C.f.* Evans (2003) for a general introduction to and discussion of dual-systems models

one comes to believe may result in an emotional response establishes one direction of influence (*e.g.,* my coming to think that I shall be attacked can result in my having an emotional response). That the automated appraisal device influences and informs cognitive processes is also well established (*e.g.,* the affective priming findings). The question, of course, remains as to whether emotion occupies the proper relationship to the operations of practical reason and decision-making.

In what follows I will examine how far consideration of emotion might go in helping us to understand how we might contend with some of the frame problems that arise in practical reason and decision-making. Ultimately, I will suggest that bringing emotion to bear might help us in understanding how we contend with a number of instances of the problems. Specifically, I will suggest that emotion might meaningfully be brought to bear in us helping to understand how we might contend with the attentional-direction, problem-sequencing, meta-planning/ends selection, and Hamlet's problem instances of the problem.

To situate discussion, I will begin by setting out a puzzle of practical reason and decision-making noted by Damasio & Bechara. Following this, I will outline their *somatic marker* model which provides an account of how emotion informs and influences the operations of practical reason and decision-making. Following this, I will consider how emotion might be brought to bear in helping us contend with some instances of the frame problem that arise in practical reason and decision-making.

A PUZZLE:

Curiously, damage to three brain structures – the amygdala, the primary somatosensory region of the parietal lobe (area SM1) and the ventromedial/orbitofrontal cortex (area VMF) -results in patients who are both "emotionally flat," "devoid of emotion" or otherwise "affectless" or "disaffected" and who exhibit profound deficits in practical reason and decision-making.

Given discussion in the previous chapter, that patients with damage to the amygdala are emotionally impaired (both in emotional learning and appraisal/response) should be of no great surprise since we have reason for thinking that much of what the emotions are is undertaken by automated appraisal processes realized in this structure. That damage to the amygdala and thus the automated learning and appraisal operations realized there result in deficits in practical reason and decision-making is of interest. That damage to area SM1 should have a role in emotional "feeling" is not wholly surprising given its function. However that it should play a role in practical reason and decision-making is unexpected. Finally, that damage to area VMF should result in deficits in practical reason and decision-making is not of particular surprise given that the frontal lobes quite generally (of which area VMF is part) are often considered to be the "seat of reason" where "executive" cognitive function resides. What is of interest is that in addition to exhibiting deficits in practical reason and decision-making, VMF-damaged patients are emotionally flat, as well.

Before setting out Bechara & Damasio's account, I will begin by discussing briefly the nature of the deficits exhibited by these patients. In so doing, it will become apparent that damage to the amygdala and area VMF result in patients whose behavior uncannily resembles that of the robots employed by Dennett in his discussion of the frame problem.[59] Simply put, damage to these areas appears to result in patients who

[59] Very briefly, Dennett's first robot (R1) failed to retrieve the battery because it fails to consider an obvious implication of its plan (*i.e.,* it fails to satisfy a suitable normative standard of correctness – its isn't "rational" enough). Because of this, it takes the bomb along too and is destroyed. The second and third robots (R1D1 & R1D2) fail because they attempt to exhaustively

are stricken with/by a practical variant of the frame problem.[60] That damage to these same areas results in concomitant deficit in emotion as well, is of interest, for it suggests that emotion might play a role in helping us to contend with some instances of the problem that arise in practical reason and decision-making.

There are two interesting ways that one can fail in contending with the frame problem (in any domain). One can expeditiously arrive at incorrect conclusions (*i.e.,* be fast but wrong) or one can (attempt to) arrive at conclusions "rationally" and thus non-expeditiously (*i.e.,* be right but slow). Interestingly, damage to area VMF results in patients whose practical decision-making vacillates (in the same subject) between these two extremes – either they choose quickly and often incorrectly or they are unable to choose at all, paralyzed by their own attempts at the exhaustive consideration of plans, options and implications.

With respect to the claim that VMF-damaged patients are, at times, paralyzed by their attempts to engage in the exhaustive consideration of options, plans, implications and outcomes (in much the manner in which I suggested an overly rational system might), Damasio provides the following account.

> I was discussing with a [VMF-damaged] patient when his next visit to the laboratory should take place. I suggested two alternative dates, both in the coming month and just a few days apart from each other. The patient pulled out his appointment book and began consulting the calendar. The behavior that ensued … was remarkable. For the better part of a half-hour, the patient enumerated reasons for and against each of the two dates: previous engagements, proximity to other engagements, possible meteorological conditions, virtually anything that one could reasonably think about

consider every implication of their plan. Since exhaustive consideration is an intractable task and the bomb is on a timer, these robots too are destroyed because ultimately their reasonings are overly "rational."

[60] While I will not directly consider the particular deficits exhibited by agnosognosics (SM1-damaged) here, I will return to this in a later section. Briefly, however, while damage to area SM1 does result in profound deficits in practical reasoning and decision-making as well as affectlessness, such patients also exhibit rather pronounced cognitive deficits particularly in judgment and belief-formation.

> concerning a simple date. Just as calmly as he had driven over the ice, and recounted that episode, he was now walking us through a tiresome cost-benefit analysis, an endless outlining and fruitless comparison of options and possible consequences. It took enormous discipline to listen to all of this … when we finally did tell him … that he should come on the second of the alternative dates … he simply said: 'That's fine.' (Damasio, 1994 p. 193)

Furthermore, with respect to the frame problem-like nature of the behavior of VMF-damaged patients, Damasio provides the following discussion of a particular patient under his care.

> He needed prompting to get started in the morning and prepared to go to work. Once at work he was unable to manage his time properly; he could not be trusted with a schedule. When the job called for interrupting an activity and turning to another, he might persist nonetheless, seemingly losing sight of his main goal. Or he might interrupt the activity he has engaged, to turn to something he found more captivating at that particular moment. Imagine a task involving reading and classifying documents of a given client. Elliott would read and fully understand the significance of the material, and he certainly knew how to sort out the documents according to similarity or disparity of their content. The problem was that he was likely, all of a sudden, to turn from the sorting task he had initiated to reading one of those papers, carefully and intelligently, and to spend an entire day doing so. Or he might spend a whole afternoon deliberating on which principle of categorization should be applied. Should it be date, size of document, pertinence to case, or another? The flow of work was stopped. One might say that the particular step of the task at which Elliot balked was actually being carried out *too well*, and at the expense of the overall purpose. One might say that Elliot had become irrational concerning the larger frame of behavior, which pertained to his main priority, while within the smaller frames of behavior, which pertained to subsidiary tasks, his actions were unnecessarily detailed. (Damasio, 1994 p. 36)

That the patient exhibits both of the ways that one can fail to adequately contend with the frame problem is apparent. At times, he decides rapidly, impulsively even, failing to consider the implications of his choices. At other times, he appears to be cognitively paralyzed by attempts at exhaustive consideration. That the patient should enter into a Rylean type regress with respect to which principle of categorization should be employed is particularly instructive in this regard. With respect to this VMF patient's real-world deficits in practical decision-making, Damasio continues,

> [A]fter repeated advice and admonitions from colleagues and superiors went unheeded, Elliot's job was terminated. Other jobs – and other dismissals – were to follow. No longer tied to regular employment, Elliot charged ahead with new pastimes and business ventures. He developed a collecting habit – not a bad thing in itself, but less than practical when the collected objects were junk. The new business ventures ranged from home building to investment management. In one enterprise, he teamed up with a disreputable character. Several warnings from friends were of no avail, and the scheme ended in bankruptcy. All of his saving had been invested in the ill-fated enterprise and all were lost. It was puzzling to see a man with Elliot's background make such flawed business and financial decisions.
>
> His wife, children and family could not understand why a knowledgeable person who was properly forewarned could act so foolishly … There was a first divorce. Then a brief marriage to a woman of whom neither family nor friends approved. Then another divorce. (Damasio 1994 p. 37)

And so, damage to area VMF results in both practical paralysis *and* expeditious but reliably poor decision-making, that is, in both of the ways that one can interestingly fail to contend with the frame problem.

With damage to the frontal lobes, one expects a degradation of cognitive function. (Luria 1980; Roberts *et.al.,* 1998, Freeman, 1957) One might naturally object that perhaps Elliot's and other VMF-damaged patients' practical paralysis and impulsivity are due to a rather general – and purely cognitive – inability to reason effectively. That is, before suggesting that emotion has anything whatsoever to do with the rather unique deficits exhibited by these patients, we need reason for thinking that these are not due to simple defects in attention, memory, or reasoning.

Of relevance to this concern are the findings that, unlike most frontal-lobe damaged patients, those with damage confined to the VMF/orbitofrontal regions consistently score well within the normal range on the standard battery of cognitive tests aimed at assessing abstract reasoning and decision-making. (Tranel *et. al.,* 1993) Specifically, in addition to scoring well within the normal range on declarative memory, attentional, spatial, and IQ tests, VMF-damaged patients (quite unlike those with damage to other regions of the frontal lobe) are: (i) As adept as normal subjects at generating options for action and alternatives in response to hypothetical situations; (ii) Found to outperform normal subjects in their abilities to generate the likely consequences of actions in hypothetical situations (*i.e.,* they provide more – in number – consequences); (iii) As adept at predicting the likely outcome of hypothetical social interactions as normal subjects and (iv) As adept as normal subjects at undertaking means-ends reasoning in hypothetical social tasks. (Damasio, 1994 pp. 46-50; Tranel *et.al.*, 1993)

Such findings are of interest, for they suggest that VMF-damaged patients are quite capable of projecting consequences, generating plans and engaging in effective means-end reasoning on hypothetical tasks – at least as well as normal subjects. However, that VMF-damaged patients are capable of undertaking these tasks, and thus that their particular deficits are not due to a rather general defect in reasoning, only adds to the mystery. For example,

> At the end of one session, after [the VMF-damaged patient] had produced an abundant quantity of options for action, all of which were valid and implementable, [he] smiled, apparently satisfied with his rich imagination, but added: 'And after all this, I still wouldn't know what to do!' (Damasio, 1994 p.49)

Two points may be drawn from this. First, the particular deficits exhibited by VMF patients cannot be attributed to standard deficits in reasoning, attention, working memory or memory. Second, that these deficits manifest when subjects are required, not merely to think about hypothetical options for action, but to act/choose is instructive, for it suggests that the particular deficits exhibited by these patients manifest late in the deliberative process. Damasio explains,

> The results strongly suggested that we should not attribute [VMF-damaged patients'] decision-making defect to lack of social knowledge, or to deficient access to such knowledge, or to an elementary impairment of reasoning, or even less, to an elementary defect in attention or working memory concerning the processing of the factual knowledge needed to make decisions in the personal and social domains. The defect appeared to set in at the late stages of reasoning, close to or at the point at which choice making or response selection must occur. (Damasio, 1994 p. 50)

Returning then to our puzzle, damage to area VMF results in patients who appear to be stricken with a quite specific deficit and one that exhibits all of the hallmarks of (a practical variant of) the frame problem. Sometimes, VMF-damaged patients act impulsively and thus disadvantageously by failing to consider the implications of their plans and actions – behaving like Dennett's robot R1. At other times, these patients (like the robots R1D1 and R1D2) become, in effect, paralyzed by their own endless and exhaustive consideration of plans, outcomes, and implications – in much the way that I argued an overly "rational" system might. By so doing, they fail to act/decide advantageously (i.e. correctly) by failing to act/decide at all. [61]

Damage to the somatosensory region of the parietal lobe (area SM1) consistently results in profound defects in both the perception/appraisal and display of emotion. With respect to the latter, Damasio notes, for example, that "emotion and feeling are nowhere to be found" in these patients. (Damasio, 1994 p. 64) While damage to region SM1 results in a number of other symptoms, most notably partial paralysis and the condition *agnosognosia* (*i.e.,* the inability to sense and/or represent feedback from one's viscera and body - to "feel" one's bodily states), they also exhibit profound deficits in practical decision-making. (Damasio, 1994 p. 65) Specifically, Damasio continues, agnosognosics "are unable to make appropriate decisions on personal and social matters, just as is the case with [VMF-damaged] patients."(Damasio, 1994 p. 67) And so,

[61] Elliot for example, "was unable to choose effectively, or he might not choose at all, or choose badly." (Damasio, 1994 p. 50)

that damage to area SM1 results in both deficits in practical decision-making and the perception and display of emotion, suggests that emotion might play some role in helping us to contend with some of the frame problems that arise in this domain.

While the role of the amygdala as an automated emotional learning and appraisal module has been discussed, it is instructive to note how damage to this region affects subjects' "real-world" practical reason and decision-making capacities. Anecdotal evidence of the real-world deficits exhibited by amgydala-damaged patients is hard to come by since cases of pure bilateral amygdala damage in humans are exceedingly rare. Only one natural condition consistently results in selective bilateral amygdala damage, namely *Urbach-Wiethe* disease (an exceptionally rare autosomal disorder resulting in the calcification of the amygdala). Those patients available for study are described as exhibiting significant deficits in real-world practical reasoning decision-making as well as demonstrating "improprieties and irrationalities" in social behavior. (Tranel & Hyman, 1990 p. 349, p. 352 and p. 354). A particular Urbach-Wiethe patient under his care, exhibited, Damasio recounts,

> A lifelong pattern of personal and social inadequacies. There is no doubt that the range and appropriateness of her emotions are impaired and that she has little concern for the problematic situations in which she gets herself. The 'folly' of her behavior is not unlike that found in [VMF-damage patients] or patients with agnosognosia. (Damasio, 1994 p. 69; *c.f.* Nahm, Damasio, Tranel & Damasio, 1993)

One might object, noting that perhaps such patients' practical deficits are attributable to some rather general defect in reasoning. If so, the concern continues, then there would be no reason to think emotion to be particularly relevant to the problem. With respect to this, Siebert, Markowitsch & Bartel (2003) studied nine *Urbach-Wiethe* patients finding that with respect to the standard battery of attentional, working memory, spatial, verbal and visual memory and executive function tests, (Siebert *et. al.*, 2003 pp. 2628-9 and p. 2631 table 4) "patients showed cognitively little deviation from normal subjects, while they differed emotionally."(Siebert *et. al.,* 2003 p. 2627)[62] And so, the profound real-world deficits in practical reasoning and decision-making exhibited by amygdala-damage patients do not appear to be due to a rather general deficit in cognitive function or reasoning capacity.

THE IOWA GAMBLING TASK

Since VMF and amygdala-damaged patients clearly exhibited deficits in "real-world" practical reason and decision-making that are undetectable by the standard battery, Bechara and Damasio developed a paradigm "designed to simulate real-life decision in terms of uncertainty" – the *Iowa Gambling Task* – in order to better understand the puzzle posed by these patients. (Bechara, Damasio & Damasio, 2003 p. 357)

While there are a number of versions of the Iowa Gambling Task (IGT), all are quite similar and all rely on the same paradigm in which there are four decks of cards (A,B,C, and D) from which subjects are to choose. The decks are "stacked" in the following manner: Per every ten cards selected from deck A, the subject will receive reward cards yielding in total $1000 and punishment cards inflicting a loss ranging from $150 to $350. In fact, for every 10 cards selected from deck A, subjects receive $1000 but are forced to pay out $1250, resulting in a net loss of $250. For every ten cards selected from deck B, there are rewards totaling $1000 but there is one punishment card of $1250.

[62] It should be noted, however, that while Siebert *et.al.,* found no remarkable deficits in the planning functions of the nine Urbach-Wiethe patients studied, the one patient studied by Tranel & Hyman (1990) did exhibit some planning deficits as well as some marked deficits in verbal memory. Urbach-Weithe disease can result in calcification of the regions surrounding the amygdala (including aspects of the hippocampus) and at least some of the discrepancies may be attributable to this.

For every 10 cards selected from this deck as well, then, there is a net loss of $250. Per every ten cards selected from deck C there is a net gain of $500 and five punishment cards inflicting losses of $25 to $75. And so, for every ten cards selected from this deck, there is a net gain of $250. Similarly, for every ten cards selected from deck D there is a net gain $500 and one large punishment card of $250. The net gain per every ten cards selected from this deck, then, is also $250.

And so, decks A and B, and decks C and D, are equivalent with respect to net losses and net gains respectively. Decks A and C have a higher frequency but smaller punishments, while decks B and D have lower frequency but larger punishments. As Bechara *et. al.,* explain, "decks A and B are disadvantageous because they cost more in the long run. Decks C and D are advantageous because they result in an overall gain in the long run."(Bechara Damasio, Damasio & Lee, 1999 p. 5474)

In addition to monitoring the cards selected by subjects, *skin conductance responses* (*SCR*s) described as "physiological indices of an anatomically controlled change in somatic state (Bechara, Damasio & Damasio, 2000 p. 299) which provides "a measure of somatic state activation" (Bechara, Damasio, Damasio & Lee, 1999 p. 299) and which are best understood as the physiological manifestation of an automated emotional response, are recorded at different intervals during the task. Specifically, *SCR*s recorded from subjects are divided into three category-epochs. (1) *Reward SCR*s that are generated after the selection of cards resulting in an immediate monetary gain; (2) *Punishment SCR*s that are generated after the selection of cards resulting in an immediate monetary loss, and (3) *Anticipatory SCR*s which "are generated previous to turning cards from any given deck, *i.e.,* during the time period the subject ponders from which deck to choose." (Bechara, Damasio, Damasio & Lee, 1999 p. 5474)

While control subjects rapidly (within twenty or so card selections) come to select from the advantageous decks, patients with damage to either the amygdala or the VMF fail to generate any aversion to the bad decks and no preferences for the good decks, exhibiting ultimately a stable preference for the bad decks. (Bechara, Damasio, Damasio & Lee, 1999 p. 5477 figure 2; Bechara, Damasio & Damasio, 2000 p. 298 figure 2; Bechara, Damasio, Damasio & Lee, 1999 pp. 5475-6; Bechara, Damasio & Damasio, 2000 p. 297; Damasio, 1994 pp. 212-222) Put another way, damage to either area VMF or to the amygdala results in patients who consistently choose poorly.

Examination of the *SCR*s of the subjects during the selection period of the task provides further insight to the roles of these areas in practical reason and decision-making. Control subjects were found to generate significant *SCR*s *after* the selection of cards regardless of their deck of origin. However, while *SCR*s were generated to cards from all decks, controls generated higher amplitude *SCR*s in response to the selection of a card that punished the subject with monetary loss. (Bechara, Damasio, Damasio & Lee, 1999 p. 5478 figure 4). Patients with amygdala damage, as expected, failed to generate *SCR*s to any of the cards selected. (Bechara, Damasio, Damasio & Lee, 1999 p. 5479) Such findings are consistent with the results obtained by Morgan & LeDoux (1995) and Morgan *et. al.,* (1993) discussed previously with respect to the role occupied by the amygdala in emotional appraisal and learning in rats. And so, in keeping with Ekman and Zajonc's models, discussed in the previous chapter, damage to the amygdala in humans significantly interferes with the generation of emotional responses to both immediately rewarding and punishing stimuli (*i.e.,* in emotional appraisal) and renders subjects unable to acquire new learned conditioned responses. Likewise, in keeping with Morgan & LeDoux's (1995) findings, damage to the VMF alone results in no impairment in the ability of patients to generate normal (*i.e.,* control-level *SCR*s) responses to immediately rewarding and punishing stimuli. Put another way, VMF-damage patients and controls (but not amygdala-damaged patients) generated *SCR*s to immediately rewarding and punishing stimuli. Patients with damage to either the amygdala alone or to both the amygdala and area VMF "were impaired severely in the generation of either reward or punishment *SCR*s." (Bechara, Damasio, Damasio & Lee, 1999 p. 5478)

When presented with the gambling task, normal subjects come to generate *anticipatory SCR*s *before* selecting cards from the various decks. The *SCR*s of the control group, Bechara *et. al.,* note "develop over time (*i.e.,* after selecting several cards from each deck, and thus encountering several instances of reward and punishment) … [and] become more pronounced before selecting cards from the disadvantageous decks (A and B)." (Bechara, Damasio & Damasio, 1999 p. 299) Of particular significance were the findings that, unlike the robust anticipatory *SCR*s generated by controls before selecting, neither amygdala-damaged nor VMF-damaged patients generated anticipatory *SCR*s before choosing.

That amygdala-damaged patients should fail to generate *anticipatory SCR*s is unsurprising since such patients have lost the capacity to both generate emotional responses to stimuli and the capacity to come to learn new emotional "inducers" or "triggers." Simply put, because they lack the underlying machinery necessary for emotional appraisal, amygdala-damaged patients fail to generate *anticipatory SCR*s for the same reason they fail to generate reward and punishments *SCR*s.

That VMF-damaged patients decide disadvantageously (*i.e.,* "myopically) and are incapable of generating anticipatory *SCR*s is puzzling, for the amygdala and thus the capacity to generate emotional responses to immediately rewarding/punishing selections is intact in these patients. Adding to the puzzle, Bechara *et. al.,* found that normal subjects come to consistently select from the "good" decks, though they are unable for some time, if ever, to explain why they do so. Specifically, upon completion of the task, 70% of normal subjects attain what Bechara *et. al.,* term the "conceptual stage" whereby subjects are able to form correct opinions about both (i) which decks are objectively "good" and which "bad" and (ii) how they *should* choose. And so, while all normal subjects rapidly come to decide advantageously, most are eventually able to form correct opinions about how they *should* choose and why.

Amygdala-damaged patients, in contrast, both consistently decide poorly and are entirely incapable of attaining the "conceptual stage." Such patients quite literally have no opinion whatsoever (nor do they even have "hunches") about the goodness or badness of any of the decks. Put another way, amygdala-damaged patients both invariably make poor decisions on the gambling task and have absolutely no idea about how they *should* choose. That damage to the amygdala should result in the inability to attain the conceptual stage (*i.e.,* to dramatically interfere with a subject's ability to arrive at correct conclusion about how she *should* choose) is of interest, for it suggests that emotion might play a role in the operations of practical reasoning.

Of further interest, is the finding that while all VMF-damaged patients consistently chose poorly, 50% attained the "conceptual stage" – forming correct opinions about which of the decks were good/bad and how they *should* go about choosing. That is, 50% of VMF-patients fixed correct beliefs about how they *should* choose, yet all (still) consistently chose disadvantageously. This result confirms the earlier discussed (behavioral/anecdotal) evidence in which VMF-damage patients, who "knew better" (and who really do appear to know better) still invariably made decidedly poor plans and decisions in real-life situations.

Having set out this puzzle, I will turn next to outlining Bechara and Damasio's proposal – the *somatic marker hypothesis* – which provides a systematic account of the role played by emotion in practical reason and decision-making.

BECHARA AND DAMASIO'S *SOMATIC MARKER HYPOTHESIS*

Bechara and Damasio propose that emotion plays a critical role in the heuristic direction of practical reason and decision-making (*i.e.,* in "cognitive guidance" Damasio, 1994 p. 130) by helping to automatically cull from further consideration plans and response options that are (likely) to be disadvantageous, while "highlighting" that is, both drawing attention to and increasing the desirability of, those options that are (likely) to be advantageous. Specifically, by exploiting the automated processes of emotional appraisal during deliberation, emotion helps to both expedite and increase the accuracy of the operations of practical reason and decision-making. While I will return to the issue of how this proposal might help us to understand how we contend with some instances of the frame problem, I will, for now, focus on outlining the somatic marker hypothesis.

Damasio explains that during deliberation but,

> *Before* you apply any kind of cost/benefit analysis to the premises and before you reason toward the solution of the problem, something quite important happens: When the bad outcome connected with a given response option comes into mind, however fleetingly, you experience an unpleasant gut feeling. (Damasio, 1994 p. 173)

Specifically, with respect to the operations of practical reason and decision-making, these "gut feelings" or somatic markers serve to, "Force attention on the negative outcome to which a given action may lead, and functions as an automated alarm signal which says: beware of danger ahead if you choose the option which leads to this outcome." (Damasio, 1994 173) Such somatic markers (*i.e.,* feeling-image signals), Damasio continues, "may lead you to reject, *immediately,* the negative course of action and thus make you choose among fewer alternatives."(Damasio, 1994 p. 174) Somatic markers, Damasio explains,

> Assist deliberation by highlighting some options … and eliminating them rapidly from subsequent consideration. You may think of it as a system for automated qualification of predictions, which acts, whether you want it or not, to evaluate the extremely diverse scenarios of the anticipated future before you. Think of it as a biasing device. (Damasio, 1994 p. 174)

By so doing, somatic markers,

> Protects you against future loss, without further ado, and then allows you *to choose from among fewer alternatives*. There is still room for cost/benefit analysis and proper deductive competence, but only *after* the automated step drastically reduces the number of options. (Damasio, 1994, p. 173)

Somatic-markers, Damasio explains, are "feelings" generated from "secondary emotions" which "have been connected, by learning, to predicted future outcomes of certain scenarios." (Damasio, 1994 p. 174) And so, during deliberation and quite automatically, "when a negative marker is juxtaposed to a particular future outcome the combination functions as an alarm bell. When a positive somatic marker is juxtaposed instead, it becomes a beacon of incentive." (Damasio, 1994 p. 174) Damasio continues,

> My idea is that somatic markers (or something like them) assist the process of sifting through such a wealth of detail – in fact, reduce the need for sifting because they provide an automated detection of the scenario components which are more likely to be relevant. (Damasio, 1994 pp. 174-5)

At base then, Bechara and Damasio's claim is a straightforward one. When abstracted from the neuro-anatomical details, the somatic marker hypothesis suggests that emotion-induced feeling is exploited *heuristically* - as a non-compensatory (*i.e.,* one-reason) cue - during deliberation. Put another way, based upon induced feeling alone, some plans or options are rejected outright and given no more consideration (*i.e.,* we stop thinking about these), while others are rendered more desirable and attentionally highlighted.

Having set out the general claim, I will next consider the proposal in greater detail. I will begin by explaining the distinction marked by Bechara and Damasio between "primary" and "secondary" emotions, which relies upon the distinction between primary and secondary inducers (of emotion).[63] Primary inducers are

> Stimuli that unconditionally, or through learning (*e.g.,* conditioning and semantic knowledge) can (perceptually or subliminally) produce states that are pleasurable or aversive. Encountering a fear object (*e.g.,* a snake), a stimulus predictive of a snake, or semantic information such as winning or losing a large sum of money are all example of primary inducers. (Bechara, Damasio & Damasio, 2003)

The primary emotions then are analogous to the basic-emotions and affect programs discussed in the previous chapter.

