Brit. J. Phil. Sci. **53** (2002), 497–509

The Search Hypothesis of Emotion

Dylan Evans

ABSTRACT

Many philosophers and psychologists now argue that emotions play a vital role in reasoning. This paper explores one particular way of elucidating how emotions help reason which may be dubbed 'the search hypothesis of emotion'. After outlining the search hypothesis of emotion and dispensing with a red herring that has marred previous statements of the hypothesis, I discuss two alternative readings of the search hypothesis. It is argued that the search hypothesis must be construed as an account of what emotions typically do, rather than as a definition of emotion. Even as an account of what emotions typically do, the search hypothesis can only be evaluated in the context of a specific theory of what emotions are.

1 *Introduction*
2 *The search hypothesis of emotion*
3 *A red herring: the frame problem*
4 *The search problem*
5 *Two readings of the search hypothesis*
6 *Two final remarks*
7 *Conclusion*

1 Introduction

How do the emotions affect reason? This question has been debated since Plato, who proposed what may be called 'the negative view of emotion' (Evans [2001]). According to the negative view, emotions usually affect reasoning for the worse. To the extent that humans can free themselves of emotion, therefore, they can become more rational. Until recently, most philosophers and psychologists have tended to agree with Plato on this matter.

In the past couple of decades, however, a growing number of thinkers have challenged the traditional consensus (see, for example, de Sousa [1987]; Frank [1988]; Damasio [1994]; Elster [1999]; Evans [2001]). They argue for what may be called 'the positive view of emotion'. According to the positive view,

emotions usually affect reasoning for the better. The positive view suggests that, other things being equal, humans will be *less* rational to the extent that they lack emotion.

The positive view of emotion is mere hand-waving, of course, unless we can spell out exactly *how* emotions are supposed to aid rationality. The various proponents of the positive view have risen to this challenge in a number of different ways. In this paper, I will examine one particular way of elucidating how emotions help reason which, for want of a better name, I will call 'the search hypothesis of emotion'.

In the next section, I provide a brief sketch of the search hypothesis of emotion. In the following section, I dispense with a red herring that has marred previous statements of the search hypothesis. I then set out the model of rational decision-making that is implicit in the search hypothesis. I go on to distinguish two possible ways of reading the search hypothesis, and conclude that the hypothesis can only be evaluated in the context of a specific theory of what emotions are.

2 The search hypothesis of emotion

Ever since Thomas Hobbes denied that reason can fix our ends or desires, an instrumental conception of rationality has dominated Western thought. According to this view, reason is, in Hume's famous phrase, a mere 'slave of the passions'. It has no motivating force, and is limited to finding the right means to attain the ends which are 'given' by the emotions.

This view has the merit of providing a neat explication of the reason-emotion distinction. Perhaps, however, it is rather *too* neat. Not only does it rule out the possibility of asking people to justify their ends—*de gustibus non disputandum est*—but it also, conversely, rules out the possibility of any emotional influence in rational choice. First, one supposes, emotions assign a subjective utility to each end, without any help from reason; then, whenever we have to make a decision or choose between a number of actions, reason calculates the expected utility for each action and automatically selects that action with the highest expected utility. Reason is thus reduced to computation, in true Hobbesian fashion.

One problem with this view of reason is that it tells us nothing about how we predict the possible outcomes of each action. In rational choice theory, the range of possible outcomes (and the conditional probabilities relating outcomes to actions) are assumed to be given, just like the subjective utilities of the outcomes. The latter are supposed to be given by the emotions; but what about the former? We cannot simply appeal to the agent's beliefs, since belief-fixation is supposed to be a rational process itself, so that simply leads to an infinite regress.