Secondary inducers are, Bechara explains,

> Generated by the recall of a personal or hypothetical emotional event or perceiving a primary inducer that generates "thoughts" and "memories" about the inducer, all of which, when brought to memory, elicit a somatic state. The episodic memory of encountering a snake, losing a large sum of money, imagining the gain of large sum of money, or hearing or looking at primary inducers that bring to memory "thoughts" pertaining to an emotional event are all examples of secondary inducers. (Bechara, Damasio & Damasio, 2003)

[63] Perhaps relating primary and secondary emotions to earlier discussion of the "primary emotions" will help to alleviate potential confusion. Both primary and secondary emotional responses, as is clear from Damasio's discussion, are undertaken by automated mechanisms of affective appraisal (*i.e.,* the mechanisms of the "basic emotions.") The sole difference between primary and secondary emotion lies in the "reality" of the inducer. Primary emotions are triggered by immediate and "real" situations, while secondary emotions are triggered by *suppositionally* entertained "thoughts" (*i.e.,* episodic memories of past situations or hypothetically entertained situations.) Put simply, secondary emotions should *not* be taken to be somehow unique with respect to the basic operations of emotional appraisal nor should they be taken in contrast with the "basic" emotions. Both primary and secondary responses are undertaken by means of the same automated mechanisms of emotional appraisal processes and are in this respect both "basic."

Bechara and Damasio argue, drawing upon LeDoux, that the amygdala is the principle region implicated in the generation of "primary emotions," that is, in emotional learning and appraisal and the triggering of emotional responses to primary inducers. Area VMF is the principal substrate implicated in the appraisal, learning and generation of secondary emotions, that is, in the triggering of emotional responses to secondary inducers (*e.g.*, memories and suppositional or hypothetical "thoughts".) Having marked this distinction and explained some necessary terminology, I can now turn to outlining Bechara and Damasio's *Somatic Marker* proposal. To do so however, some discussion is needed of what "feeling" is under this account. This, in turn, requires that I present briefly the role of area SM1 in this account.

Bechara and Damasio suggest that the generation and storage of somatosensory images are the provinces of area SM1. Specifically, area SM1 generates "on-line representations … of what our body state is now" (Damasio, 1994 p. 152) by continuously monitoring and mapping all moment-to-moment bodily/somatic occurrences. It is also maintains "off-line" representations of some of these somatic states in somatosensory pattern memory. (Damasio, 1994 p. 150). Direct pathways relay information from the viscera, musculoskeletal positioning – including facial expression and bodily stance – blood vessels, blood pressure and heart rate to the somatosensory cortices, where a "snapshot" representation, body-state map or somatosensory state image of the subjects *internal milieu* or *Korpersgefuhl* is generated on-line and in real-time. (Damasio, 1994 p. 144) This process, Damasio explains, "of continuous monitoring … of what your body is doing … [and the generation of] a multifarious view of the body landscape … is the essence of what I call a feeling." (Damasio, 1994 pp. 143-4)

And so, area SM1 generates both "on-line" images of the subject's somatic/visceral state at a given moment and serves as the repository of somatosensory feeling-image memory. In so doing, SM1 maintains a set of "off-line" somatosensory feeling memories both of our body states between emotions (*i.e.,* "when it is not shaken by emotion" and therefore what the subject's somatic homeostatic state tends to be like) and of our body-state when an emotional response is enacted.

Whenever an amygdala-realized appraisal occurs and an emotion is induced, manifold alterations in the somatic state of the subject come to manifest in the manner proposed by Ekman and Zajonc. The role of area SM1, then, is simply to create a map or somatosensory-image "snap-shot" of the particular physiological state induced. "Feelings" just are these body-image snapshots generated or mapped by area SM1.[64]

Having set out what feelings are and the mechanism by which they are generated, we can now consider the role proposed by Bechara and Damasio for area VMF. The principle function of area VMF, they argue, is in the construction of "fact-feeling" sets whereby perceptual and quasi-perceptual images (*i.e.,* recalled or hypothetical imaged content) come to be juxtaposed with and indexed by somatosensory feeling. Put another way, area VMF serves to "mark" or "tag" images (*i.e.,* both perceptual images and quasi-perceptual images of the content of "thoughts") with the feeling that is or comes to be induced by them.

With respect to the feeling-tagging of perceptual images, let us suppose that one has an automated emotional response to some stimuli (a snake, for example) which results in a primary emotional response and a set of manifold physiological changes to

[64] Note too that *qualia* need not play a role in Damasio & Bechara's account of what feelings are. Rather, feelings are simply somatosensory cortex generated imagetic representations of somatic states.

be realized. While area SM1 is busy constructing a somatic map of these induced physiological changes (*i.e.,* a "feeling" of the emotion induced), perceptual processes are going about generating a detailed perceptual image of the inducing scene. These two images, Bechara and Damasio explain, are relayed to a "convergence zone" (*i.e.,* a dedicated working memory buffer) in area VMF whereby the two images are juxtaposed and a fact-feeling set is created. The role, then, of area VMF is to "tag" incoming perceptual images (the perceptual representation of the snake scene, in this example) with the particular "feeling" induced, creating a fact/feeling pair. This fact/feeling set is then stored as a "dispositional representation," in memory, that is, in a form that can be, when retrieved, reconstructed by or reconstituted in sensory-specific cortical regions.

With respect to the tagging of secondary inducers, Bechara and Damasio offer the following account. Drawing upon Kosslyn & Koenig's model (later significantly elaborated by Milner & Goodale, 2006), Bechara and Damasio suggest that the content of "thoughts" are represented imagetically by means of the modular operations of sensory specific cortical regions. Put another way, they suggest that the content of "thoughts" are, by means of the modular operations of sensory specific cortical perceptual processes, reconstructed or reconstituted as quasi-perceptual images (*i.e.,* as visual, auditory, olfactory and so on "pictures"). And so, for example, before processing (*i.e.,* before we can "think" about them) episodic/declarative memories of a particular event (the memory of seeing a snake, for instance) are relayed to the visual cortex where a quasi-perceptual image of the content of this memory is reconstructed using the same processes that are employed directly in perception, in effect, in reverse. (Milner & Goodale, 2006)

Regarding the tagging of quasi-perceptual images, (*i.e.,* multimodal sensory cortex constructed images of the content of recalled or hypothetical "thoughts"), there are two mechanisms proposed. First, *via* the "body-loop"device, quasi-perceptual images are relayed through sensory cortices (which reconstruct the content of these thoughts as visual/auditory etc., images) directly to the amygdala for emotional appraisal. If the (quasi-perceptual) image serves an inducer or trigger, an automated emotional response is generated and, *via* the workings of SM1, a feeling image is constructed. Once relayed to the convergence zone, the "thought" (or more precisely, the quasi-perceptual image of the content of the thought) comes to be "tagged," indexed or marked by the feeling (somatosensory image) induced, in the same manner as perceptual images are marked.

In cases in which a quasi-perceptual image has been previously feeling tagged and stored in memory (*i.e.,* in cases in which a fact/feeling set exists), the feeling-memory – but not the emotional response itself - is reconstructed in area SM1 and relayed to the convergence zone whereby the full fact-feeling set is reconstituted. This "as-if" loop mechanism, by relying upon area SM1-mediate feeling memory, allows for an emotional appraisal of secondary inducers (*i.e.,* thoughts) to be undertaking without the need for direct appraisal by the amygdala. By bypassing the body-loop appraisal device – and thus the necessarily serial and time-consuming process of generating an emotional response and mapping the physiological alterations that ensue – secondary inducers (*i.e.,* episodic memories or suppositional "thoughts") may be quite rapidly appraised with respect to the feeling that has been juxtaposed with the image.

One may object to the model offered by questioning what reason we have for thinking that the fact/feeling sets generated should be preferentially encoded into memory in the manner that Damasio & Bechara suggest. Since, as we will see, the somatic marker account relies heavily upon these fact-feeling sets being retained and maintained in memory, we need some reason, the concern continues, to think that these sets should be preferentially remembered.

That emotion influences memory is well established.[65] Quite generally, and with respect to this concern, these findings confirm that emotionally significant material is preferentially remembered. With regard to the mechanism by which emotion influences the preferential encoding of material to memory, McGaugh & Cahill offer the following explanation.

The presence of the peripheral hormone epinephrine (adrenaline) is known to significantly enhance memory.[66] McGaugh & Cahill argue that since activation of the amygdala (*i.e.,* emotional appraisals) invariably result in hypothalamic and brainstem activation, which in turn results in the release of epinephrine, emotionally arousing material (*i.e.,* episodic/declarative "facts") comes to be preferentially encoded to memory. That amygdala activation is responsible for inducing the epinephrine cascade resulting in the preferential encoding of "factual" material to memory is established by Canli, Zhao, Brewer, Gabrieli & Cahill (2000). And so, LeDoux, drawing upon McGaugh and Cahill's findings explains,

[65] Bower, 1981; Bower, Gilligan & Monteiro, 1981; Gilligan 1982; Matt, Vasquez & Campbell, 1992; Watkins, Mathews, Williamson & Fuller, 1992; Ellis, Thomas & Rodriguez, 1984; Burke & Mathews, 1992; Mayer &Volanth, 1985; Mayer, Gaschke, Braverman & Evans, 1992; Mayer, McCormick & Strong, 1995; Bower, 1981; Bower, Monteiro & Gilligan 1978; Teasdale & Fogarty, 1979; Snyder & White, 1982; Bower & Mayer, 1985; Blaney, 1986; Eich, 1995; Eich & Macaulay, 1989; Fiedler, 1990; Forgas, 1991; Forgas 1993; Forgas, 1995; Forgas & Bower, 1987; Eich & Macaulay, 2000; Niedenthal & Setterlund, 1994, Ingram, 1984; Teasdale, 1983; Teasdale & Clark, 1982; Teasdale & Russell, 1983, Phelps 1998, Siebert *et. al.,* 2003.

[66] McGaugh *et. al.,* 1993; Cahill, Prins, Weber & McGaugh, 1994; Cahill *et. al.,* 1994; Cahill & McGaugh, 1995; Cahill *et. al.,* 1996; McGaugh, Cahill, & Roozendahl, 1996; McGaugh 2004, Davis, 1992; Davis, 1994, Eichenbaum, 2000; Rolls & Treves, 1998; Rolls 2000.

> When the amygdala detects an aversive emotional situation, it turns on all sorts of bodily systems, including the autonomic nervous system. The consequence of autonomic nervous system activation of the adrenal gland is the release of epinephrine [which] interacts with systems that are also active at the time, such as the hippocampal system that is forming the explicit memory of the situation. (LeDoux, 1998 p. 207)

And so, we have good reason to think that emotionally significant ("factual") material is preferentially encoded to memory - because it is emotionally significant - in the manner assumed by Damasio & Bechara's account.

Having briefly set out the machinery of Bechara and Damasio's model, I will return to the Iowa Gambling Task in order to set out the role proposed for emotion in practical reason and decision-making.

That amygdala-damaged patients should generally choose poorly (disadvantageously) is explicable in terms of their inability to generate "primary emotions." That is, since damage to the amygdala impairs patients' abilities to both perceive the emotional significance of stimuli and learn to respond (emotionally) to new stimuli, patients are incapable of appraising the emotional significance of situations, including, in this case, the outcome of any of their selections. It is for this reason that no reward or punishment *SCR*s are generated. It would be, however, highly unlikely that when the perceptual image is relayed to the convergence zone in area VMF for fact/feeling juxtaposition that there should be no feeling present to be linked. Rather, in the absence of an emotional response, it is quite reasonable to think that perceptual images would be linked to/with the subject's prevailing feeling state – in this case a somatosensory image of bodily homeostasis (*i.e.,* a baseline body state "unshaken by emotion"). And so, it would be improper to suggest that amygdala-damaged patients' deficits lie in their inability to tag perceptual images with feelings. Rather, their problem lies in the inability to generate emotional responses. As such, each selection scene is tagged with same rather mundane feeling-image – physiological homeostasis. In this respect then each selection (and thus each fact-feeling set generated) is, for these patients, entirely devoid of emotional and thus motivational significance.[67]

With respect to the question of why amygdala-damaged patients fail to choose advantageously, Bechara and Damasio offer the following explanation. Since amygdala-damaged patients are incapable of emotionally responding to objectively rewarding/punishing card selections, the fact-feeling sets generated in area VMF lack an informative feeling component – in effect, marking all "facts" with the same commonplace feeling image of homeostasis. Understood in this way, each selection and outcome is no more or less rewarding or punishing than any other. During deliberation, suppositionally entertained quasi-perceptual images of the scene of selecting from each deck are generated and subjected to a round of emotional appraisal. Since each option is, upon appraisal by the body or as-if loop devices, found to be equivalent with respect to the information provided by emotion-induced feeling (all are neutrally marked), none is taken to result in a state of affairs that is any better or worse than any other. And so, Damasio and Bechara continue, the outcome of each option is, as emotionally neutral, motivationally equivalent.

Underlying the somatic marker account is the claim that feeling is exploited heuristically in practical reason and decision-making to cull negatively marked plans and options from further consideration while highlighting for further consideration and

[67] Clearly, there should be other factors influencing the body-maps and thus the feelings created. For example, something as simple as a decrease in blood sugar would result in a change in the subject's somatic state and thus in the kind of feeling generated and juxtaposed. And so, while the "factual" aspect of amydala-damaged patients sets are likely tagged with the subjects prevailing body-state, I intend here that such states are only *emotionally* homeostatic (*i.e.,* with respect to those changes in physiology induced by emotion).

rendering more desirable those positively marked. Since amygdala-damaged patients cannot generate emotional responses, the feeling component of their fact/feeling sets is limited to the feeling-image of homeostasis. And so, since all fact/feeling sets are feeling equivalent, the feeling-*cue* cannot be exploited by the stopping and decision heuristic proposed by Damasio and Bechara's account. Simply put, the feeling-cue based heuristic cannot function because, since each feeling is neutral, the cue is uninformative.

Returning to the IGT then, during deliberation, quasi-perceptual images of each option-scenario are generated and suppositionally entertained. The *anticipated* emotional response/feeling brought to bear on each option is the same – feeling neutral. And so, it is reasonable to suggest that amygdala-damaged patients exhibit no *anticipatory SCR*s, because the suppositionally entertained images of any response option is projected to result in a neutral feeling – the same as any other. As Damasio explains, "Failure to evoke somatic states after winning or losing money would preclude the reconstitution of such somatic states when deliberating a decision with future consequences."(Damasio, 1999 p. 5480) The thought, then, of choosing from any deck is appraised as neither likely to be rewarding nor punishing – just mundanely neutral. As feeling-neutral, each option tagged is, for the amygdala-damaged patient, motivationally neutral as well – no option is taken to result in a state of affairs that is any more or less desirable than any other.

That amygdala-damaged patients are entirely incapable of forming any opinion (much less a reliably correct one) about which decks are good/bad and about how they *should* choose (*i.e.,* of attaining the "conceptual stage") is of interest, for it suggests that emotion might directly inform practical reasoning.

While Bechara and Damasio offer no explicit account of why amygdala-damaged patients should be incapable of attaining the conceptual stage, the following proposal can be offered. Since amygdala-damaged patients are incapable of finding states of affairs emotionally significant (*i.e.,* rewarding or punishing) they are radically impaired in their ability to value (and thus to "care about") any particular outcome any more or less than any other. A natural hypothesis then is to suggest that during deliberation, patients are likewise incapable of assigning a value to the outcome of any of their actions on the IGT. And so, if they are incapable of valuing any state of affairs over any other, they would fail to take precautions to guard against those options that are objectively "bad" (*i.e.,* punishing.) Put somewhat differently, if these patients can value neither their current situation nor those states of affairs that will result from their actions, then there is, from their perspective, nothing to pursue and nothing to guard against. There is, in short, no problem in need of solving. And so, one plausible conjecture for why it is that amygdala-damaged patients fail to attain the conceptual stage on the IGT is that it fails to (be taken to) present a problem. Furthermore, since they experience no emotional response to any given outcome, they can assign no value (positive or negative) to them. And so, since all states of affairs are value-neutral, perhaps such patients are unable to form an opinion about how they *should* decide/choose, because from their defective perspective, there is nothing upon which to ground such judgments.

Underlying the somatic marker hypothesis is the claim that emotion-induced feeling is exploited heuristically to both focus attention upon and fix (*i.e.,* increase or decrease from neutral) the desirability of particular response options. It is generally accepted that attentional focus, by preferentially enlisting cognitive resources, results in cognitive focus (Adolph, 2000). Simply put, we preferentially "think about" those things that we attend. That the emotional landscape of amygdala-damaged patients is profoundly flat – nothing is any more emotionally significant than anything else – suggests that attentional resources would not be preferentially directed upon material that a normal subject would find emotionally significant. Furthermore, being incapable of assigning values (positive or negative) to these, such patients are also unable to fix the desirability of those options entertained. And so, it is reasonable to suggest that these patients are doubly impaired, for not only are their quite normal cognitive resources not preferentially enlisted or directed to "think" about the material, but such patients are also incapable of valuing and thus fixing the desirability of any of the options brought to bear.

That VMF-damaged patients should decide disadvantageously is explained, Bechara and Damasio continue, in terms of their inability to effectively employ the machinery of the secondary emotions. That is, since the amygdala is intact and these patients are perfectly capable of generating reward/punishment *SCR*s in response to their selections, they are quite capable of generating primary emotions in response to learned and unlearned stimuli. They are also capable of generating feelings of the emotions induced since area SM1 is intact in all patients. The particular deficit in practical reason and decision-making exhibited by VMF-patients then, is attributed to an inability to effectively construct and maintain "fact/feeling" sets, which in turn precludes feeling from being heuristically exploited during deliberation.

And so, in the case of VMF-damaged patients, emotional responses are generated to each selection. Feeling-images of the induced physiological changes are generated in area SM1. Simultaneously, images of the scenario itself are being generated by perceptual processes. It is at this point, however, that the process breaks down, for damage to the VMF precludes these two kinds of information from being linked. Therefore, while perceptual images of each “factual” selection are generated, and emotion-induced feeling images are generated of the response to each selection, there is no fact-feeling set created. There are “facts,” and there are “feelings,” but the former are never properly tagged or marked by the latter.

In order to avoid confusion, it should be noted that since VMF – damaged patients can attain the “conceptual stage,” they are seemingly capable of pairing images and feelings in *one* sense – that of recognizing the co-occurrence of particular facts and feelings. However, what they are unable to do is to pair (*i.e.,* to assign) particular feeling tags to/with particular factual images. This in turn, results in their inability to transitively fix (“by dint of juxtaposition” with feeling) the value or desirability of any particular outcome. That is, VMF-damaged patients, while capable of recognizing that some facts and feelings are co-occurent, are unable to “pair” the two such that the “body images give to other images a *quality* of goodness or badness.” (Damasio, 1994 p. 159) It is, then, this latter sense of “pairing” that is Damasio’s focus.

During deliberation, suppositionally entertained quasi-perceptual images of selecting from each deck are generated. In normal subjects these images would be emotionally appraised and the feeling induced juxtaposed. By means of this, the desirability of the perceptual or quasi-perceptual images itself is (transitively, *via* the juxtaposition with induced feeling) fixed. VMF-damage however precludes this information from being brought to bear since the quasi-perceptual images arrive for consideration untagged by feeling. Since there are no feeling-markers affixed to the options, anticipatory feeling cannot be exploited heuristically to highlight or cull options during deliberation nor to render them respectively more or less desirable. And so, like the amygdala-damaged patients, no option is any more or less emotionally and hence motivationally significant (*i.e.,* rewarding, positive, preferable, desirable or punishing, negative, aversive) than any other. However, unlike amgdala-damaged patients who tag images with the neutral feeling of emotional homeostasis, VMF-damaged patients are incapable of marking these images with *any* emotion–induced feeling at all. Options are, for them, not just feeling-neutral but entirely emotionally feeling-less. As such, each option is, with respect to its desirability, equivalently emotionally and hence motivationally neutral.

On the IGT, normal subjects generate primary emotional responses and feelings to each card selection and, *via* the operations outlined, fact-feeling sets are created. At the outset, since there is no feeling-history to exploit, during deliberation none of the suppositionally entertained scenes of each option is emotionally highlighted (positively or negatively). As such, no option is taken to be any more or less desirable than any other. That for the first twenty or so trials subjects choose randomly supports this. However, each card selection and induced primary emotion and fact/feeling set generated provides additional information. As Bechara *et. al.,* explain, area VMF functions to,

> Establish a linkage between the disposition for a certain aspect of a situation for instance, the long-term outcome for a type of response, and the disposition for the type of emotion that in past experience has been associated with the situation. (Bechara, Damasio & Damasio, 2003 p. 360)

Over time and with repeated interactions with the environment, Damasio & Bechara explain, since the role of the VMF is to "integrate effectively all of the somatic information triggered by the amygdala" (Bechara, Damasio, Damasio & Lee, 1999 p. 5480) a summed, averaged, or otherwise integrated feeling of what selecting from each decks feels like *on balance* comes to be generated and dispositionally encoded.[68] During deliberation, quasi-perceptual images of the "scene" of selecting from each deck are generated and suppositionally entertained. *Via* appraisal by the as-if loop device the retrieved "somatic state that integrates the numerous and conflicting instances of reward and punishment encountered with individual cards draws from each of the decks " is reconstituted (Bechara, 1999 p. 5480) and the *on balance* feeling is reconstructed. This *on balance* feeling is then juxtaposed with the representation of the suppositionally entertained scene. By means of this, each option entertained comes to be marked by the feeling that has *on balance* resulted. Those options that are on balance positively feeling-marked are highlighted – they command attention and are rendered more desirable. Those that are on balance negatively marked are, as undesirable, culled from further consideration. Since, Bechara explains, feelings "influence the decision to select from, or avoid, that deck." (Bechara, 1999 p. 5474) during deliberation subjects heuristically exploit the *on-balance* feeling tags associated with each options *anticipatorily,* to direct attention upon and motivate the pursuit of those options that are likely to be more rewarding than not.

It is of interest that normal subjects decide correctly long before they have fixed a belief/opinion about which decks are good/bad and thus about how they *should* choose. This suggests that the feeling-heuristic is exploited not only to guide practical reason by limiting the number of options made available for consideration (as Damasio suggests) but also that, at least sometimes, the heuristic is exploited directly as a decision-rule.

[68] It should be noted that at this time the precise method by which feeling-images are integrated over time/trial is not known. That frequency and amplitude should be factors seems clear given the IGT findings. However, that either is more significant seems unlikely, given that no preference is exhibited between the bad decks nor between the good decks.

That is, since normal subjects come to decide and decide correctly and do so at least sometimes in the absence of belief about how they should choose, at least in some cases, it would appear that practical decision-making is itself undertaken by heuristic and automated processes.

FEELING, MOTIVATION AND "BASIC VALUE"

With respect to the question of how emotion helps in assigning or fixing the desirability of response options, Damasio offers the following account. Since it is "by dint of juxtaposition, [that] body images give to other images a *quality* of goodness or badness," (Damasio, 1994 p. 159), it is transitively, by means of juxtapositional tagging, that the goodness/desirability or badness/undesirability of particular response options comes set. (Damasio, 1994; Bechara *et. al.*, 2000). We are, Damasio contends, innately averse to particular bodily states (*i.e.,* negative feeling states). Transitively, we are averse to (*i.e.,* find undesirable) any option the outcome of which is or comes to be negatively (*i.e.,* un-preferably) feeling marked. Likewise, we are innately disposed to find other somatosensory feeling states preferable (*i.e.,* positive feeling states). *Via* feeling tagging and this set of innate dispositions to find particular states preferable, transitively we find those response options that are positively (*i.e.,* preferably) marked desirable. In this manner, the emotional tagging of response options serves to, in effect, mark/fix (and ratchet up and down from neutral) the desirability of particular response options - thereby directly motivating the pursuit or avoidance of particular response options during deliberation. (Damasio, 1994; Bechara, Damasio & Damasio, 2000 p. 297) By means of this, desirably marked response options come to preferentially command attention and cognitive resources in the manner outlined.

One might object here claiming that we have no reason to think that particular emotional responses and the feelings induced to be "valenced" (*i.e.,* positive/negative) in the manner proposed by Damasio & Bechara's account. With respect to this concern, some account is needed of the nature of these cues themselves – specifically, discussion is needed of how feelings are valenced in ways that makes them exploitable in the manner suggested by Bechara and Damasio's account.

Damasio suggests that species possess a set of innate "basic values" which "are the collection of basic preferences inherent in biological regulation." (Damasio, 1994 p. 155) This collection of preferences, Damasio continues, forms an innate and fixed feeling hierarchy in which particular feelings just are more preferable than others.[69] It is a feeling's location in the hierarchy that fixes its valence. Those preferred somatosensory states (*i.e.,* feelings) are positively valenced, while those un-preferred or aversive ones are negatively valenced. Similarly, Rolls (1998, 2000) describes an analogous hierarchy but does so in terms of an innate set of species-specific dispositions or *taxes.* And so, in much the same manner in which plants have a *taxic* disposition for states of affairs that maximize access to sunlight (*i.e.,* they are disposed to grow toward it), so too are other species taxically drawn toward certain states (*i.e.,* those prefered) and away from others (those to which we are averse).[70]

[69] It is unclear from Damasio's discussion whether the proposed hierarchy is species universal or whether there is some minor degree of individual variation in preference ordering. The latter seems, at least prima facie, more likely.

[70] The difference here might not be worth noting but the distinction between the two approaches lies in how preferred and aversive come to be determined. For Damasio what makes a state a "good" one is that it is located high in the feeling preference hierarchy. It is because it is preferred that we seek these out. Rolls takes the opposite approach suggesting that those states of affairs towards which we are highly taxically driven are those that we most prefer while those that we are highly driven to avoid we are highly averse to. That both Damasio and Rolls suggest that the preference hierarchy and taxic drives are innately fixed and species specific, suggests that ultimately there is little, if any, real difference between their claims.