An example may serve to make this problem a bit clearer. Suppose that I wish to arrange an appointment with my doctor, and he suggests two alternative days—say, next Monday or next Tuesday. If I am a rational agent (according to the instrumental view of rationality), I will calculate the expected utility of going to the doctor's on each of these days and choose that which has the higher. Before I can do this, I need to assign a conditional probability $Pr\ (\omega|x)$ to each outcome, where ω is the outcome and x is the action. Before I can do *that*, however, I must first of all list all the possible outcomes. And, to borrow a phrase, there's the rub; for who knows 'what dreams may come, when we have shuffled off this mortal coil'—or even what might ensue from arranging to see the doctor on Monday? If I have already agreed to take my kids out to the zoo on Monday, then arranging the appointment for Monday would mean having to reschedule the trip to the zoo. Rescheduling the trip to the zoo would be one outcome of arranging the appointment for Monday, but why stop there? Why not also consider the possible consequences of *that*? My kids might get annoyed with me; I might tell them off for being intransigent; this might lead them to be more flexible, which might help them to be happier later in life . . . and so on, *ad infinitum*. Since the relation *being an outcome of* is transitive, it follows that all these are also outcomes of arranging the appointment for Monday. So even for a simple decision like arranging an appointment with the doctor, the set of possible outcomes for each action is in principle unbounded. Therefore, listing the possible outcomes of any given action is a potentially endless task. Yet, if I am to make a decision, I must stop listing outcomes at some point. Looking before you leap is all very well, but the point is to leap. At some point, you must stop thinking, and start acting. Let us call this 'Hamlet's problem'.[1]

Note that we cannot solve Hamlet's problem by recourse to rational decision theory by, say, assigning a utility function to thinking-time. That would simply lead to an infinite regress, for it would mean that, before making any decision, we would have to decide how much time to spend on making it. But before we could decide *that*, we would first have to decide how much time to spend on *that* decision, and so on, *ad infinitum*. So, it seems that reason cannot even get off the ground unless some prior, non-rational procedure is used to delimit the time for decision-making, or the range of consequences to be considered, or both (or perhaps they just amount to the same thing).

According to the search hypothesis of emotion, this non-rational procedure for delimiting the range of consequences to be considered in a rational

[1] Jerry Fodor ([1987], p. 140) used this term to describe a similar, though more general kind of problem.

decision-process is governed by the emotions. The emotions, on this account, play more than one role in rational choice. Not only do they assign a subjective utility to each outcome; they also delimit the range of outcomes to be considered.

The search hypothesis of emotion is a relatively recent idea. The first person to have put it forward explicitly is, as far as I can tell, the philosopher Ronald de Sousa. In his book *The Rationality of Emotion* ([1987], p. 195), he argued that emotions limit 'the range of information that the organism will take into account, the inferences actually drawn from a potential infinity, and the set of live options among which it will choose'. In his article on emotion for the *Blackwell Companion to the Philosophy of Mind*, de Sousa ([1994], p. 276) re-iterates the idea, stating that 'in the process of rational deliberation itself, [emotions] render salient only a tiny proportion of the available alternatives and of the conceivably relevant facts'.

More recently, similar views have been expressed by a number of evolutionary psychologists. Timothy Ketelaar and Peter Todd, for example, have suggested that 'specific emotions might help to solve the problem of what information to attend to in specific environmental circumstances' (Ketelaar and Todd [2001], p. 194). Spelling out the idea in more detail, they go on to state:

> Thus, when the future outcome of various courses of action cannot be objectively 'calculated', often because there are simply too many plausible consequences to consider (i.e., the frame problem), it may pay to 'let your emotions be your guide' in selecting which course of action to pursue. ([2001], p. 194)

So far so good, or so it seems. The search hypothesis of emotion seems, on the face of it, to be a neat way of fleshing out the claim that emotions play a positive role in aiding rational choice. However, when one probes it a little, it turns out to have some serious problems. In a short while, I will look at these problems, but first I want to dispense with a red herring.

3 A red herring: the frame problem

In the passage just cited, Ketelaar and Todd refer to 'the frame problem'. De Sousa does, too. In the article on emotion in the Blackwell *Companion* ([1994], p. 276), he makes the following claim:

> [T]he number of goals that it is logically possible to posit at any particular time is virtually infinite, and the number of possible strategies that might be employed in pursuit of them is orders of magnitude larger. Moreover, in considering possible strategies, the number of consequences of any one strategy is again infinite, so that unless some drastic preselection can be effected among the alternatives their evaluation could

> never be completed. This gives rise to what is known among cognitive scientists as the 'Frame Problem': in deciding among any range of possible actions, most of the consequences of each must be eliminated from consideration a priori, i.e. without any time being wasted on their consideration. That this is not as much of a problem for people as it is for machines may well be due to our capacity for emotions.

So it seems that what I have called 'the search hypothesis of emotion' could be re-phrased in something like the following terms: *emotions help humans to solve the frame problem*. This is, in fact, more or less how Ketelaar and Todd describe their view of emotion: 'a simple emotion mechanism might help us surmount the frame problem' (Ketelaar and Todd [2001], p. 185). So, might we not refer to the search hypothesis of emotion instead as, say, 'the *frame* hypothesis of emotion'?

The reason I have chosen *not* to coin such a phrase, and to talk instead about 'the *search* hypothesis of emotion', is that there is no consensus about what exactly the frame problem *is*. The phrase was coined by John McCarthy in 1969 to denote a difficulty with a particular formalism that he and Patrick Hayes had developed for temporal reasoning in computer programs (McCarthy and Hayes [1969]). As time went on, however, the term was increasingly used by philosophers in more general ways. By the early 1980s, for example, Jerry Fodor was using the term to designate the problem of how to program digital computers to make good abductive inferences (Fodor [1983]). Patrick Hayes, who co-authored the original paper with John McCarthy in which the term first appeared, has protested that these broader definitions of the frame problem are muddying the waters. For Hayes, these broader definitions are unconnected with the original frame problem, which is simply a technical problem with a particular kind of formalism (Hayes [1987]). Philosophers such as Fodor and Dennett disagree, and the debate has raged through several edited collections of essays (Pylyshyn [1987]; Ford and Pylyshyn [1996]). My reason for not wishing to speak of 'the frame hypothesis of emotion' is pragmatic; I do not wish to get bogged down in endless disputes about what the frame problem 'really is'. It is not necessary to resolve these disputes before addressing the search hypothesis of emotion.

When de Sousa, Ketelaar and Todd refer to the frame problem, they all have in mind a much broader definition of the frame problem than that which is common in artificial intelligence. The problem that they think emotions solve is that of when to stop listing what the possible consequences of an action will be. This is a long way from what most researchers in artificial intelligence understand by the frame problem. If we wish to label the problem identified by de Sousa, Ketelaar and Todd with some technical term drawn from the jargon of artificial intelligence, we should probably call

it 'the search problem'. In the following section, I explain what the search problem is.

4 The search problem

One of the conceptual mainstays of artificial intelligence is the insightful idea, first proposed by Allen Newell, Cliff Shaw and Herbert Simon, of conceiving problem-solving as a kind of *search* (Newell and Simon [1976]). Imagine a state space consisting of many potential solutions to a problem; finding the right solution then involves searching through the state space. Now, if we had to generate the state space of potential solutions in entirety in advance of the search, we would often be stuck with Hamlet's problem, since the state space of many problems is infinite. To get round this, Newell, Shaw and Simon proposed that the state space can be expanded step by step, with the current state being tested to see whether or not it qualifies as an acceptable solution before the next expansion. If the current state space *does* qualify as an acceptable solution, the search is terminated and the problem is solved; otherwise, the state space is expanded by generating another state. The new state is then tested, and so on.

It is this step-by-step process of generating and testing that makes all the difference between the idealised models of rational choice theory and the bounded rationality of artificial intelligence systems. Instead of trying to figure out all the consequences of an action before evaluating them, the search procedure generates one consequence at a time and then evaluates it before going on to generate another one. The test it uses to evaluate the consequence must be a simple algorithm that returns a quick 'yes' or 'no'. If the answer is positive, a solution has been found and the search stops. If the answer is negative, the search continues.