In effect, Bechara and Damasio suggest, as does Rolls (1998, 2000) that these basic values and the preference hierarchy generated from them enables feeling to serve as a "common currency" – that is, a noncompensatory (*i.e.,* singular) cue - by which feelings and, thus transitively any tagged perceptual or quasi-perceptual images, may be compared and evaluated. That is, since feeling states are themselves hierarchically ranked with respect to desirability, and feeling-images tag perceptual and quasi-perceptual images, transitively "by mere dint of juxtaposition," the (perceptual and quasi-perceptual) images of situations and options themselves come to be ranked - with respect to their desirability or preferableness. By means of this value/desirability-fixing device, Damasio suggests, emotion influences practical reason and decision-making directly by marking and thus ratcheting up or down from neutral the desirability of particular response options. So doing, in turn, serves to increase or decrease motivation to pursue or avoid these (options or lines of inquiry) during deliberation.

ON EMOTION AND SOME INSTANCES OF THE FRAME PROBLEM

Having discussed a puzzle of practical reason and decision-making and outlined Bechara and Damasio's proposal, I will turn now to considering how emotion might help us in contending with some particular instances of the frame problem that arise in practical reason and decision-making.

EMOTION AND ATTENTIONAL DIRECTION

In Chapter 1, I argued that the frame problem is best understood to be a constellation of related problem kinds. One instance of which is the problem of attentional direction. Any finite and physically realized system capable of attending at all is faced with the frame problem instance of determining which "things" it is to attend and which it is to ignore. The attentional direction frame problem, then, is the puzzle of how a system could expeditiously attend to those things that are relevant while efficiently ignoring those that are not.

As I argued previously, arriving at an all-things-considered or rational (in Fodor's sense) conclusion about what is to be attended is both computationally and normatively untenable. That any system attempting to direct attention rationally has no hope of ever doing so follows for two reasons. First, any rational (in Fodor's sense) system would immediately be forced to contend with a Rylean (1949) regress with respect to arriving at a conclusion about what particular standard should be employed. That is, any rational system must, before doing anything else, arrive at a conclusion about what criterion of, for lack of a better term, "attention-worthiness" should be relied upon. Moreover, it must, if it is to be a rational system, arrive at this conclusion rationally. We have, as I argued in Chapter 3, reason for thinking that no finite system that goes about arriving at conclusions in this way could succeed.

Second, even if we assume that such a system could (somehow) determine which standard it is to apply, it would, as a rational system, be faced with a computationally intractable task. Specifically, in order to arrive at a conclusion about what "things" should be preferentially attended, such a system would need to first, consider each "thing" and second make a rational determination as to its attention-worthiness. So doing, of course, requires that each "thing" be attended and given rational consideration.

That any system engaged in a rational process of attention-direction would be effectively cognitively/attentionally paralyzed, when taken in conjunction with the uncontroversial claim that normal humans are not attentionally paralyzed, suggests that some automated heuristic is (or set of heuristics are) at play that assists in directing our limited attentional resources. There are a number of reasons for thinking this a plausible suggestion. First, that one's attention can be "grabbed" in a somewhat reflex-like manner often in the absence of belief and often in spite of one's beliefs, supports the claim that attention is directed, at least in part, by automated heuristic processes. For example, most can think of a situation in which one's attention has been drawn to something while s/he was otherwise engaged in thinking, perhaps quite deeply, about something entirely unrelated. Likewise, most have found themselves in a situation where one's attention is drawn quite automatically to something that one believes she/he should not be attending. While there are many examples, most rather lurid, seeing someone with a pronounced physically deformity provides a suitable example. One "knows" not to stare, but one's attention is drawn nonetheless. That we can decide what we *should* attend and still rapidly and automatically attend to the "wrong" things (with respect to what we have decided we should attend) provides further support for the claim that attention is at least in good part automatically and heuristically directed.

In discussing Damasio's somatic marker proposal - in which emotion, by transitively fixing the desirability of particular images and options, facilitates deliberation by highlighting and/or culling options for and from further consideration - an additional strand emerged that, while briefly touched upon above, warrants separate discussion. The influence of emotion on attention has received considerable treatment in the literature.[71] The majority of these studies focus on establishing the existence of an

[71] Etcoff & Magee, 1992; Bargh, Chaiken, Govender & Pratto, 1992; Greenwald, Klinger & Liu, 1989; Markus & Kitayama, 1991; Niedenthal, 1990; Niedenthal & Kitayama, 1994; Pratto & John, 1991; Clark & Isen, 1982; Ingram, 1984; Bower & Forgas 1987; Lang 1984; Teasdale, 1983; Niedenthal & Showers, 1991; Niedenthal & Setterlund, 1994; Martin, William & Clark, 1991; Kunst-Wilson & Zajonc, 1980; MacLeod, Mathews & Tata 1986; Mathews & Klug, 1993; Mathews & MacLeod 1985; Broadbent & Broadbent, 1988; Dagleish & Watts, 1990; Hunt & Ellis, 1999; Kraiger, Billings & Isen 1989; Berkowitz 2000.

affective/emotional attentional congruency effect – the phenomenon whereby a subject's current emotional state directly influences the stimuli preferentially attended in the subject's environment. Specifically, subjects are found to preferentially attend to those stimuli that are emotionally "congruent" with/to their current emotional state. And so, for example, anxious subjects attend preferentially to the objects of their anxieties (*e.g.*, arachnophobes preferentially attend to spiders and show significant disregard for other stimuli.) (Mathews & MacLeod, 1985; MacLeod, Mathews & Tata, 1986; MacLeod & Mathews, 1991; Beck & Clark, 1988; Martin & Williams, 1990) Similarly, employing a variant of the "affect priming" studies discussed in the previous chapter, Neidenthal & Setterlund (1994) found that subjects in whom an emotional state of "joy" had been induced preferentially attended to "happy" stimuli (*e.g.*, positively emotionally valenced words like "joy" and "cheer") while those in whom a "sad" state had been induced preferentially attended to "sad" stimuli (*e.g.*, negatively emotionally valenced words like "despair" and "weep.") The emotional/affective state attentional congruency effect is also readily apparent in subjects with generalized anxiety disorders (Mogg, Mathews & Wienman, 1987), panic disorders (Ehlers, Margraf, Davies & Roth, 1988; McNally, Reimann & Kim, 1990) and post-traumatic stress disorder (McNally, Kaspi, Reimann & Zeitlin, 1990).

Since at least some of what emotions are is explicable in terms of the operations of automated appraisal mechanisms, and the amygdala, we have every reason to think, is the structure mediating these appraisals, I will turn now to considering the role occupied by this structure in the direction of attention. That the amygdala exerts considerable influence on the operations of attentional direction and guidance is well established.[72] Holland, Han & Gallagher (2000), for example, found that the central nucleus of the amygdala (the principle output region of the structure) and its pathways to the substantia nigra and dorsal striatum are directly implicated in the direction of attention. Specifically, Holland *et. al.,* (1999) found that rats with ablated central nuclei are entirely incapable of exhibiting a "conditioned orienting response" to stimuli that have been, through conditioned learning, associated with reward and punishment. Normal rats, in contrast, exhibit a robust orienting response toward previously learned stimuli, preferentially attending to those stimuli that are rewarding and punishing.

As discussed in the previous chapter, that the amygdala is responsible for both the maintenance of emotional memory (it serves as both a mechanism of conditioned learning and repository of emotional memory) and appraisal of the emotional significance of stimuli is well established. Complete amygdala ablation, recall from previous discussion, results in an inability to respond to emotionally significant stimuli and an inability to learn to respond to new stimuli – subjects are in effect incapable of finding stimuli emotionally significant. The environment is for these subjects entirely devoid of emotional significance/value.

Lesioning of the central nucleus aspect of the amygdala (the principle output region) does not affect the learning of new stimuli nor does it affect the ability to emotionally appraise stimuli. However, by lesioning this aspect and thus effectively the connection between the amygdala and those regions directly implicated in attentional control, subjects become incapable of exploiting the information provided by the amygdala to direct attention toward emotionally significant stimuli.

[72] Helmuth, 2003; Gloor, 1992; Kapp, Whalen, Supple & Pascoe, 1992; Rolls, 1999; LeDoux, 1996; McDonald, 1992; Halgren, 1992; Kapp, Supple & Whalen, 1994; Holland & Gallagher, 1999; Holland, Han & Gallagher, 2000.

Similarly, the amygdala is directly implicated in the process whereby surprising or novel stimuli are preferentially attended. Holland & Gallagher (1993) and Han, Holland & Gallagher (1999) note, for example, that rats with ablation of the central nucleus (of the amygdala) or lesioning of its connections to the basal forebrain, exhibit no "surprise induced enhancements," failing to preferentially attend to novel and surprising stimuli in their environment. That is, while normal subjects rapidly orient/attend to those features of the environment that are novel/unusual, those with damage to the amygdala are entirely incapable of doing so.

Furthermore, Cahill *et. al.,* (1994, 1996), Adolphs *et. al.,* (1997) and Baxter & Chiba (1999) demonstrate that the substantia nigra and nucleus basilis regions of the basal forebrain are directly implicated in the regulation of attention in both sustained and selective attentional tasks. Earlier, Carli, Robbins, Evenden & Everitt (1983) found that basal forebrain lesioning significantly impairs the performance of rats on a sustained attentional task, dramatically interfering with their ability to maintain attention over time. Of particular interest are Holland & Gallagher's (1993) findings that identical behavioral results are obtained following ablation of or lesioning to the connections between the central nucleus of the amygdala and the regions implicated in attentional direction (*i.e.,* the substantia nigra and striatum). This implies, Holland & Han conclude, that while areas of the basal forebrain are responsible for the mechanics of attentional direction, it is the amygdala (*via* projections to these regions) that is responsible at least in part for determining how (*i.e.,* upon what) attention will be directed. And so, *via* automated influences on structures in the basal forebrain and striatum, the amygdala-realized emotional appraisal module, Adolphs (2000) explains, "helps to select particular aspects of the stimulus environment for disproportionate allocation of cognitive processing resources; in other words, the organism preferentially processes information about its environment that is most salient."

And so, while the details of the operations of attentional direction are still under study, it is clear that in all species that possess an amygdala or analogous limbic structures, (i) there are defined anatomical pathways from the amygdala to regions implicated in attentional control, (ii) ablation or lesioning of the amygdala or its connections results in the cessation of the orienting response toward emotionally significant stimuli, (iii) ablation or lesion of the amygdala or its connections eliminates surprise-induced enchancement of attention, and (iv) ablation of the amygdala dramatically interferes with the guidance and maintenance of attention on both sustained and selective tasks.[73]

We can, then, arrive at a conclusion about one way in which emotion might help us contend with a few of the frame problem instances that arise in these domains. Specifically, with respect to the attentional direction instance of the problem, emotion clearly plays a role in directing attentional resources upon emotionally significant material. That such an heuristic should be far more expeditious than the "rational" alternative is clear. That the contents of emotional memory are composed of unlearned and learned "triggers" both of which are evolutionarily constrained (either innately fixed or governed by species-specific learning preparedness rules) provides reason for thinking, as Ekman (1983), Izard (1991, 1993), LeDoux (1996), Damasio (1994), Panksepp (1998) and Cosmides & Tooby (1990) suggest, that emotion preferentially directs attention toward stimuli that are of evolutionary significance. The influence of emotion

[73] Muir, Dunnett, Robbins & Everitt, 1992; Muir, Everitt & Robbins, 1996; Pang, Williams, Egeth & Olton, 1993; McGaugh *et. al.,* 1992; McGaugh, Cahill & Roozendaal, 1996; McGaugh, 2004; Baxter & Chiba, 1999.

on attention serves, in effect, as LeDoux (1996) argues, as an automated "early warning system" alerting us to the presence of evolutionary significant states of affairs (*e.g.,* dangers, rewards, punishments). Insofar as attention enlists cognitive processes (Simon, 1967; Kosslyn & Koenig, 1994; Rolls, 1999; Adolphs, 2000), that is, insofar as we preferentially "think" about those things to which we attend while ignoring (*i.e.,* cognitively ignoring) those unattended, emotion serves the heuristic function of directing and limiting what it is that we "think" about and the lines of inquiry pursued. Likewise, *via* the mechanisms outlined by Damasio in which the preferable-ness or desirability of particular (images of) states of affairs and response options come to be transitively fixed, emotion serves to (by ratcheting up or down the desirability of options and thus our motivation to pursue or avoid these respectively) further facilitate attention and guide deliberation by directing the ends and lines of inquiry pursued.

It is, of course, highly unlikely that emotional significance should be the only heuristic at play directing attention. Rather, it is far more likely that there are number of such processes. It is also clear that we can, in the absence of emotionally significant stimuli, consciously direct our attention (*e.g.,* I can attend to a footnote, astrological chart, line of code, particular DNA sub-sequence or any other potentially emotionally banal thing). However, in much the way that Simon (1967), LeDoux (1996), Ekman (1983), Damasio (1994), Panksepp (1998), and Cosmides & Tooby (1990) suggest, it is also quite clear that emotional appraisals automatically interrupt these other processes. And so, regardless of what I might be thinking about and wherever else my attention might be directed, attention will be automatically and immediately drawn to emotionally significant material.

That relying on this heuristic to direct attention should be far more expeditious than the rational alternative seems apparent. However, the question remains as to how "good" such an heuristic would be. And so, we need some reason to think the attention-directing heuristic outlined to satisfy the normative horn of the frame problem. There are a number of reasons for thinking this quite likely with respect to the rather global ends of survival/fitness. The amygdala (and the limbic structures quite generally) is a phylogenically ancient brain structure that serves the same function in all animals that have one. It is far older than the neocortex and thus has had the benefit of millennia of natural selection. Given the manner in which natural selection operates it is quite reasonable to think that there were animals that directed their attention differently (*i.e.,* based upon some other criteria). It is also quite reasonable to think that the inability to preferentially attend to dangers, rewards, punishment, disgusting things, surprises and the like would have resulted in the demise of this line. And so, we have at least some reason for thinking that this automated attentional directing device makes salient those things that are, given our particular evolutionary history, relevant.

This establishes only the likelihood that relying on such an heuristic was at some point beneficial. However, with respect to the gross categories of dangers, threats, rewards, punishments, surprises and the like, we have no reason for thinking the current environment to be markedly different from that of our ancestors. That is, at least some (if not all) of the unlearned inducers/triggers are still present in the current environment – animals still growl and bare their teeth, food still rots, excrement still abounds, objects still loom, the world is still full of surprises and so on. The same would appear to hold for learned triggers. The current environment, we have every reason to think, is both full of dangers, threats, surprises, disgusting things etc., and patterned in ways that can be exploited by mechanisms of conditioned learning (*i.e.,* the current environment is not random). That we can today find states of affairs rewarding, punishing, threatening etc., and that we can associate these responses with their inducers supports this. And so, since the attentional direction heuristic is the result of evolutionary pressures and we have reason for thinking the current environment to be, with respect to the gross categories of emotionally significant things, relevantly analogous to those of the past, we also have reason for thinking that relying on this heuristic should be beneficial today. Put another way, automatically orienting toward dangers, rewards, punishments, surprises and the like, is very likely to be as useful a strategy today as it was on the savannahs of the past.

PROBLEM-SEQUENCING AND META-PLANNING

By influencing, in an automated and obligatory manner, motivation and attention, emotions might play a role in helping us contend with the problem-sequencing and, *via* this, the meta-planning puzzles.

Suppose that one is engaged in some highly cognitive and rational task – undertaking some mathematical proof, for example. Suppose further that a tiger stalks in, a fire alarm sounds, or some other (learned or unlearned) emotionally significant stimuli is introduced. Assuming that these are, in fact, "triggers" or "inducers" of the relevant sort, attention is quite automatically redirected and motivation is shifted. With respect to our mathematical aims, the presence of a tiger or the fire alarm is not likely to be directly relevant. However, with respect to fitness/survival, these are quite relevant. In this regard, then, the tiger or alarm is relevant to our mathematical endeavors, for we have here a good example in which solving the wrong problem first will make it such that one cannot solve the other. Somehow, we must expeditiously determine whether to continue with our proof and then think about the tiger, or sequence these problems the other way round.

The modest point I am making is that whatever we were thinking about (and thus whatever immediate ends we might have been pursuing), is interrupted quite automatically by the influences of the automated emotional appraisal mechanism on motivation and attentional direction. We immediately stop thinking about the proof (*i.e.,* we both cease attending to the proof itself and no longer have the immediate goal of contending with *that* puzzle) and focus instead quite automatically on the current tiger-filled situation. In so doing, we cannot help, as Adolphs (2000) suggests, but enlist those cognitive processes that were engaged in thinking about math to start thinking about the current tiger-filled situation. In so doing, a simple problem-sequencing operation has been undertaken quite automatically and without the need for our having to arrive rationally at a conclusion about this. Likewise, by automatically maintaining or sustaining attention on emotionally significant material (tigers in this case), emotion helps to keep those cognitive processes enlisted, focused on the situation at hand (and not the math), again, without the need for our having to first arrive at an all-things-considered rational conclusion. Furthermore, by means of the operations of motivational direction outlined by Damasio's proposal, we come to have a much stronger immediate desire to avoid the tiger-filled situation than we do to contend with the math puzzle.

This is, of course, not to say that attention cannot be redirected away from the tiger. There are, so it would appear, two possible means by which attentional redirection might occur. First, the feeling preference hierarchy provides one means by which this might be occur. Were some situation to obtain that is appraised to be more emotionally significant than the one under current consideration, that situation - since it is, in effect, more emotionally "pressing," (*i.e.,* the feeling generated is markedly lower or higher on the preference hierarchy and thus of lesser or greater desirability) - would come to command attentional and cognitive resources. For example, let us suppose then that we are, once again, working on our proof and a fire alarm sounds. This, assuming it is a learned trigger, would command attention and redirect immediate motivational aims - we would stop thinking about our proof. However, were a tiger to now stalk in, we might very well find this situation to be more emotionally "pressing" (*i.e.,* resulting in a feeling state to which we are more averse) than that induced by the alarm. We might, *via* the influences on motivation and attentional direction, stop thinking about the fire alarm and start thinking about the tiger. By means of the influence on motivation, attention and, *via* this, cognitive resources, might come to be redirected quite automatically in this manner.

Second, *via* the "high road" pathway proposed by LeDoux, attention might come to be redirect attention away from the tiger and back onto the proof. For example, once attention has been focused and motivational and cognitive processes engaged (in thinking about the tiger situation rather than math) we might come to realize, *via* processing by other consumer subsystems, that the tiger is, for instance, leashed, trained or stuffed. *That* scene (*i.e.,* the image constructed following further processing), when re-appraised, might no longer hold any emotional significance – it might no longer be an undesirable one or one to which the agent is averse. And so, assuming that the agent has not learned to respond emotionally to stuffed toy tigers, (*i.e.,* such things do not feature as "triggers"), the ersatz tiger would (as no longer an emotionally and hence a motivationally significant situation) also no longer commands attention.[74] And so, by freeing attentional and cognitive processes from considering this further, we return to whatever it was we were thinking about.

Again, the point here is a modest one. That we are compelled to attend to and thus "think" about emotionally significant material and stop thinking about whatever it was that previously had our attention/cognitive resources, is a simple heuristic device for problem-sequencing and, by way of this, meta-planning. Since it is such a simple heuristic (relying upon only one cue) we have good reason to think it to be highly expeditious and, in any case, every reason for thinking it far more expeditious than the computationally burdensome "rational" alternative.

META-PLANNING: PROBLEM-SETTING AND ENDS-SELECTION

Drawing upon above discussion and Bechara & Damasio's model, emotion might help us in contending with the meta-planning instance of the problem by serving as an automated problem-setting and ends-selection heuristic. As discussed previously, one natural hypothesis with respect to the deficiencies exhibited by amygdala-damaged patients is that they are, in effect, incapable of valuing states of affairs and thus of valuing and hence desiring particular states of affairs over others. That they are incapable of so doing suggest that they are also unable to set situations as posing problems for themselves. For example, if one is incapable of finding an objectively punishing outcome undesirable, then one is also incapable of taking the current situation to be a problematic one. Likewise, if one cannot "care about" or desire any state of affairs any more than any other, then one also cannot care about any of the states

[74] Even the surprise-induced enhancements would decline as we become habituated to the presence of the tiger.

of affairs that might result from one's action. Since the amygdala-damaged are indifferent to both the current situation and all of the possible outcomes (none are taken to be rewarding or punishing), from their perspective, there really is no problem posed by this or any other situation.[75]

And so, one plausible suggestion for why it is that amygdala-damaged patients are unable to attain the conceptual stage is that they fail to set the task as posing a problem for them. That is, if one *really* cannot assign a value (*e.g.,* desirable or undesirable) to any of states of affairs, then one, in effect, fails to have not losing as one of one's goals (*i.e.,* one fails to be motivated by this prospect). The amygdala-damaged, then, might fail to contend effectively with both real-world problems and those of the IGT because, from their perspective, no problems are posed (by and of them). In contrast, damage to area VMF appears to have a different effect in this regard. These patients know that situations in the world and the IGT pose problems for them. They also know that these situations require consideration and are motivated to contend with them (some, in fact, even know how best to contend with them), they just fail to do so effectively. Their deficit lies not in the inability to set situations as posing problems, but rather in the inability to effectively manage and contend with those that are set. Specifically, while damage to area VMF does not preclude subjects from appraising the emotional significance of *current* situations (*i.e.,* finding actual outcomes rewarding/punishing), it does prevent them from appraising the emotional significance of *future* (*i.e.,* hypothetically entertained projections) situations and outcomes. With

[75] The point I am belaboring here is a simple one. Someone, somewhere is renting a car at this moment. This person might get a blue, white, silver, green etc., colored car. I truly do not care what color car this person gets. Since I am indifferent, this situation poses no problem for *me.* If, however, one is entirely indifferent to *all* states of affairs and thus *all* potential outcomes, then no situation would (be taken to) present a problem.

respect to the anticipated feeling (*i.e.,* the feeling likely to result from their actions), VMF-damaged patients are no different from those with amygdala damage – all futures are emotionally and hence also motivationally equivalent.

It is then reasonable to think that emotion might help us contend with the meta-planning instance of the problem by heuristically helping us to set situations as posing problems for ourselves. A quite simple heuristic emerges from this: set as problematic those situations that are emotionally significant. Furthermore, the degree to which a situation is taken to be problematic (for the system) is given by reference to the innate feeling hierarchy. Those situations that are most emotionally significant (*i.e.,* the most immediately un/desirable) are, *via* the sequencing operation, the most emotionally and thus motivationally pressing.

Put somewhat differently, before one can even begin thinking about how best to order the sequence in which one will contend with the problems posed of them, one must be able to discern that there are problems with which s/he needs to contend. The automated mechanisms of emotional appraisal would appear to play a role here, helping to alert us to the fact that our current situation is problematic one and thus one in need of our consideration. In so doing emotion might, when taken in conjunction with the motivational role proposed by Damasio's account, help in setting the ends and goals (and thus lines of inquiry) that will be pursued. Understood in this manner, emotion, by setting situations as problematic helps to load the problem-sequencing queue by informing us that (i) some situation is, as un/desirable, problematic and (ii) that, at some point, we are to direct attentional and cognitive resources towards considering (i.e, "thinking" about) this further. In so doing, emotion helps to direct both the lines of inquiry and ends pursued and the order in which those problems posed are considered.

Simply put, with respect to meta-planning, emotion appears to help us in determining both *what* to think about (*i.e.,* with what situations we should contend – the "problematic" ones – and thus with what ends are to be pursued) and *when* we should think about it (*i.e.,* the most emotionally/motivationally "pressing" first, then the next and so on).

Relying on a simple one-cue heuristic to determine whether a situation poses a problem (for the system) and thus presents something with which it must contend, should be far more expeditious than the rational alternative whereby the "problematicity" of each situation must be rationally determined. Along the same lines as those discussed previously, so doing would require the rational system to arrive at a conclusion (and do so rationally) concerning which standard of "problematic" it is to employ. That it should be sufficiently normatively "good" to satisfy (a suitably weakened version of) the second horn of the dilemma posed by the frame problem is likely as well.

While the role proposed above for emotion in helping us to contend with some instances of the frame problem that arise in practical reason is rather indirect, (*i.e.,* it relies upon the influences of emotion on motivational and attentional direction), Damasio argues that emotional feeling is exploited *directly* in helping us to contend with the meta-planning and problem-sequencing instances of the problem. And so, with respect to the cognitive paralysis brought on by the endless consideration of options and implications exhibited by VMF-damaged patients, Damasio explains,

> An automated somatic marker would have helped the [VMF-damaged] patient in more ways than one. To begin with, it would have improved the overall framing of the problem. None of us would have spent the amount of time the patient took with this issue, because an automated somatic marker device would have helped us detect the useless and indulgent nature of the exercise. If nothing else, we would have realized

> how ridiculous the effort was. At another level, sensing the potentially wasteful approach, we would have opted for one of the alternative dates the equivalent of tossing a coin or relying on some kind of gut feeling for one or the other date. Or we might simply have turned the decision over to the person asking the question and replied that it really did not matter, that he should choose.
>
> In short, we would picture the waste of time and have it marked as negative; and we would picture the minds of others looking at us, and have that marked as embarrassing. There is reason to believe that the patient did form some of those internal "pictures" but that the absence of a marker prevented those pictures from being properly attended and considered. (Damasio, 1994 pp. 193-4)

And so, by exploiting the body-loop or as-if loop marking devices, Damasio suggests that the feeling generated of the automated appraisal of the imagined "scene" of undertaking a perseverative and exhaustive consideration of the options and implications would come to be marked negatively. By somatically "warning" us in this way (*i.e.,* marking that scene as undesirable), attentional, and *via* this cognitive resources, are drawn away from the particular planning task itself and focused onto the meta-planning task. That is, we stop thinking about particular options (*e.g.,* plans and reasons for meeting on Tuesday rather than Thursday) and start thinking about whether we should be contending with this problem at all and, if so, if we should be contending with it in this way.