This simple process can solve Hamlet's problem—or the problem of arranging an appointment to see the doctor. Instead of imagining all the consequences of each action, and then assigning utilities and conditional probabilities to each of the consequences before finally calculating the overall expected utility of each action—a potentially infinite task, as we have seen—I simply imagine the consequences of each action one at a time, and apply some test. The test might be something simple such as meeting some aspiration level. I might, for example, think of the first consequence of arranging to see the doctor on Monday—re-arranging the trip to the zoo—and ask whether that would be too much bother for me. If it is not, I arrange my appointment for Monday. Otherwise, I go on to consider other consequences of this action and/or the consequences of other actions.

Now, what makes this way of deciding rational is that—providing the right kind of test is used, and the right search strategy[2]—decisions made on the basis of a search process like this can regularly lead to decisions that closely approximate those made by an ideally rational agent operating according to classical decision theory. This, at least, is what proponents of bounded rationality maintain, following the work of Herbert Simon (Simon [1955]). Picking the right test to use for the given problem at hand, and the right search strategy, is thus crucial. This is the search problem.

The search hypothesis of emotion can now be re-stated; it is the claim that emotions enable humans to solve the search problem. Emotions prevent us from getting lost in endless explorations of potentially infinite search spaces by providing us with both the right kind of test and the right kind of search strategy for each kind of problem we must solve. Only when emotions fail us do we end up in Hamlet's situation, suffering from a severe case of analysis paralysis.

5 Two readings of the search hypothesis

There are two ways in which the search hypothesis can be construed. The first is that the hypothesis is a metaphysical claim about what emotions really *are*; the second is an empirical claim about what emotions typically *do*. There is at least one good reason, however, why the first reading is not tenable. This is that the search hypothesis, when construed as a definition of emotion, becomes vacuous. If it were intended as a definition, of course, it would imply a functionalist account of mental states, since it would define emotions by what they typically do. And what emotions typically do, according to the search hypothesis, is help us delimit the range of possible consequences to be considered in any rational decision process. If we had an independent idea about what such things might look like, then the hypothesis would be an interesting claim—perhaps a proposal to identify emotions with this antecedently discovered class of mechanisms. However, the sad fact is that we have *no idea* how such mechanisms might work. Despite over forty years of hard work, researchers in artificial intelligence are still incapable of designing mechanisms that can delimit, in advance, the range of consequences to be considered in such a way that decisions can be both rapid and rational.

[2] A 'search strategy' is defined by the criterion used to decide which state to expand first. If we represent the search process as building up a search tree, then the search strategy tells us which node of the tree to expand first. A 'breadth-first' search, for example, expands all the nodes at one level before expanding any of the nodes at the next level. A 'depth-first' search, on the other hand, always expands one of the nodes at the deepest level of the tree; only when the search hits a dead end (a nongoal node with no expansion) does the search go back and expand nodes at shallower levels. More informed kinds of search strategy—ones that use problem-specific knowledge—are also possible (Russell and Norvig [1995]).

There are, it is true, a number of interesting proposals about specific heuristics that can make rapid, rational decisions *in a given domain*, but this merely pushes the explanatory burden back one stage, to the question of how a given real-world problem is mapped onto one of the system's domains. And nobody has any idea about how such mappings can be achieved in a rapid and rational way. Invoking 'heuristics' in response to the search problem is just passing the buck. But, then, so is invoking 'emotions'. Since we have no idea about how to solve the search problem, then to claim that emotions are what help us solve it is simply to exchange one mystery for another. What purports to be an explication of the concept of emotion turns out to be just an obfuscation.

I noted previously that the search hypothesis can be construed either as a claim about what emotions are, or as a claim about what emotions do. The objection just raised shows that, as a claim about what emotions are, the search hypothesis is vacuous. But this objection is not fatal to the search hypothesis; after all, it is not the task of every hypothesis about emotion to tell us what emotions are. Some accounts of emotion purport to answer a rather different question; namely, how do emotions interact with reason? Perhaps the search hypothesis is better construed, then, as a claim about what emotions do.

This reading seems much more promising, though with one major caveat; construed as a claim about what emotions typically *do*, the search hypothesis remains vacuous unless we have some independent notion of what emotions *are*. Only then would the claim that emotions help us solve the search problem avoid substituting one mystery for another. In other words, the search hypothesis can only properly be evaluated in the context of a specific theory of emotion. We can ask, for example, whether the search hypothesis provides a good account of what emotions *do* in the context of Antonio Damasio's theory of what emotions *are*, or of Ronald de Sousa's account. But we cannot ask whether the search hypothesis provides a good account of what emotions do in the absence of such a theory. It is to their credit, then, that both of these authors have advocated the search hypothesis only in the context of their original accounts of what emotions are.

Damasio [1994], for example, argues that emotions—or at least the qualitative, conscious experience of emotions that some philosophers prefer to call 'feelings'—are 'somatic markers'. That is, they are bodily sensations, whether of a visceral or a nonvisceral kind, that are summoned up by particular thoughts or mental images. When the search hypothesis is advanced in the context of this particular theory of emotion, the following account of how emotions help rational decision-making emerges. When one starts thinking about the possible consequences of a decision, gut feelings may be triggered by particular images associated with certain consequences.

If negative, the gut feeling 'forces attention on the negative outcome to which a given action may lead, and functions as an automated alarm signal which says: Beware danger ahead if you choose the option which leads to this outcome' (Damasio [1994], p. 173). The signal may lead one to reject this option straight away, without the need for consideration of further consequences, and thus allow one to choose from among fewer alternatives.

Damasio backs up his claims with empirical data drawn from patients with various kinds of frontal lobe damage. These patients are strangely Hamlet-like. When confronted with a decision of even the most trivial nature, they may loose themselves in endless musings about the consequences of each possible action, with the result that the decision itself is postponed indefinitely. After one consultation, for example, with a patient with frontal lobe damage, Damasio asked him when he would like to arrange his next appointment:

> 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 behaviour that ensued, which was witnessed by several investigators, 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 concerning a simple date. [. . .] 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 without pounding on the table and telling him to stop, but finally we did tell him, quietly, that he should come on the second of the alternative dates. His response was equally calm and prompt. He simply said: 'That's fine'. Back the appointment book went into his pocket, and then he was off. (Damasio [1994], p. 193–4)

When these clinical data are combined with Damasio's claim that patients with frontal lobe damage are *emotionally* impaired, this lends some *prima facie* support to the claim that most of us are able to avoid analysis paralysis only because we are emotionally intact, and thus support the search *hypothesis when read in conjunction with* Damasio's theory of emotions as somatic markers.

Damasio's account leaves many questions unanswered. What, for example, is the patient really failing to do? Suppose, for the sake of argument, that Damasio's patient begins to construct a search tree in the typical manner. First, he starts by expanding the initial node of the tree into two states; go on Monday, and go on Tuesday. How he then goes about building the tree up from there depends on his search strategy. Most of us would soon stop the search, having found some particular node that 'tips the balance' towards one

or other of the initial two nodes. The patient does not stop; but *why* not? According to Damasio's theory, he does not stop because he does not generate an appropriate somatic marker. But the theory does not say exactly why this failure occurs.

There are a number of possible answers. Perhaps the patient is simply failing to apply any test to each new node of the search tree as he generates it. Or perhaps he is applying a test, but not the right one. A third possibility is that he is applying the right test, but something is malfunctioning in the test mechanism. The somatic marker theory of emotion is consistent with all of these possibilities. Does the emotional deficit of Damasio's patient involve a failure to call up a test to be applied to each new node of the search tree as it is generated, so that no test is applied at all? Or does the patient simply apply the wrong test, so that no appropriate somatic marker can be generated? Or, as a third possibility, might the patient be applying the right test, but the test mechanism fails to trigger any somatic marker? Of course, this ambiguity is hardly grounds for dismissing Damasio's theory altogether. It simply means that Damasio's theory is couched at a certain level of generality, and is consistent with a number of more detailed specifications. It does, nevertheless, provide at least a general way of fleshing out the search hypothesis and turning it into a testable theory of the relation between reason and emotion.