It is worth noting, however, that, given the role proposed by the VMF in Bechara and Damasio's model, there is no reason to think that this meta-planning or problem-sequencing heuristic should be inflexible. Rather, since the VMF functions to somatically mark or tag both perceptual and quasi-perceptual (*i.e.,* suppositional or hypothetical) images, and there is reason to think, given LeDoux's findings, that "factual" inducers may be highly complex and abstract, the feeling generated in response to such scenes could be highly context-sensitive. Through learning and the construction of fact-feeling sets, particular contextual, normative and social standards might very well be incorporated into the appraisal process. And so, in situations in which one has come to learn that a higher standard of accuracy than is usually expected applies, the feeling-marker apparatus could be sufficiently flexible to accommodate this context-specific information by means of the operations outlined. For example, one might come to learn that in some specific problem-contexts what usually constitutes an "embarrassing" display of detailed consideration does not apply.[76]

HAMLET'S PROBLEM: STOPPING AND DECISION RULES

> In situations in which there is remarkable uncertainty about the future and in which the decision should be influenced by previous individual experience, such constraints permit the organism to decide more efficiently with short time intervals. … In the absence of a somatic marker, options and outcomes become virtually equalized and the process of choosing will depend entirely on logic operations over many option-outcome pairs. The strategy is necessarily slower and may fail to take into account previous experience. This pattern of slow and error-prone decision behavior we often see in [VMF] patients. Random and impulsive decision-making is a related pattern. (Damasio, 1998)

And so, Bechara and Damasio are proposing a quite simple heuristic for facilitating decision-making and practical reason - that emotional feeling tags (*i.e.*, somatic markers) are automatically exploited as noncompensatory heuristic cues for both culling and highlighting particular plans and options. By means of this "cognitive guidance"(Damasio, 1994 p. 130) heuristic, the set of plans and options is limited in size *via* culling, while motivational, attentional and cognitive resources are directed upon those that are highlighted. And so, during deliberation, Damasio contends, the set of options upon which reason is applied is pre-limited by this automated culling/highlighting device. It is worth noting too, that *via* the "as-if loop" device set out

[76] For example, detailing the reasons why one should visit Damasio's lab on one day versus another might very well be highlighted as embarrassing, while taking that much care when deciding, for instance, when to invade Normandy would not. That the "factual" aspect of the fact/feeling set can be complex, abstract or highly contextually specific seems clear from both LeDoux's and Damasio & Bechara's discussions.

by Damasio, which bypasses the necessarily serial and thus slow “body-loop” appraisal mechanism, the tagging and culling/highlighting process should be quite expeditious.

In this respect, then, emotion-induced feeling tags serve as a cue that is exploited by an automated heuristic stopping rule which, when operational, prevents our further considering (*i.e.,* attending to, thinking about and pursuing) particular options and plans. Furthermore, once the set has been winnowed (*i.e.,* once particular options have been automatically culled from further consideration), Damasio & Bechara suggest that these same feeling tags, in effect, serve double-duty as a cue that, by transitively fixing the desirability of options themselves, is exploited by an automated heuristic decision rule. The feeling tag associated with each response option is, by ratcheting up or down (from neutral) the desirability of a particular response option, directly exploited to guide decision-making by both serving as a rudder on attention and cognition and motivating the pursuit/avoidance of particular options. That normal subjects choose correctly even when they have no opinion about how they should choose and VMF-damaged patients choose poorly even when they have correct opinions about how they should choose, suggests that feeling might be heuristically exploited as a noncompensatory cue to guide decision-making.

Fodor’s Hamlet’s challenge is the puzzle of when to stop “thinking” and choose/act. Earlier I argued that no rational (in Fodor’s sense) system could be expected to adequately contend with this problem given the regress that must be (rationally) anchored. Put another way, no finite system that attempts to arrive at a conclusion rationally could contend with this problem, for none can be expected to arrive at a rational conclusion about when to stop thinking. As such, no rational and finite system should ever be expected to choose. And so, as I argued, rational consideration results in cognitive and, by way of this, practical paralysis. Since we do choose and we do not appear to be instantiations of a rational system, I suggested that it is reasonable to think that we contend with this puzzle heuristically. Bechara and Damasio's account provides a parsimonious explanation of *one* way in which we might contend with this problem as well as providing a model by which this heuristic might be realized. Simply put, we "stop thinking" about those options that are likely to result in undesirable states of affairs – that is, those that are negatively feeling-tagged. Likewise, with respect to the decision aspect of Hamlet's puzzle, we choose or act when a suitably desirable (*i.e.,* positively feeling-marked) option is discerned.

More globally, emotion might play a further role here by helping us in determining when to stop thinking about a particular problem posed. That is, we stop thinking about a particular problematic situation when it is no longer emotionally and thus motivationally significant. Particular problematic situations would no longer be emotionally significant when they no longer obtain. Put another way, we stop thinking about a problematic situation when it ceases to pose a problem for us (*i.e.,* when the *current* situation is no longer motivationally aversive). In this respect then emotion might also serve as simple means by which we are alerted that some satisfaction condition has been met. That is, given some problem posed and some suppositionally entertained plan under consideration, we stop thinking and act when so doing results in a state of affairs in which a problem is no longer taken to exist (*i.e.,* one that is not appraised to be undesirable/aversive).

With respect to the tractability horn of the dilemma, the noncompensatory feeling-cue heuristic stopping and decision rules described by the somatic marker model should be far more expeditious than the rational alternative of cognitive paralysis.

Relying on one cue to winnow/highlight options is, of course, much less computationally burdensome than the rational alternative that requires that *all* of the available evidence be considered. Furthermore, not only would we expect one-cue strategies like those proposed by Damasio and Bechara to be much less computationally intensive than the rational (in Fodor's sense) alternative, they are also more efficient than traditional "benchmark" rational decision-making tools or strategies (*e.g.,* Dawes' rule, Franklin's rule and multiple linear regression) that consider *only* a great deal of the available information (Gigerenzer, 1999). That simple heuristic stopping and decision rules should be far more expeditious and far less computationally demanding than more complex processes that consider *all* or *most* of the available evidence, however, is of no great surprise.

While certainly expeditious, one might naturally object that no heuristic, particularly any as simple as those proposed by Damasio, could possibly be sufficiently accurate, reliable or otherwise "good" to satisfy the normative horn of the frame problem. Underlying this concern is the intuition that any gains in speed *must* come at the price of a decrease in accuracy. That is, the concern runs, since noncompensatory heuristics are simple, computationally frugal, rapid and, by design, fail to consider all of the available information (in fact, considering very little), they *cannot* adequately satisfy the normative horn of the dilemma.

While the IGT and the real-world deficits exhibited by VMF and amygdala-damaged patients are compelling, there may be resistance to the idea that a noncompensatory heuristic like those underlying Bechara & Damasio's proposal could be adequately normatively "good." Such resistance might lead one to reject the proposal outright (and look to another explanation) on the grounds that *in principle* the one-cue decision-heuristic posited is far too simple to possibly satisfy the normative horn of the dilemma. And so, we need reason for thinking, independently of Bechara and Damasio's claims, that one-reason decision-making could be sufficiently accurate.

In Chapter 3, I argued that the normative standard underlying Fodor's pessimistic conclusion should be weakened in light of its *in principle* unsatisfiability. While I did not specify *how* this normative rationality principle should be weakened, as this was not my aim, *that* it should be weakened to something that is at least satisfiable by a finite and physically realized system seems clear. In assessing the normative "goodness" of any necessarily irrational (in Fodor's sense) procedure, we cannot demand the impossible – and thus should not demand *complete* correctness. Rather, all that is needed is reason for thinking such heuristics to be "good enough."

With respect to this, that an heuristic is accurate/correct at a rate that is only slightly better than chance, it is worth noting, would be sufficient to establish the modest point that that heuristic *helps* in contending with the frame problem, assuming of course that is also expeditious. We have, however, reason for thinking very simple heuristics like those proposed by the somatic marker account to be capable of far greater accuracy than chance.

Gigerenzer (1999) investigated the accuracy, robustness and computational requirements of a number of decision-making strategies. Both simple heuristics that relied upon one-cue (noncompensatory) stopping and decision rules and complex and computationally intensive "rational" strategies were examined. The aim was to determine how accurate a simple heuristic can be that (1) fails to incorporate all of the available evidence and (2) is noncompensatory, that is, that relies upon only one cue. While we need not go into the particular details, Gigerenzer's team found that with respect to accuracy, the three simple "fast and frugal" one-cue heuristic rules investigated outperformed the computationally intensive "rational" strategies (*e.g.*, Dawes' rule, Franklin's rule and Multiple linear regression) that consider all of the available information. (Gigerenzer *et.al.*, 1999 p.87) Likewise, by eliminating the problem of data "over-fitting" which plagues classical (rational) strategies, these simple heuristic strategies were also found to be remarkably robust. (Gigerenzer *et. al.*, 1999, pp. 109-110, pp. 127-136)

For present purposes, the upshot of Gigerenzer's findings is that "more is not always better" when it comes to determining both when to stop "thinking" about options and when to decide/choose among them. Rather, the evidence suggests that under conditions of uncertainty "less is more" since one-cue stopping and decision-rules can be as accurate and flexible as computationally intensive rational strategies and are decidedly more computationally lean and expeditious. He explains,

> Fast and frugal heuristics that embody simple psychological mechanisms can yield inferences about a real-world environment that are at least as accurate as standard linear statistical strategies embodying classical properties of rational judgment. This result liberates us from the widespread view that only "rational" algorithms, from Franklin's rule to multiple linear regression, can be accurate. (Gigerenzer, 1999 p. 95)

It should be noted that I make no claims about the psychological plausibility of the particular heuristics investigated by Gigerenzer's team. For present purposes, I need not weigh in on the question of whether these particular heuristics are employed by us. What is significant is the result that, contrary to our intuitions, one-reason stopping and decision heuristics can be as accurate as the classical rational alternatives. And so, since very simple heuristics can be both expeditious and accurate, we have no reason for thinking that the feeling-based heuristic stopping and decision rules outlined by Bechara & Damasio necessarily violate the normative horn of the frame problem.

One of the ways that fast and frugal heuristics gain their competitive advantage is by exploiting cues with high "ecological validity." The ecological validity of a cue, Gigerenzer explains, "is the relative frequency with which the cue correctly predicts the criterion defined with respect to the reference class."(Gigerenzer *et. al.,* 1999 p.84) Simply put, a cue has a high ecological validity if it is highly predictive of some other criterion. For example, that a city is a national capital is a very good indicator that it has a very large population. As such, this cue has a high ecological validity with respect to population size.

We have reason for thinking the ecological validity of feeling-markers likely to be high. This follows for a number of reasons. First, Bechara's findings on the IGT are instructive in this regard. That normal subjects generate anticipatory *SCR*s before selecting from any deck and that higher amplitude *SCR*s are recorded before subjects select from "bad" decks, suggests that *anticipatory* feeling is, in fact, highly predictive of the *objective* disadvantageousness of options and outcomes on the IGT.

Second, the IGT results aside, the operations of emotional learning and appraisal undertaken by the amygdala are the result of millennia of evolutionary pressures (*i.e.,* the amygdala and other limbic structure are phylogenically ancient). This suggests that the kind of stimuli that feature as *unlearned* triggers or inducers of emotion should be those that are evolutionarily significant and thus relevant to survival/fitness. Furthermore, drawing upon Seligman's learning preparedness account, the particular learning rules undertaken in the amygdala are, we have reason to think, evolutionarily fixed, as well. This suggests that the kinds of things that result in an emotional response (both the learned and unlearned triggers) are those that are of evolutionary significance and thus relevant to survival/fitness. One must, of course, be careful, for what had a high ecological validity in some past environment need not today.

As discussed previously, the current environment presents many of the same *unlearned* triggers as those of past environments. These unlearned inducers still correlate with the same kinds of (evolutionarily significant) problems posed of our ancestors. For example, both then and now, rotten food and excrement abound and cause illness or death. This suggests that, with respect to the *unlearned* inducers, it should be as beneficial to mind these now as it was then.

Furthermore, it seems quite reasonable to suggest that, with respect to emotional significance, the current environment should not be radically different from that of our ancestors. Then, as now, there are rewards, punishments, dangers, threats, disgusting things, surprises and so on. Since the underlying nature of these do not appear to have radically changed, if it was beneficial to mind these then, it should also be beneficial to do so today. That is to say, with respect to the conditioned/*learned* inducers, while the particular kinds of things that might come to be learned to be rewarding/punishing *etc.,* might change with the environment (*e.g.,* a cappuccino versus a quick sip at a feces-free watering-hole), *that* the environment still affords *us* rewards, punishments *etc.,* has not.

At base, however, the validity of the cues exploited by the heuristics outlined by Bechara and Damasio rely upon iterative interactions with and feedback from an agent's particular environment, as the IGT results indicate. This suggests that the ecological validity of feeling markers should increase over time. Consider the IGT as an example. At the outset, subjects sample randomly from each of the decks. This is so, Bechara & Damasio explain because the integrated fact-feeling set is highly uninformative. In essence, the agent has had insufficient interactions with the environment to generate a sufficiently informative *on balance* feeling for each deck. And so, at the outset, no deck is taken to be any more desirable than any other. Over time, however, normal subjects come to choose from the "good" decks. This is explicable by their having had sufficient interaction with and feedback from (*i.e.,* feedback of their emotional responses to) the environment to generate exploitable *on balance* feeling tags for each deck/option. And so, over time the integrated *on balance* feeling that comes to be associated with each response option (*i.e.,* each deck) should become a more reliably indicator of how it will likely feel to enact each response option. That the *on balance* feeling is generated from individual appraisals of the outcome of particular card/deck selections, suggests that, over time and with repeated interaction with a stable environment, how an option comes to feel *on balance* should come to be highly predictive of whether it is likely to be *on balance* rewarding or punishing. That all normal subjects come to generate unique and substantial *SCR*s before choosing from the bad decks, suggests that the *on balance* feeling that comes to be associated with these decks is highly indicative of an objectively punishing outcome. That feeling cues come to, over time, correctly predict the likely outcome of selecting from each deck (with respect to reward and punishment) provides good reason for our thinking feeling to, over time, come to be a cue with a sufficiently high ecological validity.

It is, then, at least reasonable to suggest that the ecological validity of feeling markers should increase over time with interaction with non-random environments, like the IGT. Our environment, we have reason to think, is a sufficiently stable one. It is also quite reasonable to think that each normal human has been, since infancy, generating and fine-tuning the *on balance* feeling component of myriad integrated fact-feeling sets.

If so, then it is quite reasonable to think that feeling should be, in normal adults, a cue with a sufficiently high (and ever increasing) validity.

Clearly, there are limits to the usefulness of this particular cue and thus limits to the heuristic itself. For example, in cases in which there is no fact-feeling history upon which to rely (*i.e.,* in truly novel "factual" situations that are not analogous to *anything* that we have previously experienced), feeling would be quite uninformative. However, it is also quite reasonable to think this might be quickly remedied. Any completely *outré* situation would, precisely because it is so novel, preferentially command attention and thus cognitive resources in the manner outlined.

Likewise, to the extent that the current environment does differ from that of our evolutionary past, one would expect characteristic breakdowns of the emotion-based heuristics to result. That is, in much the same manner in which Kahneman and Tversky's (1972) findings point to the existence of particular (evolutionary based) cognitive biases, it would not be unreasonable to suggest the same with respect to the emotions. However, while an interesting and highly plausible conjecture (that the *current* environment should result in some characteristic breakdowns of the emotion-based heuristics proposed), there are no studies (analogous to those undertaken by Kahneman & Tversky) directly considering this question. Discerning those modern-day situations in which the emotion-based heuristic rules outlined are less beneficial, or even detrimental, would be, of course, of great interest. So doing would also go some distance in helping us to better understand the extent and limits of the normative "goodness" of the emotion-based heuristics discussed.

ASSEMBLING A MODEL: GLOBAL WORKSPACE THEORY AND SOMATIC MARKERS

Damasio and Bechara's account readily lends itself to incorporation into the existing Global Workspace model of cognition discussed in Chapter 3. Sketching this will, I think, help to clarify our subsequent discussion of how emotion might help us to understand how it is that we contend with some instances of the frame problem that arise in practical reason and decision-making.

By integrating the two accounts, a picture emerges of how emotion might interact with and facilitate practical reason and decision-making in ways relevant to contending with some of the frame problems that emerge in these domains. And so, in addition to the set of modularly realized consumer systems proposed by Baars' account, each of which undertakes local computational transformations on representations in the global workspace and release these back into the workspace whereby other systems can (if, drawing upon Barrett's metaphor, they accept representations of that "shape") engage in further processing, there are the three additional sub-systems outlined by Damasio and Bechara's model.

And so, given the role occupied by emotion in motivational and attentional direction, emotion might initially guide or act as a rudder on cognition by, in effect, releasing into the workspace representations of some problematic scene. So doing might effectively drive out or put on temporary hold whatever was in the workspace and thus whatever it was that we were thinking about including whatever immediate ends we might have been pursuing. By means of processes undertaken by manifold perceptual consumer sub-systems, objects or situational features might come to be recognized. Images of the refined scene would, in turn, be relayed back for another round of appraisal in the manner suggested by LeDoux's account. This refined image will, in turn, either serve as an inducer or not. If so, the refined image will be attended/relayed to the global workspace and tagged with the feeling induced. So doing would effectively mark the desirability or problematicity of the image of that scene. *Via*, the motivational influences outlined by Damasio, this highlighting mechanism would serve to help direct the ends and thus the lines of inquiry pursued. This, in turn, would (re)direct attentional resources on the problem posed thus maintaining it in the working memory of the global workspace. Once the problematicity of the situation is fixed and motivational and attentional resources directed upon it (thus maintaining the problem in the workspace), manifold cognitive processes (*i.e.,* other consumer sub-systems) would, once again, be enlisted in the manner proposed by Baars and Carruthers.

If we assume that the situation, in fact, is an undesirable one (*i.e.,* it poses a problem for the system), manifold processes engaged in planning such as those set out by Carruthers (2006), Milner & Goodale (2006) and Kosslyn & Koenig (1996) would be enlisted. During deliberation about what one is to do, representations of the response options are, once retrieved or generated (as Damasio 1994 and Forgas 2000 suggest), made available to the global workspace.[77] *Via* the modularly realized operations of sensory-specific cortical regions, outlined by Milner & Goodale (2006) and Kosslyn & Koenig (1996) and elaborated by Carruthers (2006), quasi-perceptual images of the content of these representations are constructed and internally rehearsed. While we need not go into the details of the workings of this mechanism here, the idea is that the content of propositionally formulated "thoughts" come to be, *via* the interface between

[77] Incorporating Baars, Milner & Goodale (2006) and/or Carruthers' (2006) proposal with respect to how a modular system might go about generating plans and response options, would allow for the workings of this stage of the deliberative process to be more fully developed. Providing a detailed account of Carruthers' proposal and incorporating this into the model as well will, for the sake of brevity, not be undertaken here.

cortical regions (implicated in planning and executive function) and early sensory areas, "translated" into auditory/visual/olfactory etc. images – (*i.e.,* multiple sense modality "pictures.") That is, the same modularly realized operations that are engaged in perceptual processing are exploited, Milner & Goodale and Kosslyn & Koenig argue, in effect, in reverse to reconstruct quasi-perceptual images of the content of entertained "thoughts." The amygdala, we know, receives input from the thalamus and sensory specific cortical areas (*e.g.,* visual, auditory, olfactory areas). And so, *via* reconstruction in sensory specific cortical regions and the interface with these to the amygdala (*i.e., via* LeDoux's "high road" pathway) suppositionally entertained response options (*i.e.,* "thoughts") come to be emotionally appraised. If the quasi-perceptual image (of the content of the representation entertained) serves as an inducer or trigger (of a primary emotion) then an emotional response will be induced and area SM1 will get busy generating a feeling-map of the physiological alterations that unfold. This feeling is then juxtaposed with the image (of the content of the representation entertained) and the "scene" comes to be somatically marked. So doing transitively fixes the desirability of the response option or resulting "scene" which in turn, given the innate set of dispositions/*taxes* for particular states of affairs, automatically motivates the pursuit or avoidance of that particular option, end, or line of inquiry. And so, by means of these operations, internally rehearsed suppositionally entertained plans for contending with the current posed problem are appraised with respect to the feeling likely to result.

There are a number of possibilities for how the heuristic stopping and decision rules proposed by the somatic marker account might be incorporated into the global workspace framework. Drawing upon Barrett's enzyme metaphor, one possibility is that somatic/feeling "tagging" might affix (metaphorical) tags to representations.

Negative feeling markers might be of such a metaphorical "shape" that they make the representation to which they are affixed no longer process-able by other consumer subsystems – in effect resulting in these being ignored. Similarly, a positive tag would do quite the opposite, allowing particular consumer subsystems to continue processing the representation and do so with the additional information provided by the tag. Or, more in keeping with Damasio's line, the heuristic culling and highlighting operation might be undertaken by some subsystem that actively scans the workspace looking for negatively marked representations, removing from the set any that it finds. So doing would leave only those sufficiently positively marked options remaining, highlighting them in this way. Likewise, *via* both the feeling marking apparatus and the feeling preference hierarchy, the desirability of options would be "by dint of juxtaposition" transitively fixed, as well. We are, *via* the *taxic* dispositions outlined by Rolls and Damasio's account, motivated to pursue those positively highlighted options the objects/outcomes of which are preferably marked (*i.e.,* those outcome states of affairs that are desirable), while avoiding those the outcomes of which are negatively marked (*i.e.,* those states of affairs that are undesirable). Alternatively, and more in keeping with LeDoux's model of emotional appraisal and Milner & Goodale's model of action-plan construction, we might suppose that the output of those sub-systems engaged in response option generation might incorporate a cycle of emotional appraisal into their operations. And so, each response option or plan generated would be, before it is released to the global workspace for further processing, emotionally appraised. Those that are or come to be negatively tagged would not be released.

While a number of possibilities can be offered, it is ultimately an empirical matter as to how these particular heuristics are implemented. Similarly, the precise nature of these heuristics is also unclear. And so, for example, with respect to the stopping/culling operation, it may be that all options tagged with feelings below some threshold are culled. It may be that only the "worst" (*i.e.,* the least preferably marked with respect to the feeling hierarchy – or most undesirable) options are culled. Likewise, with respect to the heuristic decision-rule proposed by Damasio, it may be that the "best" option (as given by the feeling preference hierarchy – the most desirable) is chosen. It might be that the first positively marked option is selected or that the first positively marked option above some threshold is selected (some form of feeling *satisficing*). This too is an empirical matter. Damasio and Bechara make no conjecture and the IGT findings are not instructive in this regard. And so, while we have good reason to think the operations outlined by Damasio to be readily incorporable into the existing global workspace framework, a complete model cannot, given the available evidence, be precisely specified. Discerning the particular stopping/decision rules employed would allow a more complete model to be set out. Seemingly, some progress might be made on this question by testing subjects on a variant of the IGT in which all options are "good" but some are better than others.

SOME EARLY RESULTS FROM ROBOTICS

Additional support for the claim that emotion, *via* the mechanisms outlined by LeDoux, Damasio & Bechara and Rolls, may help us to contend with some of the frame problems that arise in practical reason and decision-making can be found in some very recent work in the field of robotics (Canamero, 2003; Picard, 1997, 2002; Slomin 2001; Petta, 2003). Germane to this, Shanahan (2006) constructed a robotic system that incorporated rudimentary emotional processes, specifically a scaled-down version of the body-loop appraisal device, into the framework of Baars' global workspace model.

Shanahan's robot possessed an amygdala-analogue (based upon a network model of amygdala function proposed by Canamero) that, following a set of learning/training trails, set about appraising the "emotional" significance of objects in its environment (a simple box-world type environment). This amygdala-analogue was incorporated into the global workspace architecture along the lines set out by LeDoux and Bechara and Damasio. First, visual representations of the environment are relayed to the amygdala-analogue and its appraisals are exploited to highlight particular features. In so doing and, *via* the framework outlined by global workspace architecture, a set of processes aimed at generating an initial (rather reflexive) response option are enlisted. In keeping with Damasio's model, instead of merely reactively undertaking this action, the system holds this action-plan "on veto" until a body-loop appraisal-analogue is undertaken. And so, instead of merely reacting to its environment, initial responses or action-plans are stayed until the proposed action is "internally rehearsed" and an appraisal of this plan - considered *suppositionally* - is undertaken. During deliberation about what it is to do, the robot exploits *anticipated* emotion to guide response selection in the manner outlined by the somatic marker model. And so, before acting, Shanahan's robot considers what it will likely be like "emotionally" to have undertaken the plan under consideration. And like Damasio's model, those plans the suppositional outcome of which come to be (*via* body-loop appraisal by the amygdala-analogue) negatively marked are, as undesirable, rejected from further consideration (*i.e.,* the motivational veto is maintained) while those positively marked are highlighted.

Clearly, the robot's cognitive/emotional architecture is quite simplified with respect to both Baars framework and Bechara and Damasio's model. And so, for example, the integrative role of the VMF is either entirely absent or subsumed by the amygdala itself – a departure from Damasio's proposal in either case. Likewise, neither feeling nor somatic feedback play a role. Furthermore, while LeDoux's apparatus of conditioned learning in incorporated into the amygdala-analogue, none of the learning mechanisms proposed by Damasio's model are included, thus preventing emotion-induced feeling and thus feeling-markers from being integrated and updated on-the-fly following interactions with environment. The robot's test environment is quite simple, as well – far simpler than any "real-world" situation. It is, however, worth noting that this environment is no simpler than the blocks-world environments that gave rise initially to the frame problem. These caveats noted, that this robot can expeditiously and successfully seek out "good" objects while avoiding the "bad" is a quite promising result. Specifically, by incorporating the machinery of emotional appraisal and a body-loop analogue device that appraises suppositionally considered plans and exploits emotions in an anticipatory fashion to guide response selection as Damasio and Bechara's model suggests, the robot is capable of successfully contending with its environment without effecting an exhaustive search of the problem space. And so, while fully acknowledging the toy status of this result, that a robotic system incorporating emotion (*i.e.,* a scaled-down version of Damasio & Bechara's proposal) into its decision-making or response-selection machinery can effectively and expeditiously decide what to do seems a promising early result.

SOME OPEN QUESTIONS

Having outlined a number of ways that emotion might help us contend with some of the frame problems that arise in practical reason I will turn now to consider a few open questions. Specifically, I have not explicitly considered how emotion might help us contend with the search and inference instances of the problem (*i.e.,* finding the "right stuff" in memory and making the right inferences without having to search/consider the totality of memory/the problem space).

Earlier, in considering a possible objection to the somatic marker account, I argued that emotion facilitates the encoding of factual (declarative/episodic) material to memory. This faciliatory effect of emotion on encoding is well documented and we have a plausible neurological mechanism (Cahill & McGaugh's *epinephrine hypothesis*) by which such influences are realized. And so, that the influence of emotion on encoding should be explicable in terms of the operations of a rather simple heuristic encoding rule – preferentially remember those events that are attended by an emotional response – seems to be a plausible conjecture.

If so, emotion might help us in contending with the memory-encoding instance of the frame problem.[78] Any finite system capable of remembering at all is faced, given that it cannot remember everything, with the puzzle of determining what it will remember. Arriving at a rational conclusion about this is, for the reasons outlined in Chapter 3, untenable. And so, I argued it is plausible to suggests that some heuristic or set of heuristics should be relied upon to direct the kinds of material preferentially encoded to memory. Emotional significance does appear to be one such heuristic.

Compared to the rational alternative for contending with the encoding problem, whereby the system must arrive at a rational conclusion about which things it should remember, which requires that a suitable standard for "remember-worthiness" be rationally discerned, the simple emotion-based heuristic rule should be far more expeditious. And so, with respect to the tractability horn of the dilemma posed by the

[78] Clearly, I am not suggesting that emotion should be the *only* heuristic at play guiding memory encoding.

frame problem, a system relying on such a simple heuristic should fare much better than a rational one.

Prima facie it is reasonable to suggest that preferentially remembering where, when and in what circumstances (*i.e.,* the episode or event) in one's past one has encountered particular dangers, threats, rewards, punishments, surprises and the like should afford some benefit to practical reason and decision-making. That is, relying upon this encoding heuristic whereby the details of an emotionally significant event are preferentially remembered, should be beneficial to, at the very least, the ends of survival/fitness. Preferentially remembering where, when and details of the scene in which one was mugged seems to be, at least prima facie, a good strategy – if for no other reason than to help us to avoid going down *that* dark alley again in the future.

However, we also have reason to think that relying upon this heuristic in at least some cases might be disadvantageous. Simply put, there appear to be some events/episodes that are preferentially remembered because they are emotionally significant and which are either (i) irrelevant or (ii) we would be better off not remembering. For example, with respect to the former, "flashbulb memories" or the "JFK effect" (or the Space Shuttle Challenger or 9/11 effects for those of younger generations) is well documented. (Schacter, 1996; 2001) Nearly every American alive at the time remembers precise details about the moment in which they learned that Kennedy had been assassinated. It is, however, unclear how preferentially remembering *this* should be relevant to practical reason or decision-making generally. Likewise, studies of posttraumatic stress disorder provide suitable examples of instances in which relying upon this encoding heuristic might be detrimental. Given the current conjecture that PTSD patient's troubles are the result of their incessant rehearsal of the memory of a particularly emotionally significant event, there appear to be cases in which the facilitatory effect of emotion on memory is detrimental. PTSD patients, it is not unreasonable to suggest, would fare better were they capable of forgetting (*i.e.*, or at least not preferentially remembering in the first place) the details of the particular traumatic/emotionally significant event.

Intuitively, it seems quite likely that relying on this encoding heuristic should be, at least on balance, more beneficial than detrimental. At least prima facie it appears likely that relying on the emotion-based encoding heuristic outlined should be more beneficial than relying upon a strategy that randomly (*i.e.*, at chance) preferentially encodes material to memory. That an heuristic succeeds (*i.e.*, is accurate or otherwise normatively "good") at a rate only slightly better than chance is sufficient to secure the modest point that that heuristic should *help* in contending with the frame problem – assuming, of course, that it is also expeditious. However, while relying upon this simple encoding heuristic is likely to be, on balance, more beneficial than not, given the lack of available empirical evidence in support of this (modest) normative conjecture, no claim is warranted.

And so, while emotion directly influences the kind of episodic/declarative material encoded preferentially to memory, which in turn suggests that emotion might be brought to bear to explain how the process of selective memory encoding might be expedited, there is no empirical evidence directly bearing on the normative question. Relying on the heuristic outlined might be, with respect to accuracy, on balance beneficial, detrimental or a normative wash (on balance neither helpful nor harmful).

Damasio and Bechara's discussion of the nature of fact-feeling sets lends itself naturally to the idea that just as feelings can be recovered by attending to particular facts, so too might facts be preferentially retrieved by attending to feeling. This, when taken in conjunction with the ample empirical evidence that emotion directly influences the kinds of material retrieved from memory[79] suggests that emotion might be heuristically exploited to expedite search, perhaps by preferentially enlisting particular search heuristics aimed at bringing emotionally congruent material to bear.[80] Given the evidence, it is quite reasonable to think that a subject's emotion state is heuristically exploited to direct search of memory. It is also quite reasonable to think that such an heuristic look-up device would be far more expeditious than one relying upon the exhaustive consideration of the contents of memory. However, that this heuristic should afford a benefit to practical reason and decision-making is unclear. That is, it seems quite plausible to suggest that bringing to bear (recalling from memory) those situations in which one contended (successful or unsuccessfully) with a similar problem kind (*i.e.,* an emotionally analogous situation) during deliberation should expedite and afford a normative benefit to practical reason. However, it also seems quite likely that the set of material preferentially brought to bear by this heuristic *alone* might be overly broad. That is, that at least some of the material that is brought to bear should be relevant to some current problem seems probable, but that some (if not much more) irrelevant

[79] Bower, Gilligan & Monteiro 1981; Bower, 1981; Gilligan 1982; Teasdale & Fogarty 1979; Ingram, 1984; Teasdale, 1983; Teasdale & Clark, 1982; Teasdale & Russel, 1983; Bower & Mayer, 1985, Blaney, 1986; Eich, 1995; Eich & Macaulay, 1989; Fiedler, 1990; Forgas, 1991; Forgas 1993; Forgas 1995; Forgas & Bower, 1987; Eich & Macaulay, 2000; Eich, Macaulay & Ryan, 1994; Eich & Metcalf 1989; Niedenthal & Stetterlund, 1994

[80] The *affective state dependency effect*, holds that emotional states function as context cues facilitating or inhibiting recall accuracy in much the same manner as situational or environmental cues have long been known to do. Simply put, subjects in whom a particular emotional state ("happy," for example) has been induced will preferentially retrieve from memory (recall) emotionally congruent memories ("happy" ones). Furthermore, while the subject's current emotional state facilitate the retrieval of emotionally like material from memory, it is also found to interfere with or inhibit the retrieval of emotional incongruent material. And so, not only will a "happy" subject preferentially retrieve "happy" memories (and do so faster), they will have difficulty in recalling, for example, "sad" memories.

material should be brought to bear seems likely as well. And so, while we have reason to think both (i) that emotion is likely heuristically exploited to help direct memory search (and thus to help limit in part the material brought to bear) during deliberation, and (ii) that such an heuristic should be far more expeditious than the rational alternative, it is less clear that relying upon it should afford a normative benefit to practical reason.

And so, that, *via* its influences on memory search, emotion should play some further heuristic role in practical reasoning seems likely. However, there is insufficient evidence to support the claim, at this time, that such an heuristic would adequately satisfy the normative horn of the puzzle. So doing would require a much deeper understanding of the structure, workings and organization of human memory as well as an account of how these operations interact with those of practical reason. In addition, much more would need to be learned about the nature and specificity of the particular search heuristic proposed. These are decidedly busy areas of ongoing research.

Likewise, with respect to the inference problem it is reasonable to think that emotion should be exploited to preferentially enlist particular processing strategies. That VMF damage and amygdala damage results in impairment of practical reason (*i.e.*, both groups are impaired in this regard) supports this conjecture. That emotion influences the kinds of processing strategies and "cognitive styles" employed is also well established.[81] It is not unreasonable to think that, with respect to the tractability horn of

[81] Schwarz & Clore, 1983 & 1988; Isen, Johnson, Mertz & Robinson 1985; Isen 1987 & 1993; Isen & Daubman, 1984; Isen, Niedenthal & Cantor, 1992; Isen & Kahn, 1993; Isen, Daubman & Nowicki, 1987; Murray, Sujan, Hart & Sujan 1990; Arkes, Herren & Isen, 1988; Dunn & Wilson, 1990; Isen & Geva, 1987; Isen 1988; Isen & Patrick, 1983; Isen, Pratkanis, Slovic & Slovic 1984; Berg 1991; Schwarz, Bless & Bohner 1991; Carnevale & Isen 1986; Nygren, Isen, Taylor & Dulin 1996; Johnson & Tversky 1983; Mayer, Gaschke, Braverman & Evans 1992; Forgas & Bower 1987; Forgas 2001.

the puzzle, relying on an emotion-based heuristic should be more expeditious than the rational alternative of exhaustive inference generation and consideration. However, that such an heuristic would satisfy the normative horn is, at this time, unclear. And so, while emotion does appear to automatically influence cognitive "style" and the kind of processing strategies preferentially employed (Isen, 1993; Forgas, 2001) which in turn suggests that emotion might be brought to bear to explain how such processes might be expedited, there is no direct evidence bearing on the normative question.

That the linkages between particular emotions and particular learned and unlearned inferential strategies should be amenable to evolutionary explanation seems quite likely, as Panksepp (1995; 1998) and Cosmides & Tooby (1990) argue. Establishing the normative component however would, drawing upon Gigerenzer's discussion, require our having reason for thinking emotion (or emotion induced feeling) to be a suitably ecological valid cue with respect to the selection of inferential strategies. Since there is no direct evidence bearing on this particular question, beyond providing some reason why it is likely that emotion should be relevant to this instance of the puzzle, to not overstep the available evidence, I will refrain from making any specific claim. With respect to this, we must then wait for studies to be undertaken that examine both the nature of the particular cognitive strategies employed and the question of whether the particular cognitive/inferential strategies ("styles") preferentially enlisted when agents are, for example, "afraid," help agents in better contending with the particular threatening situation or problem posed.

With respect to the planning problem, two points bear mention. We have little reason to think emotion to play a direct role, other than perhaps in motivation and end-selection, in the construction of short-term plans. (*e.g.,* it is unlikely that in generating a simple plan to pick up a glass of water, for instance, that emotion should feature heavily). That damage to the amygdala and area VMF leaves this capacity intact establishes this. That the anecdotal evidence suggests that such patients are capable of generating and acting upon short-term plans (*i.e.,* "myopic") plans is clear, as well. However, we do have reason to think that emotion might play a role in motivation and meta-planning (problem setting, sequencing and ends selection). If so, it is plausible to suggest that emotion might help to frame what we think about (what problems we consider and thus what plans we go about generating) and how we think about it. As such, emotion might be relevant to the processes whereby complex plans come to be generated. Insofar as the construction of a complex plan requires the setting of intermediate goals or sub-goals (*i.e.,* the setting of sub-problems along the way), suggests, given its role in meta-planning, that emotion might play some role here as well. That is, if the construction of complex non-trivial (*i.e.,* multi-stage) plans, requires one to engage in a kind of meta-planning (*i.e.,* to arrive at conclusions about which sub-problems are relevant and thus which sub-goals to pursue), then emotion might plausibly be brought to bear on this question. Furthermore, it would not be unreasonable to think that the influences of emotion on attention and motivation should play some role in the processes of complex plan generation, as well. That is, by directing attentional resources inwardly upon those features of a suppositionally entertained quasi-perceptual images "scenes" under consideration (in the manner outlined by Kosslyn & Koenig 1996, perhaps) that are emotionally (and thus motivationally) significant, emotion might be relied upon to help us locate some of the potential "problems" that a sub-plan might generate. *Via* motivational influences and the culling and highlighting operation discussed, then, such sub-plans might come to be pruned or modified in light of this information (*i.e.,* their desirability).

That cycles of emotional appraisal should be undertaken during the operations of complex plan construction (*e.g.,* to automatically cull particular sub-plans from further consideration) seems plausible, given the mechanics of both Damasio and Bechara's model and those outlined by Milner & Goodale. Forgas (1993, 2001), in outlining his *affect infusion model*, makes a similar conjecture. Unfortunately, aside from providing reason for thinking the influences of emotion to increase as the novelty and complexity of the problem posed increases (*i.e.,* affect highly "infuses" complex problem solving while only minimally influencing the operations by which we contend with trivial/common problem kinds), Forgas provides little discussion as to *how* this is realized and undertaken.

However, even if emotion does feature in this way in the operations of non-trivial planning, it is still entirely unclear that this should be a sufficiently normative "good" strategy. That we have reason to think that the ecological validity of feeling cues should increase over time with interactions with a stable environments suggests that if emotion plays a role in the generation of complex plans in the manner outlined, then it is likely to - over time - come to be beneficial strategy. That is, if it plays this role, then relying on emotion should come to be a normatively good strategy with respect to the culling and highlighting of sub-goals and sub-plans. While a highly plausible conjecture, there is currently no empirical evidence bearing on the question. That at least some of the behavioral deficits exhibited by patients with damage to area VMF or the amygdala might be attributable to an inability to engage in complex or multi-stage planning (*e.g.,* intermediate problem fixing and sub-goal selection) seems plausible.

Specifically, that patients are prone to both impulsivity and perseverativeness during real-world planning and problem-solving, suggests that they may be impaired in recognizing sub-problems and/or in fixing sub-goals during a complex planning task. That the emotional tagging heuristic is of normative value, given earlier discussion, seems quite plausible. Likewise, that this heuristic should be of value in complex planning seems a quite plausible suggestion. However, for a number of reasons discussed earlier, while we have reason for thinking it a likely a normatively "good" strategy, there is no direct empirical evidence bearing on this specific question (*i.e.,* the question of how normatively "good" the influences of emotion on the operations of planning and plan construction might be). And so, as an empirical matter, we must wait on the evidence before the issue can be fully resolved.

I have also purposefully focused on instances of the frame problem that arise in practical reason, leaving theoretical inference and belief-formation largely unconsidered. There is, however, reason for thinking that emotion might be relevant to at least one instance of the frame problem that arises in this domain.

Drawing upon discussion by DeSousa, Dennett (1981) argues that our doxastic processes are involved in two distinct kinds of activities. We are, claims Dennett, when we talk of the activity of belief-fixing really speaking of two mistakenly conflated kinds of mental activities. Sometimes we are referring to the activity of *belief*-fixing (proper) and sometimes we are referring to the activity of *opinion* fixing. Concerning the distinction, Dennett explains, "animals have *beliefs* about this and that, but they don't have *opinions*. They don't have opinions because they don't *assent*. Making up your mind is coming to have an opinion." (Dennett, 1981 p. 304) Two further points are raised by Dennett in differentiating these activities. First, unlike in the case of belief-fixing, coming to have an opinion about something is best understood in terms of a "commitment to" or a "'wager on" the likely truth of some entertained sentence. Second, unlike in the case of belief-fixation, all instances of opinion-fixing are,

> Species of judgment, and while such judgments arise from beliefs and are ultimately to be explained by one's beliefs, such judgments themselves are *not* beliefs … but *acts*, and these acts initiate states that are also not states of belief, but something rather like commitment. [T]his decision [to assent] is first of all an act, an exemplary case of doing something … what is important is just that it is a choice point that terminates a process of deliberation or consideration that is not algorithmic, but at best heuristic. At some point we just stop deliberating … and leap. (Dennett, 1981 p. 303)

If, as Dennett suggests, opinion fixing is an *act* – a something that we *do*, then the activity of assent should be considered a practical matter. Since we have reason for thinking emotion to help us contend with some of the frame problem instances arising in practical reason and practical decision-making, if assenting is an act (that is, if part of our doxastic processes are practical) then it would not be unreasonable to think that emotion might help us to understand how we contend with at least some of the problems that arise in that domain as well. That relying on an emotion-based heuristic to help us contend with the decision-problem of assent in theoretical reason would be expeditious is apparent. That emotion influences judgment is, as noted earlier, well established in the literature. That relying on such an heuristic is sufficiently normatively "good" is, however, unclear. And so, while it is quite reasonable to think that emotion should play a considerable role in the operations of belief-fixation, given the lack of evidence (with regard to the normative horn or aspect of the puzzle), no conclusion is warranted.

CONCLUSION

I have set out in this chapter to see how far the consideration of emotion might take us in helping us to understand how we might contend with some of the frame problems that arise in practical reason and decision-making. I have not, of course, claimed that emotion alone solves outright the frame the problem. Rather, I have suggested that we have reason to think that some light might be shed on the question of how we contend with some instances of the frame problem arising in practical reason and decision-making by bringing emotion to bear. And so, the aim of this chapter has been a modest one – to provide reason for thinking that some progress, with respect to aims of both understanding and eventually modeling the operations of practical reason and decision-making, might be made by bringing emotion to bear on the issue.

To this end, I suggested that emotion might help us quite directly in contending with the attentional direction instance of the problem. Given that any finite system's attentional resources must be limited, it is reasonable to think that this instance of the problem should be contended with heuristically. Emotional significance appears to be *one* such heuristic, likely operating in tandem or competition with a number of others. Specifically, drawing upon Damasio's account, we have reason to think that the emotional tagging or marking operation directly influences motivation, which in turn helps to focus attention upon emotionally (and hence motivationally) significant material, by increasing/decreasing (from neutral) the desirability of (images of) particular states of affairs. In addition to the account provided by Damasio, we also have reason for thinking that agents' current emotional states directly influence the kinds of material that are preferentially attended. Since the operations of emotional appraisal are automated, autonomous and plausibly modularly realized, any system that exploits the information provided by emotional appraisal to direct attention should be far more expeditious than a rational (in Fodor's sense) one. We also have reason for thinking that relying upon this attentional direction heuristic (*i.e.,* attend preferentially to that material that is emotionally and hence motivationally significant) should be a sufficiently normatively "good" strategy to satisfy the second horn of the dilemma posed by the frame problem.

By directly influencing motivation and attention, we also have reason for thinking that emotion might fruitfully be brought to bear on the problem-sequencing instance of the puzzle. Specifically, I argued that by directing attention upon emotionally significant material and by ratcheting up (from neutral) the desirability of particular images (of states of affairs and options), cognitive resources would come to be preferentially enlisted and brought to bear. So doing, it is reasonable to suggest, would result not only in a shift in motivation (*i.e.,* the ends pursued) and attention but also a redirection of cognitive resources – that is, we would come to preferentially think about the emotionally (and motivationally) pressing problem posed. This in turn, instantiates a simple problem-sequencing heuristic: contend with emotionally significant situations first. Taken in conjunction with the feeling preference hierarchy/ *taxic* dispositions outlined by Damasio and Rolls respectively, the automated operations of emotional appraisal might help in fixing the order in which problems are to be contended. That is, it would appear that automated emotional appraisals are exploited to direct attention, motivation and cognitive resources upon the most emotionally (and thus motivationally) "pressing" problems first (*i.e.,* the most un/desirable state of affairs), then the next and so on. Such a sequencing heuristic, since it relies upon only one cue – the transitive marking of the desirability of states of affairs by feeling-image juxtaposition – should be far less computationally burdensome and thus more far more expeditious that the rational (in Fodor's sense) alternative. With respect to the second horn of the puzzle, we have reason for thinking that relying upon this heuristic to be sufficiently normatively good, as well. Likewise, relying on this heuristic should be a far normatively better strategy than the rational alternative that would fail to arrive at any conclusion at all about how its problems should be sequenced.

Moreover, emotion appears to help in contending with the meta-planning instances of the frame problem. Specifically, in addition to the assistance offered in this regard by influencing attentional direction and focus, somatic markers also appear to play a direct role in warning us that our current line of inquiry or planning is undesirable (*e.g.,* fruitless or indulgent). By means of the influences of this automated warning system motivational goals are shifted thus drawing attention away from the particular planning task or line of inquiry with which we are currently engaged, redirecting them upon the broader problem of whether we should be engaged in contending with this particular problem at all and, if so, in this way. Furthermore, once this (meta-planning) end is set, this problem attended, and cognitive resources are directed upon it, plans for contending with that problem might come to be generated. Whatever options are generated (*via* the mechanisms proposed by Baars) would then be subject to additional rounds of emotional appraisal in the manner outlined by the somatic marker hypothesis. Since emotion-induced feeling is exploited as a noncompensatory (*i.e.,* singular) cue to both guide the ends, goals and lines of inquiry pursued and to (re)direct attentional and cognitive resources in the manner outlined, relying on this rather simple heuristic for contending with the meta-planning instance of the puzzle should be far less computationally intensive and thus far more expeditious than the rational alternative. It should be, for the reasons discussed, sufficiently normatively "good" to satisfy the second horn of the dilemma, as well. Again, relying on such an heuristic should be, at the very least, a normatively better strategy than the rational one since no rational system could be relied upon to arrive at any conclusion at all with respect to the ends that it should pursue.

Similarly, I suggested that emotion might serve a role as a problem-setting heuristic that, in effect, helps alert us to the fact that some situation (real or imagined) is, as emotionally and thus motivationally significant (*i.e.,* un/preferably tagged and thus un/desirable), one with which we should contend. And so, the automated emotional appraisal device and its influences on motivation, attentional and cognitive direction might provide *one* means by which to explain how we might contend with the end/goal setting instance of the problem.

Consideration of emotion, I argued, might help us to understand how we might contend with the Hamlet's problem instance of the puzzle by serving an automated and heuristic "stopping rule" function in deliberation. As a noncompensatory strategy, this stopping rule should be far more expeditious than the rational alternative. We also have reason to think it a strategy that adequately satisfies that normative horn of the puzzle. Along similar lines, I suggested that emotion might helps us to contend with the decision-problem instance of the frame problem by serving as an automated and heuristic decision-rule. Compared to the rational alternative, the noncompensatory (*i.e.,* single cue) strategy outlined here should be far less computationally burdensome and far more expeditious. Likewise, with respect to the normative aspect of the puzzle, we have reason to think this to be sufficiently normatively "good."

Finally, I considered a set of open questions, conjecturing that while consideration of emotion should, for a number of reasons, hold promise for our coming to understand how we might contend with these instances of problem, with respect to the normative issue, the current empirical evidence is insufficient to warrant any specific claim.

CONCLUSION

I began, in Chapter 1, by considering the frame problem. Having first set out the original and formal version of the puzzle that arose in artificial intelligence, I next provided an overview of some of the philosophical interpretations of the problem. Following this, I set out the underlying structure of the puzzle, focusing on explaining and clarifying the nature of the dual horns of the dilemma that all instances of the problem pose. From this discussion, I concluded that "the" frame problem is not reducible to any one particular problem, but instead is a constellation of related problem instances.

Having set out the puzzle's underlying structure and a number of instances of it (*e.g.,* attentional direction, memory encoding/search, inference, ends-selection, meta-planning, problem-sequencing and Hamlet's problem), I turned in Chapter 2 to consider in greater detail Fodor's version of the problem and its application to computational psychology. In presenting Fodor's arguments for the pessimistic conclusion, I mark and set about disentangling a number of distinct argument strains.

Following this exegesis and clarification, I focused in Chapter 3 on evaluating Fodor's argument in support of the pessimistic conclusion. Specifically, I argued that Fodor's arguments rely upon an unsatisfiable and untenable *normative* principle of rationality. Having set out this principle and argued that no finite and physically realized system could satisfy it, I turn next to consider Fodor's *descriptive* claim. Specifically, Fodor contends that because we *at least sometimes* arrive at conclusion rationally – that is, that we at least sometimes satisfy the normative rationality condition that he sets out – we are (at least sometimes) instantiations of a rational system. Since, he continues, rational systems cannot be modeled, it follows straightaway that our minds cannot be modeled either. I argue, however, that we have no reason for thinking that *we* ever arrive at conclusions rationally (in Fodor's sense) and hence that it is false that we are (ever) instantiations of a rational system. Since it is this descriptive claim that directly undergirds and motivates Fodor's argument for the pessimistic conclusion, and we have no reason for thinking it to be true, (and reason for thinking it false) we have, I argued, no reason for concluding that our doxastic processes are as a matter of principle unmodelable.

Rather, I argued that it is much more reasonable to conclude that we should reject Fodor's normative rationality principle in light of its unsatisfiability, replacing it instead by a weaker one that is at least in principle satisfiable by the likes of *us*. *Any* weakening of Fodor's normative rationality principle, I argued, effectively undermines the set of in principle arguments offered by Fodor against the massive modularity and heuristics approaches. And so I conclude, by weakening the normative rationality principle, as it appears that we must, it no longer follows that one must reject *tout court* "irrational" processes like heuristics and modules. Since there is no reason to think that modules and heuristics, though "irrational" (in Fodor's sense), should be incapable of satisfying a suitably weakened version of the normative rationality principle and these operations are modelable (by design or by definition), we have no reason for thinking that these approaches should be untenable.

Having argued that both the massive modularity and heuristics approaches should not be rejected because they fail to satisfy Fodor's (unsatisfiable) normative rationality principle, I next considered an alternative model – namely Baars' Global Workspace framework. Following an overview of Baars' proposal, I considered how bringing this approach to bear might help us in understanding how a system might contend with some of the particular engineering (*i.e.,* in practice as opposed to in principle) challenges set out by Fodor. To this end, I considered Gigerenzer's "fast and frugal" heuristics approach and Barrett's discussion of modules as metaphorical enzymatic devices. These, I suggested, when integrated into the general framework of the Global Workspace model, might provide a plausible account (and an alternative to that proposed by Fodor) of how at least some of the interesting activities of mind might come to be modeled.