6 Two final remarks

Before concluding, I wish to make two final remarks. The first concerns the fact that there is, at present, no consensus about what emotions are. Indeed, many philosophers now doubt that there is any good single definition of emotion. A consensus is emerging to the effect that the very diversity of phenomena referred to as 'emotions' precludes any single definition of the term. Paul Griffiths, for example, argues that 'the general category of emotion subsumes three different kinds of psychological state', and concludes that 'the general concept of emotion has no role in any future psychology' ([1997], pp. 245, 247). Jon Elster concurs. After suggesting that the emotions 'may not be a coherent and theoretically useful concept', he goes on to speculate that 'the unruly category of "the emotions" encompasses several, internally homogeneous classes of phenomena' (Elster [1999], p. 241). In a similar vein, Aaron Ben-Ze'ev ([2000], p. 3) states that 'No single essence is necessary and sufficient for all emotions', and Ronald de Sousa ([1994], p. 270) makes similar gloomy remarks about 'the sheer variety of phenomena covered by the word "emotion" '.

The very heterogeneity of the emotions might seem to pose a further problem for the search hypothesis, construed as a claim about what emotions

typically do. For if it is true that the term 'emotion' refers to a variety of natural kinds rather than a single kind, construing the search hypothesis as a claim about what emotions do may appear to reduce the hypothesis to the trite claim that a heterogeneous variety of mechanisms help us to solve the search problem. Once again, however, this is only a problem for the search hypothesis when it is considered in isolation from the context of any specific theory of emotion, or—more pertinently, if no single theory of emotion is to be found—any specific theory of a particular *class* of emotions. The search hypothesis is not necessarily committed to the view that there is some single definition of emotion; all that it requires is that, if such a definition can be found, whatever that definition turns out to be, it cannot be (on pain of circularity) that emotions are things that help us solve the search problem.

The second of my final remarks concerns the relationship between the search hypothesis and the computational theory of mind. From the discussion so far, it may appear as if the search hypothesis, however it is read, is predicated on the assumption that humans make decisions in a way that closely resembles the way that current systems in artificial intelligence do. Only if the decision-making process in humans consists of gradually exploring a search space is it possible to claim that emotions help humans to avoid doing so indefinitely. This, in turn, seems to pre-suppose the computational theory of mind. Indeed, part of the appeal of the search hypothesis of emotion for some may be that it promises a coherent computational account of emotion—something that many cognitive scientists are increasingly concerned to provide.

Despite these appearances, the search hypothesis is not committed to the computational theory of mind. Although the metaphor of 'exploring a search space' by 'expanding the search tree one node at a time' is couched unambiguously in the jargon of artificial intelligence, one can still think of it as a useful way of describing certain thought processes without being a fully paid-up, card-carrying computationalist. Damasio's account of the search hypothesis is a case in point. From his account of the role that somatic markers play in certain decision-making processes, it is clear that he construes some decision-making at least as consisting of something very like the search process described. It is also clear, however, that he takes this as a *phenomenological* description; it is meant as an account of the verbal thoughts and mental images that pass through the subject's *conscious awareness*. As such, it is agnostic about the precise nature of the unconscious processes that support this conscious superstructure. This unconscious substrate may well consist of rule-governed transformations of syntactic representations, as the computational theory of mind has it, but then again it may not. The fact that Damasio can make use of the search hypothesis without committing himself to the computational theory of mind, and do so

coherently, is enough to refute the view that the search hypothesis stands or falls with computationalism.