In Chapters 4 and 5, I turned attention to considering some instances of the problem that arise in practical reason and decision-making. Specifically, I consider the question of how much progress might be made in our understanding of these activities by bringing emotion to bear on the frame problems that arise in this domain. I noted at the outset of this discussion two principal prima facie objections or challenges that must be met before it would be reasonable to think that emotion should be relevant to the frame problems arising in this domain.

I considered, first, a set of related prima facie objections to the very idea of bringing emotion to bear on the problem. The "irrelevance" objection, I argued, is grounded upon a particular view of what the emotions are – specifically, one in which *all* emotions are taken to be either belief-identical or belief-dependent. And so, the objection runs, if *all* emotions are belief-dependent (or belief-identical) and if, as Fodor suggests, the operations of belief-fixation are infected with/by frame problems, then bringing emotion to bear on the problems that arise in any other domain (*i.e.,* practical reason and decision-making) would serve *only* (and at best) to bring an operation that is already unmodelable, tainted, and stricken by the frame problem to bear on those instances of the problem that arise elsewhere. Therefore, the objection concludes, the very idea of bringing emotion to bear on the frame problem (in any domain) is viciously circular, an obvious non-starter, and an otherwise implausible proposal.

Following discussion aimed at eliminating a potential source of confusion with respect to the term "cognitivism," I provided an overview of the propositional attitude approach to emotion and then considered a representative sample of the empirical evidence relied upon to support the cognitivist account of emotion. Having outlined this, I next argued that given both the inability of the cognitivist account to contend with a set of standard puzzles and the problematic nature of much of the empirical evidence relied upon, we have no reason for thinking that *all* emotion is belief-dependent as the cognitivists claim. And so, while it is possible (and perhaps even quite likely) that *some* of what the emotions are are belief-dependent, we have, I concluded, no reason for thinking them all to be. With respect to the "irrevelance" objection, if at least some of what the emotions are are belief-*in*dependent, then there would be no reason to think that these operations, when brought to bear, should necessarily import the belief-fixing frame problems into other domains.

I considered next the automated appraisal or "non-cognitivist" account of emotion, concluding that while it is likely incomplete as an account of what the emotions are, it does provide a partial response to this question. And so, following a review of the evidence, I concluded that *at least some* emotions appear to be realized by automated, autonomous (*i.e.,* belief-independent or non-cognitive) and plausibly modular learning and appraisal operations. Given the simplicity of the learning and appraisal operations outlined by LeDoux, Damasio & Bechara, Ekman and Zajonc and their apparent modular nature, we have reason for thinking that these operations should be amenable to modeling. We have then, I concluded, no reason to think that bringing these to bear should serve only to import a prior frame-problem-infected operation into this other domain. Rather, since at least some of what the emotions are are undertaken by automated, autonomous and modularly realized processes, it follows that, if these inform and influence practical reason and decision-making in the proper manner, that is, in ways relevant to helping us contend with the dual horns of the dilemma, then emotion might help us in contending with some of the frame problems that arise in the practical domain.

In Chapter 5, I considered the second principal objection to the idea of bringing emotion to bear, which reduces to the claim that emotion, even if undertaken by automated and autonomous operations and even if it does in fact influence practical reason and decision-making, fails to inform these processes in ways that are relevant to helping us contend with the frame problems that arise. That is, the challenge concludes, even if emotion does inform practical reason and decision-making, its influences are either detrimental or are on balance no more beneficial than detrimental.

And so, in the final chapter I considered the issues of *if* and if so *how* emotion might help us in contending with some of frame problems that arise in the practical domain. This required that reason be provided for our thinking that emotion might help us in contending with *both* horns of the dilemma (speed and accuracy). With the aim, then, of seeing just how far we might get by bringing emotion to bear on some instances of the problem arising in practical reason and decision-making, I began by presenting both an overview of some germane and representative empirical findings and Bechara & Damasio's somatic marker hypothesis. Having set this out, I suggested that emotion might be *one* means by which we contend with the attentional-direction, problem-sequencing, ends-selection, meta-planning and Hamlet's problems instances of the puzzle.

That the emotion-based heuristics, as simple and noncompensatory, should be far less computationally demanding and thus far more expeditious than the rational (in Fodor's sense) alternative appears clear. And so, I concluded that bringing emotion to bear on some instances of the puzzle should help us in contending with the tractability horn of the dilemma. Furthermore, I argued that we also have reason for thinking these heuristic operations to increase the accuracy of the operations of practical reason and decision-making. This, in turn, suggests that the emotion-based heuristics outlined are sufficiently normatively "good" to satisfy a suitably weakened normative rationality principle and thus the second horn of the dilemma posed by the frame problem.

It would appear then that at least some of what the emotions are are undertaken by automated, autonomous and modularly realized (and hence frame problem un-infected) operations of emotional learning and appraisal. We also have reason for thinking these to inform and influence practical reason and decision-making in ways that would appear to help us contending with *both* horns of the dilemma posed (*i.e.,* by both expediting and increasing the accuracy of these operations). It follows from this that bringing emotion to bear on the puzzle might provide *one* promising approach for our coming to understand how we contend with some of the frame problems that arise in the practical domain.

Following this discussion, I considered a number of open questions. I argued there that we have good reason for thinking that emotion-based heuristics are exploited quite automatically to influence and direct the operations of memory encoding and search, non-trivial planning and assent or opinion-fixation in ways that should be far more expeditious than the rational alternative. That is, we have reason, I concluded, for thinking that emotion should be highly relevant in helping us to contend with the tractability horn of the dilemma as posed by these instances of the problem. We also, I argued, have reason to think it likely that the emotion-based heuristics outlined should be sufficiently normatively "good" and thus that relying on these heuristics should help us in meeting the second horn of the dilemma as well. However, while this seems a plausible conjecture, given the lack of empirical evidence with respect to the normative issue, no conclusion was warranted with respect to these particular instances of the puzzle.

Emotion is no panacea. It does not "solve outright" the frame problem. I, of course, never claimed that it is or that it does. Rather, I argued for a modest point – that consideration of emotion helps us to understand how we might contend with *some* of the frame problems that arise in the practical domain. We have, I argued, good reason to think that it, *via* automated, autonomous and modular learning and appraisal operations, helps to expedite and increase the accuracy of practical reason and decision-making. Bringing emotion to bear on practical matters and integrating the emotional learning and appraisal operations outlined into Baars' existing Global Workspace framework, provides, I suggest, *one* plausible approach by which to explain how at least some of the "interesting" activities of mind might come to be modeled.

BIBLIOGRAPHY

Adelman, L. (1994). Molecular computation of solutions to combinatorial problems. *Science*, *266*, 1021-1024.

Adolphs, R. (2002). Neural mechanisms for recognizing emotions. *Current Opinion in Neurobiology, 12*, 169-178.

Adolphs, R. (2002). Recognizing emotion from facial expressions: Psychological and neurological mechanisms. *Behavioral and Cognitive Neuroscience Review, 1* , 21-61.

Adolphs, R. (1999). Social cognition and the human brain. *Trends in Cognitive Science*, 469-479.

Adolphs, R. (1999). The human amygdala and emotion. *The Neuroscientist, 5*, 125-137.

Adolphs, R., Cahill, L., Schul, R., & Babinsky, R. (1997). Impaired declarative memory for emotional material following bilateral amygdala damage in humans. *Learning & Memory, 4*, 291-300.

Adolphs, R., Damasio, H., Tranel, D., Cooper, G., & Damasio, A. (2000). A role for somatosensory cortices in the visual recognition of emotion as revealed by three-dimensional lesion mapping. *Journal of Neuroscience, 20*, 2683-2690.

Adolphs, R., Denburg, N., & Tranel, D. (2000). Impaired emotional declarative memory following unilateral amygdala damage. *Learning and Memory, 7*, 180-186.

Adolphs, R., Tranel, D., & Baron-Cohen, S. (2002). Amygdala damage impairs recognition of social emotions from facial expressions. *Journal of Cognitive Neuroscience, 14*, 1264-1274.

Adolphs, R., Tranel, D., Damasio, H., & Damasio, A. (1995). Fear and the human amygdala. *Journal of Neuroscience, 15*, 5879-91.

Adolphs, R., Tranel, D., Damasio, H., & Damasio, A. (1994). Imparied recognition of emotion in facial expression following bilateral damage to the human amygdala. *Nature, 372*, 669-672.

Adolphs, R., Tranel, D., Hamann, S., Young, A., Calder, A., Phelps, E., et al. (1999). Recognition of facial emotion in nine individuals with bilateral amygdala damage. *Neuropsychologia, 37*, 1111-1117.

Aggleton, J., & Passingham, R. (1981). Syndrome produced by lesions of the amygdala in monkeys (Macaca Mulatta). *Journal of Comparative and Physiological Psychology, 95*, 961-77.

Akiyama, T., Kato, M., Muramatsu, T., Umeda, S., Saito, F., & Kashima, H. (2007). Unilateral amygdala lesions hamper attentional orienting triggered by gaze direction. *Cerebral Cortex, 17,* 2593-2600.

Amorapanth, P., LeDoux, J., & Nader, K. (2000). Different lateral amygdala outputs mediate reactions and actions by a fear-arousing stimulus. *Nature Neuroscience, 3*, 74-79.

Arkes, H., Herren, L., & Isen, A. (1988). The role of potential loss in the influence of affect on risk-taking behavior. *Organizational Behavior and Human Decision Processes, 42* , 181-193.

Averill, J. (1980) A constructivist view of emotion. In R. Plutchik & H. Kellerman (eds.) *Emotion: Theory, research and experience: Vol 1 Theories of emotion* (pp. 305-339). New York: Academic Press.

Baars, B. (1988). *A Cognitive Theory or Consciousness.* Cambridge: Cambridge University Press.

Baars, B. (2002). The conscious access hypothesis: origins and recent evidence. *Trends in Cognitive Science, 6,* 47 – 52.

Baddelely, A. (1992). Working Memory. *Science*, 556-559.

Baral, C., Kreinovich, V., & Trejo, R. (2000). Computational complexity of planning and approximate planning in the presence of incompleteness. *Artificial Intelligence, 122*, 241-267.

Bargh, J., Chaiken, S., Govender, R., & Pratto, F. (1992). The generality of the automatic attitude effect. *Journal of Personality and Social Psychology, 62*, 893-912.

Barklow, J., Cosmides, L., & Tooby, J. (1992). *The Adapted Mind: Evolutionary Psychology and the Generation of Culture.* New York: Oxford University Press.

Barrett, H. (2005). Enzymatic computation and cognitive modularity. *Mind & Language, 20,* 259-287.

Baxter, M., & Chiba, A. (1999). Cognitive functions of the basal forebrain. *Current Opinion in Neurobiology, 9*, 178-183.

Bechara, A., Damasio, A., Damasio, H., & Anderson, S. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, 7-15.

Bechara, A., Damasio, H., & Damasio, A. (2000). Emotion, decision-making and the orbitofrontal cortex. *Cerebral Cortex, 10*, 295-307.

Bechara, A., Damasio, H., & Damasio, A. (2003). Role of the amygdala in decision-making. *Annals of the New York Academy of Sciences, 985*, 356-369.

Bechara, A., Damasio, H., Damasio, A., & Lee, G. (1999). Different contributions of the human amygdala and ventromedial prefrontal cortex to decision-making. *Journal of Neuroscience, 19*, 5473-5481.

Bechara, A., Damasio, H., Damasio, A., & Tranel, D. (1996). Failure to respond automatically to anticipated future outcomes following damage to prefrontal cortex. *Cerebral Cortex, 6*, 215-225.

Bechara, A., Damasio, H., Tranel, D., & Anderson, S. (1998). Dissociation of working memory from decision-making within the human prefrontal cortex. *Journal of Neuroscience*, 428-437.

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. (1997). Deciding advantageously before knowing the advantageous strategy. *Science*, 1293-1295.

Bechara, A., Tranel, D., & Damasio, H. (2000). Characterization of the decision-making deficit of patients with ventromedial prefrontal cortex lesions. *Brain, 123*, 2189-2202.

Bechara, A., Tranel, D., Damasio, H., Adolphs, R., Rockland, C., & Damasio, A. (1995). Double dissociation of conditioning and declarative knowledge relative to amygdala and hippocampus in humans. *Science, 269*, 1115-1118.

Beck, A., & Clark, D. (1988). Anxiety and depression: An information processing perspective. *Anxiety Research, 1*, 23-36.

Benenson, Y. (2001). Programmable and autonomous computing machines made of biomolecules. *Nature, 414,* 430-434.

Berkowitz, L. (2000). *Causes and Consequences of Feelings.* Cambridge: Cambridge University Press.

Berry, D. (1987). The problem of implicit knowledge. *Expert Systems, 4,* 144-151.

Blaney, P. (1986). Affect and memory: A review. *Psychological Bulletin, 99*, 229-246.

Bless, H., Bohner, G., Schwarz, N., & Stack, F. (1990). Mood and persuasion: A cognitive response analysis. *Personality and Social Psychology Bulletin, 16*, 331-345.

Bless, H., Clore, G., Schwarz, N., Golisano, V., Rabe, C., & Wolk, M. (1996). Mood and the use of scripts: Does being in a happy mood really lead to mindlessness? *Journal of Personality and Social Psychology, 71*, 665-679.

Boden, M. (1990). *The Philosophy of Artificial Intelligence.* Oxford: Oxford University Press.

Boden, M. (1996). *The Philosophy of Artificial Life.* Oxford: Oxford University Press.

Bolte, A., Goschke, T., & Kuhl, J. (2003). Emotion and intuition: Effects of positive and negative mood on implicit judgments of semantic coherence. *Psychological Science, 14*, 416-421.

Boolos, G., & Jeffery, R. (1999). *Computability and Logic.* Cambridge: Cambridge University Press.

Boussaoud, D., Desimone, R., & Ungerleider, L. (1990). Pathways for motion analysis: Cortical connections of the medial superior temporal and fundus of the superior temporal visual areas in the macaque. *Journal of Computational Neurology, 296*, 462-495.

Bower, G. (1981). Mood and Memory. *American Psychologist 36*, 129-148.

Bower, G., & Forgas, J. (1987). Mood effects on person-perception judgments.

Bower, G., & Mayer, J. (1985). Failure to replicate mood-dependent retrieval. *Bulletin of the Psychonomic Society, 23*, 39-42.

Bower, G., & Mayer, J. (1989). In search of mood-dependent retrieval. *Journal of Social Behavior and Personality, 4*, 121-156.

Bower, G., Gilligan, S., & Monteiro, K. (1981). Selectivity of learning caused by affective states. *Journal of Experimental Psychology: General 110*, 451-473.

Bower, G., Monteiro, K., & Gilligan, S. (1978). Emotional mood as a context for learning and recall. *Journal of Verbal Learning and Verbal Behavior, 17*, 573-585.

Broadbent, & Broadbent. (1988). Anxiety and attentional biases: state and trait. *Cognition and Emotion 2*, 165-183.

Buccino, G., Binkofski, F., & Riggio, L. (2004). The mirror neuron system and action recognition. *Brain and Language, 89*, 370-376.

Burgess, N., Jeffery, K., & O'Keefe, J. (1999). *The Hippocampal and Parietal Foundations of Spatial Cognition.* Oxford: Oxford University Press.

Burke, M., & Mathews, A. (1992). Autobiographical memory and clinical anxiety. *Cognition & Emotion, 6*, 23-35.

Cacioppo, J. (1993). The Physiology of Emotion. In M. Lewis, & J. Haviland, *Handbook of Emotions* (pp. 119-142). New York: Guilford Press.

Cahill, L., & McGaugh, J. (1995). A novel demonstration of enhanced memory associated with emotional arousal. *Consciousness and Cognition, 4*, 410-421.

Cahill, L., Haier, R., Fallon, J., Alkire, M., Tang, C., Keator, D., et al. (1996). Amygdala activity at encoding correlated with long-term free recall of emotional information. *Proceedings of the National Academy of Sciences* .

Cahill, L., Prins, B., Weber, M., & McGaugh, J. (1994). Beta-andrenergic activation and memory for emotional events. *Nature, 371*, 702-704.

Calhoun, C., & Solomon, R. (1984). *What is an Emotion? Classical Readings in Philosophical Psychology.* New York: Oxford University Press.

Campeau, S., & Davis, M. (1995). Involvement of subcortical and cortical afferents to the lateral nucleus of the amygdala in fear conditioning measured with fear-potentiated startle in rats trained with auditory and visual conditioned stimuli. *Journal of Neuroscience, 15*, 2312-2327.

Campeau, S., Miserendino, M., & Davis, M. (1992). Intra-amygdala infusion of M-methyl-D-Aspartate receptor antagonist AP5 blocks acqusition but not expression of fear-potentiated startle to an auditory conditioned stimulus. *Behavioral Neuroscience, 106*, 569-574.

Canamero, L. (2003). Designing emotions for activity selection in autonomous agents. In R. Trapple, P. Petta, & S. Payr (eds.), *Emotion in humans and artifacts* (pp. 115-148). Cambridge: MIT Press.

Canli, T., Zhao, Z., Brewer, J., Gabrieli, J., & Cahill, L. (2000). Event-related activation in the human amygdala associates with later memory for individual emotional experiences. *Journal of Neuroscience, 20*, 91-95.

Carli, M., Robbins, T., Evenden, J., & Everitt, B. (1983). Effects of lesions to ascending noradrenergic neurons on performance of a 5-choice serial reaction time task in rats: implications for theories of dorsal noradrenergic bundle function based on selective attention and arousal. *Behavioral Brain Research, 9*, 361-380.

Carnevale, P., & Isen, A. (1986). The Influence of postive affect and visual access on the discovery of integrative solutions in bilateral negotiation. *Organizational Behavior and Human Decision Processes 37*, 1-13.

Carruthers, P. (2003). Moderately massive modularity. In A. O'Hear, *Mind and Persons.* Cambridge: Cambridge University Press.

Carruthers, P. (2003). On Fodor's problem. *Mind and Language, 18*, 502-523.

Carruthers, P. (2004). Practical reasoning in a modular mind. *Mind and Language, 19*, 259-278.

Carruthers, P. (2006). *The Architecture of Mind: Massive Modularity and the Flexibility of Thought.* New York: Oxford University Press.

Carruthers, P., Stich, S., & Siegal, M. (2002). *The Cognitive Basis of Science.* Cambridge: Cambridge University Press.

Cermak, L. (1994). *Neuropsychological Explorations of Memory and Cognition.* Berlin: Springer Verlag.

Challis, B., & Krane, R. (1988). Mood induction and the priming of semantic memory in a lexical decision task: Asymmetric effects of elation and depression. *Bulletin of the Psychonomic Society, 26*, 309-312.

Cherniak, C. (1992). *Minimal Rationality.* Cambridge: MIT Press.

Cherniak, C. (1983). Rationality and the structure of human memory. *Synthese, 57*, 163-186.

Cherniak, C. (1988). Undebuggability and cognitive science. *Communications of the ACM 31*, 402-412.

Chiappe, D., & Kukla, A. (1996). Context selection and the frame problem. *Behavioral and Brain Sciences, 19*, 529-530.

Christianson, S. (1992). *The Handbook of Emotion and Memory.* Hillsdale.

Ciaramelli, E., Muccioli, M., Ladavas, E., & di Pellegrino, G. (2007). Selective deficit in personal moral jugdment following damage to ventromedial prefrontal cortex. *Social Cognitive and Affective Neuroscience, 2,* 84-92.

Claparede, E. (1911). Recognition of "Me-ness". In D. Rapaport, *Organization and pathology of thought* (pp. 58-75). New York: Columbia University Press.

Clark, A. (2001). *Mindware: An Introduction to the Philosophy of Cognitive Science.* New York: Oxford University Press.

Clark, D., & Teasdale, J. (1985). Constraints on the effects of mood on memory. *Journal of Personality and Social Psychology, 48*, 1596-1608.

Clark, M., & Isen, A. (1982). Toward understanding the relationship between feeling states and social behavior. In A. Hastorf, & A. Isen, *Cognitive Social Psychology* (pp. 73-108). New York: Elsevier.

Clore, G., Schwarz, N., & Conway, M. (1994). Cognitive causes and consequences of emotion. In R. Wyer, & T. Srull, *Handbook of Social Cognition* (pp. 323-417). Hillsdale: Lawrence Erlbaum.

Cook, S. (1983). An overview of computational complexity. *Communications of the ACM 26*, 401-408.

Dagleish, T., & Watts, F. (1990). Biases of attention and memory in disorders of anxiety and depression. *Clinical Psychology Review 10*, 589-604.

Damasio, A. (1994). *Descartes' Error: Emotion, Reason and the Human Brain*. New York: Avon Books.

Damasio, A. (2003). *Looking for Spinoza: Joy, Sorrow, and the Feeling Brain.* New York: Harcourt.

Damasio, A. (1999). *The Feeling of What Happens: Body and Emotion in the Making of Consciousness.* New York: Harcourt.

Damasio, A., Tranel, D., & Damasio, H. (1990). Individuals with sociopathic behavior caused by frontal damage fail to respond automatically to social stimuli. *Behavioral Brain Research 41*, 81-94.

Damasio, H., Grabowski, T., Damasio, A., Tranel, D., Boles-Ponto, L., Watkins, G., et al. (1993). Visual recall with eyes closed and covered activates early visual cortices. *Society for Neuroscience Abstracts, 19*, 1603.

Davidson, R. (1993). The neuropsychology of emotion and affective styles. In M. Lewis, & J. Haviland, *Handbook of Emotion* (pp. 143-154). New York: Guilford Press.

Davidson, R., Ekman, P., Sharon, C., Senulis, J., & Friesen, W. (1990). Approach withdrawal and cerebral asymmetry: emotional expression and brain physiology. *Journal of Personality and social psychology*, 330-341.

Davis, M. (1994). The role of amygdala in emotional learning. *Int. Review of Neurobiology, 36*, 225-266.

Davis, M. (2000). The role of the amygdala in conditioned and unconditioned fear and anxiety. In J. Aggleton, *The Amygdala* (pp. 213-288). Oxford: Oxford University Press.

Davis, M. (1992). The role of the amygdala in fear-potentiated startle: Implications for animal models of anxiety. *Trends in Pharmacological Science, 13*, 35-41.

DeHouwer, J., Hermans, D., & Eelen, P. (1998). Affective and identity priming with episodically associated stimuli. *Cognition and Emotion, 12,* 145-169.

Delise, D., Squire, L., Bihrle, A., & Massman, P. (1992). Componential analysis of problem-solving ability: performance of patients with frontal lobe damage and amnesic patients on a new sorting test. *Neuropsychologia 30*, 683-697.

Dennett, D. (1981). *Brainstorms: Philosophical Essays on Mind and Psychology.* Cambridge: MIT Press.

Dennett, D. (1987). Cognitive wheels: The frame problem of AI. In Z. Pylyshyn, *The Robot's Dilemma* (pp. 40-64). Norwood: Ablex Publishing Corporation.

Dennett, D. (1991). *Consciousness Explained.* Boston: Little Bron and Company.

Dennett, D. (1996). Producing future by telling stories. In K. Ford, & Z. Pylyshyn, *The Robot's Dilemma Revisited* (pp. 1-8). Norwood: Ablex Publishing Corporation.

Depue, R., & Iacono, W. (1989). Neurobehavioral aspects of affective disorder. *Annual Review of Psychology*, 457-492.

Desimone, R., & Ungerleider, L. (1989). Neural mechanisms of visual processing in monkeys. In F. Boller, & J. Grafman, *Handbook of Neuropsychology Vol 2* (pp. 267-299). Amsterdam: Elsevier Press.

de Sousa, R. (1987). *The Rationality of Emotion.* Cambridge: MIT Press.

Dietrich, E., & Fields, C. (1996). The role of the frame problem in Fodor's modularity of mind thesis: A case study of rationalist cognitive science. In K. Ford, & Z. Pylyshyn, *The Robot's Dilemma Revisited* (pp. 9-24). Norwood: Ablex Publishing Corporation.

Dolan, R. (2002). Emotion, Cognition and Behavior. *Science*, 1191-1194.

Dreyfus, H. (1999). *What Computers Still Can't Do: A Critique of Artifical Reason.* Cambridge: MIT Press.

Dreyfus, H., & Dreyfus, S. (1987). How to stop worrying about the frame problem even though it's computationally insoluable. In Z. Pylyshyn, *The Robot's Dilemma* (pp. 95-112). Norwood: Ablex Publishing Corporation.

Duhem, P. (1977). *The Aim and Structure of Physical Theory.* New York: Atheneum.

Dunn, D., & Wilson, T. (1990). When the stakes are high: A limit to the illusion of control effect. *Social Cognition, 8*, 305-323.

Ehlers, A., Margraf, J., Davies, S., & Roth, W. (1988). Selective processing of threat cues in subjects with panic attacks. *Cognition and Emotion, 2*, 201-219.

Eibl-Eibesfeldt, I. (1973). The expressive behavior of the deaf-and-blind born. In M. von Cranach & I. Vine (eds.) *Social Communication and Movement*. New York: Academic Press.

Eich, E. (1995). Mood as a mediator of place dependent memory. *Journal of Experimental Psychology: General, 124*, 293-308.

Eich, E. (1995). Searching for mood dependent memory. *Psychological Science 6*, 67-75.

Eich, E., & Macaulay, D. (2000). Fundamental factors in mood dependent memory. In J. Forgas, *The Role of Affect in Social Cognition* (pp. 109-130). New York: Cambridge University Press.

Eich, E., & Metcalf, J. (1989). Mood dependent memory for internal versus external events. *Journal of Experimental Psychology: Learning, Memory and Cognition 15*, 443-455.

Eich, E., Macaulay, D., & Ryan, L. (1994). Mood dependent memory for events of the personal past. *Journal of Exerimental Psychology: General, 123*, 201-215.

Eichenbaum, H. (2002). *The Cognitive Neuroscience of Memory.* Oxford: Oxford University Press.

Ekman, P. (1992). Are there basic emotions? *Psychological Review 99*, 550-565.

Ekman, P. (1982). *Emotion in the Human Face.* Cambridge: Cambridge University Press.

Ekman, P. (1977). Biological and cultural contributions to body and facial movement. In J. Blacking (ed.). *Antrhopology of the Body.* London: Academic Press, pp. 34-84.

Ekman, P., & Davidson, R. (1994). *The Nature of Emotion: Fundamental Questions.* New York: Oxford University Press.

Ekman, P., Friesen, W., & Simmons, R. (1985). Is the startle reaction an emotion? *Journal of Personality and Social Psychology 49*, 1416-1426.

Ekman, P., Levinson, R., & Friesen, W. (1983). Autonomic nervous system activity distinguishes among emotions. *Science, 221*, 1208-1210.

Ekman, P. (1973) Darwin and cross cultural studies of facial expression. In P. Ekman (ed.) *Darwin and Facial Expression: A Century of research in review*. New York: Academic Press.

Ekman, P., & Friesen, W. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17,* 124-129.

Ellis, H., Thomas, R., & Rodriguez, I. (1984). Emotional mood states and memory: Elaborative encoding, semantic processing, and cognitive effort. *Journal of Experimental Psychology: Learning, Memory and Cognition, 10*, 470-482.

Etcoff, N., & Magee, J. (1992). Categorical perception of facial expressions. *Cognition, 44*, 227-240.

Etherington, D., Kraus, S., & Perlis, D. (1991). Limited scope and circumscriptive reasoning. In K. Ford, & P. Hayes, *Reasoning Agents in a Dynamic World: The Frame Problem* (pp. 43-54). Greenwich: Jai Press.

Evans, J. (1996). Deciding before you think: Relevance and reasoning in the selection task. *British Journal of Psychology*, 223-240.

Evans, J. (1984). Heuristic and analytic processes in reasoning. *British Journal of Psychology 75*, 451-468.

Evans, J. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Science, 7*, 454-459.

Evans, J., & Over, D. (1996). Rationality in the selection task: Epistemic utility versus uncertainty reduction. *Psychological Review, 103*, 356-363.

Falls, W., Miserendino, M., & Davis, M. (1992). Extinction of fear-potentiated startle: Blockade by infusion of NMDA antagonist into the amygdala. *Journal of Neuroscience, 12*, 854-863.

Feigenbaum, J., & Rolls, E. (1991). Allocentric and egocentric spatial information processing in the hippocampal formation of the behaving primate. *Psychobiology 19*, 21-40.

Ferre, P. (2003). Effects of level of processing on memory for affectively valenced words. *Cognition and Emotion, 17*, 859-880.

Fessler, D., Pillsworth, E., & Flamson, T. (2004). Angry men and disgusted women: An evolutionary approach to the influences of emotion on risk taking. *Organizational Behavior and Human Decision Processes, 95*, 107-123.

Fetzer, J. (1991). The Frame problem: Artificial intelligence meets David Hume. In K. Ford, & P. Hayes, *Reasoning Agents in a Dynamic World: The Frame Problem* (pp. 55-70). Greenwich: Jai Press.

Fiedler, K. (1990). Mood-dependent selectivity in social cognition. In W. Stroebe, & M. Hewstone, *European Review of Social Psychology, Vol 1.* New York: Wiley.

Fiedler, K. (1991). On the task, measure, and the mood in research on affect and social cognition. In J. Forgas, *Emotion and Social Judgments.* Cambridge: Cambridge University Press.

Fiedler, K. (2000). Toward an integrative account of affect and cogntion phenomena using the BIAS computer algorithm. In J. Forgas, *Feeling and Thinking: The Role of Affect and Social Cognition* (pp. 223-252). New York: Cambridge University Press.

Foa, E., Feske, V., Murdoch, T., Kozak, M., & McCarthy, P. (1991). Processing of threat related information in rape-victims. *Journal of Abnormal Psychology, 100*, 156-162.

Fodor, J. (1987). Modules, frames, fridgeons, sleeping dogs, and the music of the spheres. In Z. Pylyshyn, *The Robot's Dilemma* (pp. 139-149). Norwood: Ablex Publishing Corporation.

Fodor, J. (1985). Precis of The Modularity of Mind. *Behavioral and Brain Sciences, 8*, 1-5.

Fodor, J. (1983). *Representations: Philosophical Essays on the Foundations of Cognitive Science.* Cambridge: MIT Press.

Fodor, J. (1975). *The Language of Thought.* Cambridge: Harvard University Press.

Fodor, J. (2001). *The Mind Doesn't Work that Way: The Scope and Limits of Computational Psychology.* Cambridge: MIT Press.

Fodor, J. (1983). *The Modularity of Mind.* Cambridge: MIT Press.

Ford, K., & Hayes, P. (1991). Framing the frame problem. In K. Ford, & P. Hayes, *Reasoning Agents in a Dynamic World: The Frame Problem* (pp. ix-xiv). Greenwich: Jai Pres.

Ford, K., & Hayes, P. (1991). *Reasoning Agents in a Dynamic World: The Frame Problem.* Greenwich: Jai Press.

Ford, K., & Pylyshyn, Z. (1996). *The Robot's Dilemma Revisited: The Frame Problem in Artificial Intelligence.* Norwood: Ablex Publishing.

Forgas, J. (1991). Affective influence on partner choice: Role of mood in social decisions. *Journal of Personality and Social Psychology 61*, 708-720.

Forgas, J. (1991). *Emotion and Social Judgment.* New York: Pergammon Press.

Forgas, J. (1995). Emotion in social judgment: Review and a new affect infusion model (AIM). *Psychological Bulletin, 117*, 39-66.

Forgas, J. (2000). *Feeling and Thinking: The Role of Affect in Social Cognition.* Cambridge: Cambridge University Press.

Forgas, J. (1995). Mood and judgment: The affect infusion model (AIM). *Psychological Bulletin, 117*, 39-66.

Forgas, J. (1989). Mood effects on decision-making strategies. *Australian Journal of Psychology, 41*, 197-214.

Forgas, J. (1998). On feeling good and getting your way: Mood effects on negotiation strategies and outcomes. *Journal of Personality and Social Psychology, 74*, 565-577.

Forgas, J. (1993). On making sense of odd couples: Mood effects on the perception of mismatched relationships. *Personality and Social Psychology Bulletin, 19*, 59-71.

Forgas, J. (1995). Strange couples: Mood effects on jugdments and memory about prototypical and atypical targets. *Personality and Social Psychology Bulletin, 21*, 747-765.

Forgas, J. (2001). *The Handbook of Affect and Social Cognition.* New York: Erlbaum Associates.

Forgas, J., & Bower, G. (1987). Mood effects on person-perception judgments. *Journal of Personality and Social Psychology*, 53-60.

Forgas, J., & Laham, S. (2004). The interaction between affect and motivation in social judgments and behavior. In J. Forgas, K. Williams, & W. Von Hippel, *Social Motivation.* Cambridge: Cambridge University Press.

Frank, R. (1988). *Passions within Reason: The Strategic Role of the Emotions.* New York: W.W. Norton and Co.

Frankish, K. (2004). *Mind and Supermind.* Cambridge: Cambridge University Press.

Franklin, S. (1999). *Artificial Minds.* Cambridge: MIT Press.

Freeman, W. (1957). Frontal lobotomy 1936-1956: A follow-up study of 3000 patients from one to twenty. *American Journal of Psychiatry, 113*, 877-886.

Freeman, W. (2002). *The Last Resort.* New York: Pressman.

Fridja, N. (1993). Moods, emotion episodes and emotions. In M. Lewis, & J. Haviland, *Handbook of Emotions* (pp. 381-404). New York: Guilford Press.

Fridja, N. (1986). *The Emotions.* New York: Cambridge University Press.

Fridja, N., Manstead, A., & Bem, S. (2000). *Emotions and Beliefs: How Feelings Influence Thoughts.* Cambridge: Cambridge University Press.

Fuster, J. (1997). *The Prefrontal Cortex: Anatomy, Physiology, and Neuropsychology of the Frontal Lobe.* New York: Lippincott-Raven.

Gallagher, M., Kapp, B., Musty, R., & Driscoll, P. (1977). Memory formation: evidence for a specific neurochemical system in the amygdala. *Science, 198*, 423-425.

Garey, M., & Johnson, D. (1979). *Computers and Intractability: A Guide to NP-Completeness.* New York: W.H. Freeman and Co.

Gewirtz, J., Falls, W., & Davis, M. (1997). Normal conditioned inhibition and extinction of freezing and fear-potentiated startle following electrolytic lesions of medial prefrontal cortices in rats. *Behavioral Neuroscience, 111*, 712-726.

Gigerenzer, G. (2000). *Adaptive Thinking: Rationality in the Real World.* Oxford: Oxford University Press.

Gigerenzer, G. (2001). *Bounded Rationality: The Adaptive Toolbox.* Cambridge: MIT Press.

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky (1996). *Psychological Review, 103*, 592-596.

Gigerenzer, G., & Todd, P. (1999). *Simple Heuristics That Make Us Smart.* New York: Oxford University Press.

Gilovich, T., D., G., & Kahneman, D. (2002). *Heuristics and Biases: The Psychology of Intuitive Judgment.* Cambridge: Cambridge University Press.

Gloor, P. (1992). Role of amygdala in temporal lobe epilepsy. In J. Aggleton, *The Amygdala: Neurobiological Aspects of Emotion, Memory and Mental Dysfunction* (pp. 339-352). New York: Wiley-Liss.

Glymour, C. (1987). Android epistemology and the frame problem: Comments on Dennett's "Cognitive Wheels". In Z. Pylyshyn, *The Robot's Dilemma* (pp. 65-76). Norwood: Ablex Publishing.

Glymour, C. (1996). The adventure among the asteroids of angela android, series 8400XF with an afterwood on planning, prediction, learning, the frame problem, and a few other subjects. In K. Ford, & P. Hayes, (eds.) *Reasoning Agents in a Dynamic World: The Frame Problem*. Greenwich: Jai Press.

Grattan, L. & Eslinger, P. (1992). Long-term psychological consequences of childhood frontal lobe lesion in patient DT. *Brain and Cognition, 20,* 185-195.

Green, J., Nystrom, L., Engell, A., Darley, J., & Cohen, J. (2004). The neural bases of cognitive conflict and control on moral judgment. *Neuron, 44,* 389-400.

Hayes, *The Robot's Dilemma Revisited* (pp. 25-34). Norwood: Ablex Publishing.

Goble, L. (2001). *The Blackwell Guide to Philosophical Logic.* Oxford: Blackwell Publishers.

Goel, V., Grafman, J., Tajik, J., Gana, S., & Danto, D. (1997) A study of the performance of patients with frontal lobe lesions in a financial planning task. *Brain, 120,* 1805-1822.

Goldman, A. (1986). *Epistemology and Cognition.* Cambridge: Harvard University Press.

Goldman, A. (1983). Epistemology and the theory of problem solving. *Synthese, 55*, 21-48.

Goldman, A. (1993). *Readings in Philosophy and Cognitive Science.* Cambridge: MIT Press.

Goldman, A. (2006). *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading.* Oxford: Oxford University Press.

Goodale, M., Meenan, J., Bulthoff, H., Nicolle, D., Murphy, K., & Racicot, C. (1994). Seperate neural pathways for the visual analysis of object shape in perception and prehension. *Current Biology, 4*, 604-610.

Goodman, N. (1983). *Fact, Forecast and Fiction.* Cambridge: Harvard University Press.

Greenspan, P. (2000). Emotional Strategies and Rationality. *Ethics, 110*, 469-487.

Greenspan, P. (1993). *Emotions and Reasons: An Inquiry into Emotional Justification.* New York: Routledge.

Greenspan, P. (1981). Emotions as Evaluations. *Pacific Philosophical Quarterly, 62*, 158-69.

Greenwald, A., Klinger, M., & Liu, T. (1989). Unconscious processing of dichoptically masked words. *Memory & Cognition, 17*, 35-47.

Griffiths, P. (1997). *What Emotions Really Are.* Chicago: University of Chicago Press.

Gustafson, D. (1964). *Essays in Philosophical Psychology.* New York: Anchor Books.

Halgren, E. (1992). Emotional neurophsiology of the amygdala within the context of human cognition. In J. Aggleton, *The Amygala: Neurobiological Aspects of Emotion, Memory and Mental Dysfunction* (pp. 191-228). New York: Wiley-Liss.

Han, J., Holland, P., & Gallagher, M. (1999). Disconnection of amygdala central nucleus and substantia innominata/nucleus basalis disrupts increments in conditioned stimulus processing. *Behavioral Neuroscience, 113*, 143-151.

Han, J., McMahan, R., Holland, P., & Gallagher, M. (1997). The Role of an amygdala-nigrostriatal pathway in associative learning. *Journal of Neuroscience, 17*, 3913-3919.

Harel, D. (1996). *Algorithmics: The Spirit of Computing.* New York: Addison-Wesley.

Harre, R. (1986). *The Social Construction of Emotions.* Oxford: Blackwell Publishers.

Haugeland, J. (1987). An overview of the frame problem. In Z. Pylyshyn, *The Robot's Dilemma* (pp. 77-94). Norwood: Ablex Publishing.

Haugeland, J. (1985). *Artificial Intelligence: The Very Idea.* Cambridge: MIT Press.

Haugeland, J. (1997). *Mind Design II: Philosophy, Psychology, Artificial Intelligence.* Cambridge: MIT Press.

Haugh, B. (1991). Omniscience isn't needed to solve the frame problem. In K. Ford, & P. Hayes, *Reasoning Agents in a Dynamic World: The Frame Problem* (pp. 105-132). Greenwich: Jai Press.

Hayes, P. (1987). What the Frame Problem is and isn't. In Z. Pylyshyn, *The Robot's Dilemma* (pp. 123-138). Norwood: Ablex Publishing.

Hayes, P., Ford, K., & Agnew, N. (1996). Epilog: Goldilocks and the Frame Problem. In K. Ford, & Z. Pylyshyn, *The Robot's Dilemma Revisited* (pp. 135-138). Norwood: Ablex Publishing.

Helmuth. (2003). Fear and trembling in the amygdala. *Science, 300*, 568-569.

Holland, P., & Gallagher, M. (1999). Amygdala circuitry in attentional and representational processes. *Trends in Cognitive Sciences, 3*, 65-73.

Holland, P., & Gallagher, M. (1993). Anygdala central nucleus disrupts increments, but not decrements, in conditioned stimulus processing. *Bevarioal Neuroscience, 107*, 246-253.

Holland, P., Han, J., & Gallagher, M. (2000). Amygdala central nucleus lesions alter performance in a selective attention task. *Journal of Neuroscience, 20*, 6701-6706.

Horty, J. (2001). Nonmonotonic Logic. In L. Goble, *Blackwell Guide to Philosophical Logic* (pp. 336-361). Oxford: Blackwell Publishers.

Hunt, R., & Ellis, H. (1999). *Fundamental of Cognitive Psychology.* New York: McGraw-Hill.

Ingram, R. (1984). Towards an information processing analysis of depression. *Cognitive Therapy and Research, 8*, 443-478.

Irani, K., & Myers, G. (1983). *Emotion: Philosophical Studies.* New Jersey: Haven Books.

Isen, A. (1985). Asymmetry of happiness and sadness in effects on memory in normal college students. *Journal of Experimental Psychology: General, 114*, 388-391.

Isen, A. (1993). Positive affect and decision making. In M. Lewis, & J. Haviland, *Handbook of Emotions* (pp. 261-278). New York: Guilford Press.

Isen, A., & Daubman, K. (1984). The influence of affect on categorization. *Journal of Personality and Social Psychology, 47*, 1206-1217.

Isen, A., & Geva, N. (1987). The influence of positive affect on acceptable level of risk. *Organizational Behavior and Human Decision Processes, 39*, 145-154.

Isen, A., & Means. (1983). The influence of positive affect on decision-making strategy. *Social Cognition 2*, 18-31.

Isen, A., & Patrick, R. (1983). The effects of positive feeling on risk taking: When the chips are down. *Organizational Behavior and Human Decision Processes, 31*, 194-202.

Isen, A., Daubman, K., & Nowicki, G. (1987). Positive affect facilitates creative problem solving. *Journal of Personality and Social Psychology, 52*, 1122-1131.

Isen, A., Johnson, M., Mertz, E., & Robinson, G. (1985). The Influence of positive affect on the unusualness of word associations. *Journal of Personality and Social Psychology, 48*, 1413-1426.

Isen, A., Niedenthal, P., & Cantor, N. (1992). The influence of positive affect on social categorization. *Motivation and Emotion, 16*, 65-78.

Isen, A., Nygren, T., & Ashby, F. (1987). The influence of positive affect on the perceived utility of gains and losses. *Journal of Personality and Social Psychology, 55*, 710-717.

Isen, A., Pratkanis, A., Slovic, P., & Slovic, L. (1984). The influence of positive affect on risk preference. *92nd Annual Meeting of the American Psychological Association, Toronto* .

Isen, A., Rosenzweig, A., & Young, M. (1991). The influence of positive affect in clinical problem solving. *Medical Decision Making, 11*, 221-227.

Isen, A., Shalker, T., Clark, M., & Karp, L. (1978). Affect, accessibility of material in memory and behavior: a cognitive loop? *Journal of Personality and Social Psychology, 36*, 1-12.

Izard, C. (1992). Basic emotions, relations among emotions, and emotion-cognition relations. *Psychological Review, 99*, 561-565.

Izard, C. (1993). Organizational and motivational functions of discrete emotions. In M. Lewis, & J. Haviland, *Handbook of Emotions* (pp. 631-642). New York: Guilford Press.

Izard, C. (1991). *The Psychology of Emotions.* New York: Plenum Press.

James, T., Culham, J., Humphrey, K., Milner, D., & Goodale, M. (1997). Ventral occipital lesions impair recognition but not object-directed grasping: and fMRI study. *Brain 126, No. 11*, 2463-2475.

Janlert, L. (1987). Modeling change - the frame problem. In Z. Pylyshyn, *The Robot's Dilemma* (pp. 1-40). Norwood: Ablex Publishing.

Janlert, L. (1996). The frame problem: freedom or stability? With pictures we can have both. In K. Ford, & Z. Pylyshyn, *The Robot's Dilemma Revisited* (pp. 35-48). Norwood: Ablex Publishing.

Johnson, E., & Tversky, A. (1983). Affect, generalization and the perception of risk. *Journal of Personality and Social Psychology, 45*, 20-31.

Johnson, G. (1992). *In the Palaces of Memory.* New York: Vintage Books.

Kahn, B., & Isen, A. (1993). The influence of positive affect on variety-seeking among safe, enjoyable products. *Journal of Consumer Research, 20*, 257-270.

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review, 103*, 582-591.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment Under Uncertainty: Heuristics and Biases.* Cambridge: Cambridge University Press.

Kapp, B., Supple, W., & Whalen, P. (1994). Effects of electrical stimulation of the amygdaloid central nucleus on neocortical arousal in rabbits. *Behvarioral Neuroscience, 108*, 81-93.

Kapp, B., Whalen, P., Supple, W., & Pascoe, J. (1992). Amygdaloid contributions to conditioned arousal and sensory information processing. In J. Aggleton, *The Amygdala: Neuronbiological Aspects of Emotion, Memory and Mental dysfunction* (pp. 229-254). New York: Wiley-Liss.

Kastner, S., De Weerd, P., Desimone, R., & Ungerleider, L. (1998). Mechanisms of directed attention in human extrastriate cortex as revealed by functional MRI. *Science, 282*, 108-111.

Kastner, S., Pinsk, M., De Weerd, P., Desimone, R., & Ungerleider, L. (1999). Increased activity in human visual cortex during directed attention in absence of visual stimulation. *Neuron, 22*, 751-761.

Kluver, H., & Bucy, P. (1939). Preliminary analysis of functions of the temporal lobes in monkeys. *Archives Neurological Psychiatry, 42*, 979-1000.

Kluver, H., & Bucy, P. (1937). 'Psychic blindness' and other symptoms following bilateral temporal lobectomy in rhesus monkeys. *American Journal of Physiology 119*, 352-353.

Kornblith, H. (1997). *Naturalizing Epistemology.* Cambridge: MIT Press.

Kosslyn, S., & Koenig, O. (1995). *Wet Mind: The New Cognitive Neuroscience.* New York: Free Press.

Kosslyn, S., Alpert, N., Thompson, W., Maljkovic, V., Weise, S., Chabris, C., et al. (1993). Visual mental imagery activates topographically organized visual cortex: PET investigations. *Journal of Cognitive Neurosciences, 5*, 263-87.

Kosslyn, S., Shin, L., Thomson, W., McNally, R., Rauch, S., Pitman, R., et al. (1996). Neural effects of visualizing and perceiving aversive stimuli: a PET investigation. *Neuroreport, 7*, 1569-1576.

Kraiger, K., Billings, R., & Isen, A. (1989). The influences of positive affective state on task perceptions and satisfaction. *Organizational Behavior and Human Decision Processes, 44*, 12-25.

Kunst-Wilson, W., & Zajonc, R. (1980). Affective discrimination of stimuli that cannot be recognized. *Science, 207*, 557-558.

Kyburg, H. (1996). Dennett's Beer. In K. Ford, & Z. Pylyshyn, *The Robot's Dilemma Revisited* (pp. 49-60). Norwood: Ablex Publishing.

Lambert, K., & Shanks, D. (1997). *Knowledge, Concepts and Categories.* Cambridge: MIT Press.

Lang, P. (1984). Cognition in emotion: concept and action. In C. Izard, J. Kagan, & R. Zajonc, *Emotions, Cognitions, and Behavior* (pp. 192-228). Cambridge: Cambridge University Press.

Lazarus, R. (1984). On the Primacy of Cognition. *American Psychologist, 39*, 124-129.

LeBar, K. (1998). Role of the amygdala in emotional picture evaluation as revealed by fMRI. *Journal of Cognitive Neuroscience, 108* .

LeDoux, J. (1992). Emotion in the amygdala. In J. Aggleton, *The Amygdala: Neurobiological Aspects of Emotion, Mystery and Mental Dysfunction* (pp. 339-51). New York: Wiley Liss.

Ledoux, J. (1994). Emotion, memory and the human brain. *Scientific American, 270*, 34.

LeDoux, J. (1995). Emotion: Clues from the brain. *Annual Review of Psychology, 46*, 209-235.

LeDoux, J. (1993). Emotional memory systems in the brain. *Behavioral Brain Research, 58*, 69-79.

LeDoux, J. (1993). Emotional Networks in the Brain. In M. Lewis, & J. Haviland, *Handbook of Emotions* (pp. 109-118). New York: Guilford.

LeDoux, J. (1998). *The Emotional Brain.* New York: Simon & Schuster.

LeDoux, J., Romanski, L., & Xargoraris, A. (1989). Indelibility of subcortical emotional memories. *Journal of Cognitive Neuroscience, 1*, 238-243.

Lewis, H., & Papadimitriou, C. (1998). *Elements of the Theory of Computation.* New Jersey: Prentice Hall.

Lewis, H., & Papadimitriou, C. (1978). The efficiency of algorithms. *Scientific American*, 96-109.

Lewis, M., & Haviland, J. (1993). *Handbook of Emotions.* New York: Guilford Press.

Liang, K., Hon, W., & Davis, M. (1994). Pre and posttraining infusion of N-methyl-D-aspartate receptor antagonists into the amygdala impair memory in an inhibitory avoidance task. *Behavioral Neuroscience, 110*, 241-253.

Loewenstein, G., & Schkade, D. (1999). Out of control: Visceral influences on behavior. *Organizational Behavior and Human Decision Processes*, 272-292.

Lormand, E. (1996). The holorobophobe's dilemma. In K. Ford, & Z. Pylyshyn, *The Robot's Dilemma Revisited* (pp. 61-88). Norwood: Ablex Publishing.

Luce, M., Bettman, J., & Payne, J. (1997). Choice processing in emotionally difficult decisions. *Journal of Experimental Psychology: Learning, Memory and Cognition, 23*, 384-405.

Luria, A. (1980). Frontal Lobe Syndrome. In P. Vinken, & G. Burns, *Handbook of Clinical Neurology: Volume 2 Localization in Clinical Neurology* (pp. 725-757). New York: Press.

Macdonald, J., Stefanovic, D., & Stojanovic, M. (2008) Smart DNA: Programming the molecule of life for work and play. *Scientific American,* November.

MacLean, P. (1993). Cerebral evolution of emotion. In M. Lewis, & J. Haviland, *Handbook of Emotion* (pp. 67-83). New York: Guilford Press.

MacLeod, A., Byrne, A., & Valentine, J. (1996). Affect, emotional disorder and furture directed thinking. *Cognition and Emotion, 10*, 69-86.

MacLeod, C., & Mathews, A. (1991). Biased cognitive operations in anxiety: accessibility of information or assignment of processing priorities? *Behavioral Research & Therapy, 29*, 599-610.

MacLeod, C., Mathews, A., & Tata, P. (1986). Attentional Bias in emotional disorders. *Journal of Abnormal Psychology, 95*, 15-20.

MacLeod, C., Tata, P., & Mathews, A. (1987). Perception of emotionally valenced information in depression. *British Journal of Psycology, 26*, 67-68.

Maguire, E. (1997). Hippocampal involvement in human topographical memory: evidence from functional imaging. *Philosophical Transcripts of the Royal Society*, 1475-1480.

Marcus, G. (2001). *The Algebraic Mind: Integrating Connectionism and Cognitive Science.* Cambridge: MIT Press.

Markus, H., & Kitayama, S. (1991). Culture and self: Implications for cognition, emotion and motivation. *Psychological Review, 98*, 224-253.

Marr, D. (1982). *Vision.* New York: Freeman Press.

Martin, M., Williams, R., & Clark, D. (1991). Does anxiety lead to selective processing of threat-related information? *Behavioral Research and Therapy, 29*, 147-160.

Mathews. (1993). Biases in emotional processing. *The Psychologist: Bulletin of the British Psychological Society 6*, 493-499.

Mathews, A., & Klug, D. (1993). Emotionality and interference of color naming in anxiety. *Behavioral Research and Therapy, 29*, 147-160.