7 Conclusion

The search hypothesis of emotion is the claim that emotions enable humans to solve the search problem. In other words, emotions prevent us from getting lost in endless explorations of potentially infinite search spaces by providing us with both the right kind of test and the right kind of search strategy for each kind of problem we must solve. The search hypothesis thus offers an account of the relationship between emotions and reason, according to which emotions play a positive role in aiding reason to make good decisions. However, the hypothesis is vacuous unless we have some independent account of emotion to flesh it out. It can only be assessed, then, in the context of some particular theory or other about what emotions are.

Acknowledgements

I would like to thank Nicholas Humphrey and two anonymous referees for their invaluable comments on an earlier version of this paper.

Biomimetics Group
Department of Mechanical Engineering
University of Bath
Bath BA2 7AY
UK
d.evans@bath.ac.uk

References

Ben-Ze'ev, A. [2000]: *The Subtelty of Emotions*, Cambridge, MA: MIT Press.

Damasio, A. R. [1994]: *Descartes' Error: Emotion, Reason and the Human Brain*, London: Papermac.

de Sousa, R. [1987]: *The Rationality of Emotion*, Cambridge, MA: MIT Press.

de Sousa, R. [1994]: 'Emotion', in S. Guttenplan (*ed.*), 1994, *A Companion to the Philosophy of Mind*, Oxford: Blackwell, pp. 270–6.

Elster, J. [1999]: *Alchemies of the Mind: Rationality and the Emotions*, Cambridge: Cambridge University Press.

Evans, D. [2001]: *Emotion: the Science of Sentiment*, Oxford: Oxford University Press.

Fodor, J. A. [1983]: *The Modularity of Mind: An Essay on Faculty Psychology*, Cambridge, MA & London: MIT Press.

Fodor, J. A. [1987]: 'Modules, frames, fridgeons, sleeping dogs, and the music of the spheres', in Z. W. Pylyshyn (*ed.*), *The Robot's Dilemma: the Frame Problem in Artificial Intelligence*, Norwood, NJ: Ablex, pp. 139–49.

Ford, K. M. and Pylyshyn, Z. W. (*eds*) [1996]: *The Robot's Dilemma Revisited: The Frame Problem in Artificial Intelligence*, Norwood, NJ: Ablex.

Frank, R. H. [1988]: *Passions within Reason: the Strategic Role of the Emotions*, New York: Norton.

Griffiths, P. E. [1997]: *What Emotions Really Are: The Problem of Psychological Categories*, Chicago & London: University of Chicago Press.

Hayes, P. [1987]: 'What the frame problem is and isn't', in Z. W. Pylyshyn (*ed.*), *The Robot's Dilemma: the Frame Problem in Artificial Intelligence*, Norwood, NJ: Ablex, pp. 123–37.

Ketelaar, T. and Todd, P. M. [2001]: 'Framing our thoughts: ecological rationality as evolutionary psychology's answer to the frame problem', in H. R. Holcomb III (*ed.*), *Conceptual Challenges in Evolutionary Psychology: Innovative Research Strategies*, Norwell, MA: Kluwer Academic Publishers, pp. 179–211.

McCarthy, J. and Hayes, P. J. [1969]: 'Some philosophical problems from the standpoint of artificial intelligence', in B. Meltzer and D. Michie (*eds*), *Machine Intelligence 4*, Edinburgh: Edinburgh University Press, pp. 463–502.

Newell, A. and Simon, H. A. [1976]: 'Computer science as empirical inquiry: symbols and search', *Communications of the Association for Computing Machinery*, **19**, pp. 113–126.

Pylyshyn, Z. (*ed.*) [1987]: *The Robot's Dilemma: The Frame Problem in Artificial Intelligence*, Norwood, NJ: Ablex.

Russell, S. and Norvig, P. [1995]: *Artificial Intelligence: A Modern Approach*, Upper Saddle River, NJ: Prentice Hall.

Simon, H. [1955]: 'A behavioural model of rational choice', *Quarterly Journal of Economics*, **69**, pp. 99–118.