Mathews, A., & MacLeod, C. (1985). Selective processing of threat cues in anxiety states. *Behavioral Research & Therapy, 23*, 563-569.

Mathews, A., Mogg, K., May, J., & Eysenck, M. (1989). Implicit and explicit memory bias in recall. *Behavioral Research and Therapy, 21*, 233-239.

Matt, G., Vasquez, C., & Campbell, W. (1992). Mood congruent recall of affectively toned stimuli: A meta-analytic review. *Clinical Psychology Review,12*, 227-255.

Matthews, A. (1983). Biases in emotional processing. *The Psychologist: Bulletin of the British Psychological Society, 6*, 493-499.

Matthews, A., & MacLeod, M. (1994). Cognitive approaches to emotion and emotional disorders. *Annual Review of Psychology 45*, 25-50.

Mayer, J., & Volanth, A. (1985). Cognitive involvement in the mood response system. *Motivation and Emotion 9*, 261-275.

Mayer, J., Gaschke, Y., Braverman, D., & Evans, T. (1992). Mood-congruent judgment is a general effect. *Journal of Personality and Social Psychology, 63*, 119-132.

Mayer, J., McCormick, L., & Strong, S. (1995). Mood-congruent memory and natural mood: New evidence. *Personality and Social Psychology Bulletin*, 736-746.

McCabe, P., Schneiderman, N., Jarrell, T., Gentile, C., Teich, A., Winters, R., et al. (1992). Central Pathways involved in differential classical conditioning of heart rate responses. In E. Gormenzano, *Learning and Memory: The Behavioral and Biological Substrates* (pp. 321-46). Hillsdale: Erlbaum Press.

McCarthy, J. (1980). *Circumscription: A form of nonmonotonic reasoning.* Stanford Artifical Intelligence Laboratory (Memo: AIM-334).

McCarthy, J., & Hayes, P. (1969). Some philosophical problems from the standpoint of artificial intelligence. In B. Meltzer, & D. Michie, *Machine Intelligence Volume 4* (pp. 463-502). Edinburgh: University of Edinburgh Press.

McDermott, D. (1987). We've been framed: Or, why AI is innocent of the frame problem. In Z. Pylyshyn, *The Robot's Dilemma* (pp. 113-122). Norwood: Ablex Publishing.

McDonald, A. (1992). Cell types and intrinsic connections of the amygdala. In J. Aggleton, *The Amygdala: Neurobiological aspects of emotion, memory and mental dysfunction* (pp. 67-96). New York: Wiley-Liss.

McGaugh, J. (2004). The amygdala modulates the consolidation of memories of emotionally arousing experiences. *Annual Review Neuroscience, 27*, 1-28.

McGaugh, J., Cahill, L., & Roozendaal, B. (1996). Involvement of the amygdala in memory storage: interaction with other brain systems. *Proceedings of the National Academy of Science, 93*, 13508-13514.

McGaugh, J., Intrioni-Collison, I., Cahill, L., Catellano, C., Dalmaz, C., Parent, M., et al. (1993). Neuromodulatory systems and memory storage: role of the amygdala. *Behavioral Brain Research, 58*, 81-90.

McGaugh, J., Intrioni-Collison, I., Cahill, L., Kim, M., & Liang, K. (1992). Involvement of the amygdala in neuromodulatory influences on memory storage. In J. Aggelton, *The Amygdala: neurobiological aspects of emotion, memory and mental dysfunction* (pp. 431-451). New York: Wiley.

McGaugh, J., Kaiser, T., & Sarter, M. (1992). Behavioral vigilance following infusions of 192 IgC-saporin into the basal forebrain: selectivity of the behavioral impairment and relation to cortica AChE-positive fiber density. *Behavioral Neuroscience, 110*, 247-265.

McNally, R., Kaspi, S., Riemann, B., & Zeitlin, S. (1990). Selective processing of threat cues in postraumatic stress disorder. *Journal of Abnormal Psychology, 99*, 398-402.

McNally, R., Riemann, B., & Kim, E. (1990). Selective processing of threat cues in panic disorder. *Behaviour Research and Therapy, 28*, 407-412.

Mellars, B., & McGraw, A. (2001). Anticipated emotions as guides to choice. *Current Directions in Psychological Science, 10*, 210-214.

Mellars, B., Schwartz, A., Ko, H., & Ritov, I. (1997). Decision Affect Theory: Emotional reactions to the outcomes of risky options. *Psychological Sciences, 8*, 423-449.

Milham, M., Banich, M., Claus, E., & Cohen, N. (2003). Practice-related effects demonstrate complementary roles of anterior cingulate and prefrontal cortices in attentional control. *Neuroimage, 18*, 483-493.

Milner, A., & Goodale, M. (2006). *The Visual Brain in Action.* New York: Oxford University Press.

Mineka, N. (1994). The effects of high and low trait anxiety on implicit and explicit memory tasks. *Cognition and Emotion, 8*, 147-164.

Mineka, S., & Sutton, S. (1992). Cognitive biases and emotional disorders. *Psychological Sciences 3* , 65-69.

Minksy, M. (1988). *The Society of Mind.* New York: Simon & Schuster.

Miserendino, M., Sananes, C., Melia, K., & Davis, M. (1990). Blocking of acquisition but not expression of conditioned fear-potentiated startle by NMDA antagonists in amygdala. *Nature, 345*, 716-718.

Mogg, K., Mathews, A., & Weinman, J. (1987). Memory Bias in clinical anxiety. *Journal of Abnormal Psychology, 96*, 221-238.

Morgan, M., & LeDoux, J. (1995). Differential contributions of dorsal and ventral medial prefrontal cortex on acquisition and extinction of conditioned fear. *Behavioral Neuroscience, 109*, 681-88.

Morgan, M., Romanski, L., & LeDoux, J. (1993). Extinction of emotional learning: contribution of medial prefrontal cortex. *Neuroscience Letters, 163*, 109-13.

Morgenstern, L. (1991). Knowledge and the frame problem. In K. Ford, & P. Hayes, *Reasoning Agents in a Dynamic World: The Frame Problem* (pp. 133-170). Greenwich: Jai Press.

Morgenstern, L. (1996). The problem with solutions to the frame problem. In K. Ford, & Z. Pylyshyn, *The Robot's Dilemma Revisited* (pp. 99-134). Norwood: Ablex Publishing.

Muir, J., Dunnett, S., Robbins, T., & Everitt, B. (1992). Attentional functions of the forebrain cholinergic systems: effects of intraventricular hemicholinium, physostigmine, basal forebrain lesions and intracortical grafts on a multliple choice serial reaction time task. *Experimental Brain Research, 89*, 611-622.

Muir, J., Everitt, B., & Robbins, T. (1996). The cerebral cortex of the rat and visual attention function: dissociable effects of mediofrontal, cingulate, anterior dorsolateral, and parietal cortex lesions on a five-choice serial reaction time task. *Cerebral Cortex, 6*, 470-481.

Naccaches, & Dahaene. (2001). The cognitive neuroscience of consciousness. *Cognition, 79*.

Nader, K., Majidishad, P., Amorapanth, P., & LeDoux, J. (2001). Damage to the lateral and central, but not other, amygdaloid nuclei prevents the acquisition of auditory fear conditioning. *Learning and Memory, 8*, 156-63.

Newell, A., & Simon, H. (1972). *Human Problem Solving.* New Jersey: Prentice Hall.

Newell, A., Shaw, J., & Simon, H. (1995). Chess-Playing Programs and the Problem of Complexity (1958). In E. Feigenbaum, & J. Feldman, *Computers and Thought.* New York: McGraw Hill.

Niedenthal, P. (1990). Implicit perception of affective information. *Journal of Experimental Social Psychology, 26*, 505-527.

Niedenthal, P., & Kitayama, S. (1994). *The Heart's Eye: Emotional influences on perception and attention.* San Diego: Academic Press.

Niedenthal, P., & Setterlund, M. (1994). Emotion congruence in perception. *Personality and Social Psychology Bulletin, 20*, 401-411.

Niedenthal, P., & Showers, C. (1991). The perception and processing of affective information and its influences on social judgment. In J. Forgas, *Emotion and Social Judgment* (pp. 125-143). New York: Pergamon Press.

Niedenthal, P., Halberstadt, J., & Setterlund, M. (1997). Being happy and seeing "happy": Emotional state mediates visual word recognition. *Emotion and Cognition, 11*, 403-432.

Nolte, J. (1999). *The Human Brain: An Introduction to its Functional Anatomy.* Baltimore: Mosby Press.

Nutter, T. (1991). Focus of attention, content and the frame problem. In K. Ford, & P. Hayes, *Reasoning Agents in a Dynamic World: The Frame Problem* (pp. 171-188). Greenwich: Jai Press.

Nygren, T., Isen, A., Taylor, P., & Dulin, J. (1996). The influence of positive affect on the decision rule in task situations: Focus on outcome (and especially avoidance of loss) rather than probability. *Organizational Behavior and Human Decision Processes*, 59-72.

Oatley, K. (1993). Social construction in emotions. In M. Lewis, & J. Haviland, *Handbook of Emotions* (pp. 341-352). New York: Guilford Press.

Ohman, A. (1993). Fear and anxiety as emotional phenomena: Clinical phenomenology, evolutionary perspectives, and information-processing mechanisms. In M. Lewis, & J. Haviland, *Handbook of Emotions* (pp. 511-536). New York: Guilford Press.

Ohman, A., & Mineka, S. (2001). Fears, phobias and preparedness: Toward an evolved model of fear and fear learning. *Psychological Review, 108*, 483-522.

Ortony, A., & Turner, T. (1990). What's basic about basic emotions? *Psychological Review, 97*, 315-331.

Owen, A., Downes, J., Sahakian, B., Polkey, C., & Robbins, T. (1990). Planning and spatial working memory following frontal lobe lesions in man. *Neuropsychologia, 28*, 1021-34.

Pankespp, J. (1991). Gray zones at the emotion/cognition interface: A commentary. *Cognition and Emotion, 4*, 289-302.

Panksepp, J. (1992). A critical role for "affective neuroscience" in resolving what is basic about basic emotions. *Psychological Review, 99*, 554-560.

Panksepp, J. (1996). Affective neuroscience: A paradigm to study the animate circuits for human emotions. In R. Kavanaugh, B. Zimmerberg, & S. Fein, *Emotions: Interdisciplinary approaches* (pp. 29-57). Mahwah: Lawrence Erlbaum Press.

Panksepp, J. (1998). *Affective Neuroscience: The Foundations of Human and Animal Emotions.* New York: Oxford University Press.

Panksepp, J. (1988). Brain emotional circuits and psychopathologies. In M. Clynes, & J.

Panksepp, J. (1993). Neurochemical control of mood and emotions: Amino acids to neuropeptides. In M. Lewis, & J. Haviland, *Handbook of Emotions* (pp. 87-108). New York: Guilford Press.

Panksepp, J. (1995). The emotional brain and biological psychiatry. *Advances in Biological Psychology, 1*, 263-286.

Papadimitriou, C. (1993). *Computational Complexity.* New York: Addison Wesley.

Papineau, D. (1996). *The Philosophy of Science.* Oxford: Oxford University Press.

Parrot, W., & Sabini, J. (1990). Mood and memory under natural conditions: Evidence for mood incongruent recall. *Journal of Personality and Social Psychology, 59*, 321-336.

Pecchinenda, A., & Smith, C. (1996). The affective significance of skin conductance activity during a difficult problem-solving task. *Cognition and Emotion, 10*, 481-503.

Picard, R. (1997) *Affective Computing.* Cambridge: MIT Press,

Picard, R. (2002). What does it mean for a computer to "have" emotion? In R. Trappl, P. Petta, & S. Payr, *Emotions in Humans and Artifacts* (pp. 213-236). Cambridge: MIT Press.

Pinker, S. (1997). *How the Mind Works.* New York: W.W. Norton & Co.

Pitkanen, A., Savander, V., & LeDoux, J. (1997). Organization of intra-amygdaloid circuitries in the rat: an emerging framework for understanding the functions of the amygdala. *Trends in Neuroscience, 20*, 517-523.

Pitkanen, A., Stefanacci, L., Farb, C., Go, C., LeDoux, J., & Amaral, D. (1995). Intrinsic connections of the rat amygdaloid complex: Projections originating in the lateral nucleus. *Journal of Comparative Neurology, 356*, 288-310.

Plutchik, R. (1993). Emotions and their vicissitudes: Emotions and psychopathology. In M. Lewis, & J. Haviland, *Handbook of Emotion* (pp. 53-66). New York: Guilford Press.

Plutchik, R. (1994). *The Psychology and Biology of Emotions.* New York: Harper Collins.

Plutchik, R., & Kellerman, H. (1980). *Emotion: Theory, Research and Experience: Volume 1 Theories of Emotion.* New York: Academic Press.

Plutchik, R., & Kellerman, H. (1980). *Emotion: Theory, Research and Experience: Volume 2 Emotions in Early Development.* New York: Academic Press.

Plutchik, R., & Kellerman, H. (1980). *Emotion: Theory, Research and Experience: Volume 3 Biological Foundations of Emotion.* New York: Academic Press.

Pollock, J. (1997). Reasoning about change and persistence: A solution to the frame problem. *Nous, 31*, 143-169.

Post, E. (1943). Formal reductions of the general combinatorial decision problem. *American Journal of Mathematics, 65*, 197-215.

Pratto, F., & John, O. (1991). Automatic vigilance: The attention-grabbing power of negative social information. *Journal of Personality and Social Psychology, 61*, 380-391.

Priest, G. (2001). *An introduction to non-classical logic.* Cambridge: Cambridge University Press.

Prigatano, G., & Schacter, D. (1991). *Awareness of deficit after brain injury.* New York: Oxford University Press.

Pugmire, D. (1998). *Rediscovering Emotion.* Edinburgh: Edinburgh University Press.

Quine, W. (1980). *From a Logical Point of View: Nine Logico-Philosophical Essays.* Cambridge: Harvard University Press.

Quine, W. (1960). *Word and Object.* Cambridge: MIT Press.

Raymond, J., Fenske, M., & Tavassoli, N. (2003). Selective attention determines emotional responses to novel visual stimuli. *Psychological Sciences, 14*, 537-542.

Rempel-Clower, N., Zola, S., Squire, L., & Amaral, D. (1996). Three cases of enduring memory impairment after bilateral damage to hippocampal formation. *Journal of Neuroscience, 16*, 5233-5255.

Revelle, W., & Loftus, D. (1992). The implications of arousal effects for the study of affect and memory. In S. Christianson, *The Handbook of Emotion and Memory: Research and Theory.* Hillsdale: Lawrence Elrbaum.

Rey, G. (1997). *Contemporary Philosophy of Mind.* Oxford: Blackwell Publishers.

Roberts, A., Robbins, T., & Weiskrantz, L. (1998). *The Prefrontal Cortex: Executive and cognitive functions.* Oxford: Oxford University Press.

Rolls, E. (2002). A theory of emotion, its functions, and its adaptive value. In R. Trappl, P. Petta, & S. Payr, *Emotions in Humans and Artifacts* (pp. 11-34). Cambridge: MIT Press.

Rolls, E. (2000). Precis: Brain and Emotion. *Behavioral and Brain Science, 23:2*, 178-182.

Rolls, E. (1999). *The Brain and Emotion.* New York: Oxford University Press.

Rolls, E., & Treves, A. (1998). *Neural Networks and Brain Function.* Oxford: Oxford University Press.

Roozendaal, B., & McGaugh, J. (1996). The memory-modulatory effects of glucocorticoids depend on an intact stria terminalis. *Brain Research, 709*, 243-250.

Rorty, A. (1980). *Explaining Emotions.* Berkeley: University of California Press.

Rosen, J., Hamerman, E., Sitcostske, M., Glowa, J., & Schulkin, J. (1996). Hyperexcitability: Exaggerated fear-potentiated startle produced by partial amygdala kindling. *Behavioral Neuroscience, 110*, 43-50.

Rosen, J., Hitchcock, J., Miserendino, M., Falls, W., Campeau, S., & Davis, M. (1992). Lesions of the perirhinal cortex but not of the frontal, medial prefrontal, visual, or insular cortex block fear-potentiated startle using a visual conditioned stimulus. *Journal of Neuroscience, 12*, 4623-4633.

Rosenberg, F., & Ekman, P. (1994) Coherence between expressive and experiential system in emotion. *Cognition and Emotion, 8,* 201-229.

Russel, S., & Wefald, E. (1991). *Do the Right Thing: Studies in Limited Rationality.* Cambridge: MIT Press.

Ryle, G. (1949). *The Concept of Mind.* Chicago: University of Chicago Press.

Sakai, K., & Miyashita, Y. (1991). Neural organization for the long-term memory of paired associates. *Nature, 354*, 152-155.

Salmaso, D., & Denes, G. (1982). Role of the frontal lobes on an attentional task: A signal detection analysis. *Perceptual and Motor Skills, 54*, 1147-1150.

Schachter, S., & Singer, J. (1962). Cognitive, social, and physiological determinates of emotional state. *Psychological Review, 69*, 379-399.

Schacter, D. (1996). *Searching for Memory: The Brain, the Mind, and the Past.* New York: Basic Books.

Schacter, D. (2001). *The Seven Sins of Memory: How the Mind Forgets and Remembers.* New York: Houghton Mifflin.

Schwarz, & Bless. (1991). Happy and mindless but sad and smart? In J. Forgas, *Emotion and Social Judgment* (pp. 55-71). Oxford: Pergamon Press.

Schwarz, N. (2000). Emotion, cognition and decision-making. *Cognition and Emotion, 14*, 433-440.

Schwarz, N. (1990). Feelings as information: Informational and motivational functions of affective states. In E. Higgins, & R. Sorrentino, *Handbook of Motivation and Cognition: Foundations of Social Behavior (Vol 2)* (pp. 527-561). New York: Guilford Press.

Schwarz, N., & Clore, G. (1988). How do I feel about it? The information function of affective states. In K. Fiedler, & J. Forgas, *Affect, Cognition and Social Behavior* (pp. 42-62). Toronto: Hogrefe.

Schwarz, N., & Clore, G. (1983). Mood, misattribution and jugdments of wellbeing: informative and directive functions of affective states. *Journal of Personality and Social Psychology* , 513-523.

Schwarz, N., Bless, B., & Bohner, G. (1991). Mood and persuasion: affective states influence the processing of persuasive communications. In M. Zanna, *Advances in Experimental Social Psychology Vol. 24* (pp. 161-199). San Diego: Academic Press.

Sedikes, C. (1992). Mood as a Determinant of Attentional Focus. *Cognition and Emotion, 6*, 129-148.

Seligman, M. (1970). On the generality of the laws of learning. *Psychological Review, 77*, 406-418.

Seligman, M. (1971). Phobias and Preparedness. *Behavioral Therapy, 2* , 307-320.

Seligman, M., & Hager, J. (1972). *Biological Bounderies of Learning.* New York: Appelton-Century.

Shallice, T., & Burgess, P. (1991). Deficits in Strategy Application Following Frontal Lobe Damage in Man. *Brain, 114*, 727-741.

Shanahan, M. (1997). *Solving the Frame Problem: A Mathematical Investigation of the Common Sense Law of Inertia.* Cambridge: MIT Press.

Shanahan, M. (2006). A cognitive architecture that combines internal simulation with a global workspace. *Consciousness and Cognition, 15,* 433-449.

Shanahan, M., & Baars, B. (2005). Applying global workspace theory to the frame problem. *Cognition, 98,* 157-176.

Shanks, D., & Abelson, R. (1977). *Scripts, Plans, Goals and Understanding.* Hillsdale: Lawrence Erlbaum.

Siebert, M., Markowitsch, H., & Bartel, P. (2003). Amygdala, affect and cognition: Evidence from 10 patients with Urbach-Wiethe disease. *Brain, 126,* 2627-2637.

Simon, H. (1967). Motivational and Emotional Controls of Cognition. *Psychological Review, 74*, 29-39.

Simon, H. (1983). *Reason in Human Affairs.* Stanford: Stanford University Press.

Slomin, A. (2001). Beyond shallow models of emotion. *Cognitive Processing, 2,* 177-198.

Snyder, M., & White, P. (1982). Mood and memories: Elation, depression, and the remembering of the events of one's life. *Journal of Personality, 50*, 149-167.

Solomon, R. (1993). The Philosophy of Emotions. In M. Lewis, & J. Haviland, *Handbook of Emotion* (pp. 3-16). New York: Guilford Press.

Solomon, R. (1976). *The Passions: Emotions and the meaning of life.* Indianapolis: Hackett Publishing Company.

Sperber, D., & Wilson, D. (1996). Fodor's frame problem and relevance theory. *Behavioral and Brain Science, 19*, 530-532.

Sperber, D., & Wilson, D. (1987). Precis of Relevance: Communication and Cognition. *Behavioral and Brain Science, 10*, 697-754.

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and Cognition.* Oxford: Blackwell Publishing.

Stanovich, K. (1999). *Who is Rational: Studies of individual differences in reasoning.* New York: Lawrence Erlbaum Associates.

Stockmeyer, L. (1987). Classifying the computational complexity of problems. *Journal of Symbolic Logic, 52*, 1-43.

Stockmeyer, L., & Chandra, A. (1979). Intrinsically difficult problems. *Scientific American*, 140-157.

Stroop, J. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-662.

Teasdale, J. (1983). Negative thinking in depression: Cause, effect, or reciprocal relationship? *Advances in Behavior Research and Therapy, 5*, 3-25.

Teasdale, J., & Fogarty, S. (1979). Differential effects of induced mood on retrieval of pleasant and unpleasant events from episodic memory. *Journal of Abnormal Psychology 88*, 248-257.

Teasdale, J., & Russell, M. (1983). Differential effects of induced mood on the recall of positive, negative and neutral words. *British Journal of Clinical Psychology, 33*, 889-924.

Teasdale, P., & Barnard, J. (1993). *Affect, Cognition and Change.* Hillsdale: Lawrence Erlbaum.

Teasdale, P., & Fogarty, S. (1979). Differential effects of induced mood on retrieval of pleasant and unpleasant events from episodic memory. *Journal of Abnormal Psychology, 88*, 248-257.

Temple, C. (1997). *Developmental Cognitive Neuropsychology.* New York: Psychology Press.

Tooby, J., & Cosmides, L. (1990). The Past Explains the Present: Emotional Adaptations and the Structure of Ancestral Environments. *Ethology and Sociobiology, 11*, 375-424.

Tranel, D. (1993). The covert learning of affective valence does not require structures in hippocampal or amygdala. *Journal of Cognitive Neurosciences, 5*, 79-88.

Tranel, D., & Hyman, B. (1990). Neuropsychological correlates of bilateral amygdala damage. *Archives of Neurology, 47*, 349-55.

Trappl, R., Petta, P., & Payr, S. (2002). *Emotions in Humans and Artifacts.* Cambridge: MIT Press.

Turner, T., & Ortony, A. (1992). Basic emotions: Can conflicting criteria converge? *Psychological Review, 99*, 566-571.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5*, 207-232.

Ucros, C. (1989). Mood state-dependent memory: a meta-analysis. *Cognition and Emotion, 3*, 139-167.

Velten, E. (1968). A laboratory task for induction of mood states. *Behavioral Research and Therapy, 6*, 473-482.

Voytko, M. (1996). Cognitive functions of the basal forebrain cholinergic systems in monkeys: memory or attention. *Behavioral Brain Research, 75*, 13-25.

Voytko, M., Olton, D., Richardson, R., Gorman, L., Tobin, J., & Price, D. (1994). Basal forebrain lesions in monkeys disrupt attention but not learning and memory. *Journal of Neuroscience, 14*, 167-186.

Wagamaar, W. (1986). A study of autobiographical memory over 6 years. *Cognitive Psychology, 18*, 225-252.

Walsh, D. (2001). *Naturalism, Evolution and Mind.* Cambridge: Cambridge University Press.

Watkins, P., Mathews, A., Williamson, D., & Fuller, R. (1992). Mood-congruent memory in depression: Emotional priming or elaboration? *Journal of Abnormal Psychology, 101*, 581-586.

Watkins, P., Vache, K., Verney, S., & Mathews, A. (1996). Unconscious mood-congruent memory bias in depression. *Journal of Abnormal Psychology, 105*, 34-41.

Watts, F., McKenna, F., Sharrock, R., & Trezise, L. (1986). Colour naming of phobia-related words. *British Journal of Psychology, 77*, 97-108.

Weiskrantz, L. (1986). *Blindsight: A Case Study and Implications.* Oxford: Oxford University Press.

Weiskrantz, L. (1990). Outlooks for blindsight: explicit methodologies for implicit processes. *Proceedings of the Royal Society London B239*, 247-278.

Williams, J., & Scott, J. (1988). Autobiographical memory in depression. *Psyhological Medicine, 18*, 689-695.

Williams, J., Mathews, A., & MacLeod, C. (1996). The emotional Stroop task and psychopathology. *Psychological Bulletin, 120*, 3-24.

Wilson, R., & Keil, F. (2001). *The MIT Encyclopedia of Cognitive Science.* Cambridge: MIT Press.

Wyer, R., & Srull, T. (1994). *Handbook of Social Cognition.* Hillsdale: Lawrence Erlbaum.

Zajonc, R. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist, 35*, 151-175.

Zajonc, R. (1984). On the primacy of affect. *American Psychologist, 39*, 117-123.

Zigmond, M., Bloom, F., Landis, S., Roberts, J., & Squire, L. (1999). *Fundamental Neuroscience.* New York: Academic Press.

Zola-Morgan, S., & Squire, L. (1993). Neuroanatomy of amnesia. *Annual Review of Neuroscience, 16*, 547-563.

Zola-Morgan, S., Squire, L., Amaral, D., & Suzuki, W. (1989). Lesions of perirhinal and parahippocampal cortex that spare the amygdala and hippocampal formation produce severe memory impairment. *Journal of Neuroscience, 9*, 4355-4370.

Zola-Morgan, S., Squire, L., Clower, R., & Rempel, N. (1993). Damage to perirhinal cortex but not the amygdala exacerbates memory impairment following lesions to the hippocampal formation. *Journal of Neuroscience, 12*, 1582-2596.