Ratio (new series) XVIII 3 September 2005 0034–0006

EASY'S GETTIN' HARDER ALL THE TIME: THE COMPUTATIONAL THEORY AND AFFECTIVE STATES

Jason Megill and Jon Cogburn

Abstract

We argue that A. Damasio's (1994) Somatic Marker hypothesis can explain why humans don't generally suffer from the frame problem, arguably the greatest obstacle facing the Computational Theory of Mind. This involves showing how humans with damaged emotional centers are best understood as actually suffering from the frame problem. We are then able to show that, paradoxically, these results provide evidence for the Computational Theory of Mind, and in addition call into question the very distinction between easy and hard problems in the contemporary philosophy of mind.

In my view, the frame problem is a lot of what makes cognition so hard to understand . . . cognitive science without the frame problem is Hamlet without anybody much except Polonius.

J. A. Fodor (2000, p. 42)

We begin by posing two important questions: (1) what makes the human mind free from the so-called 'frame problem' that besets all extant attempts at artificial intelligence, alternatively, how did nature solve this problem for us, and (2) are *all* of us in fact free from the frame problem? By examining the frame problem as a problem besetting humans as well as machines, we are able to draw two implications for attempts to understand cognition and consciousness: (1) the Computational Theory of Mind is much more plausible than generally assumed, and (2) a complete explanation of our cognitive abilities requires appeal to the phenomenology of emotion.

In section one, we explain the frame problem and how it bedevils contemporary attempts to computationally model cognitive abilities. In section two, we outline an intriguing attempt to explain human emotion from the realm of neuroscience: A. Damasio's (1994) Somatic Marker hypothesis. In section three, drawing on our discussion of the frame problem and Damasio,

we answer the two questions posed above. In section four, we sketch the surprising implications of our argument.

I. The frame problem

D. C. Dennett (1984) asks us to consider three hypothetical robots: R1, R1D1, and R2D1.

R1's designers assigned R1 what looked to be a relatively simple task: R1's battery was placed on a wagon in a room containing a time bomb, and R1 was ordered to retrieve its battery before the bomb detonated. R1 deftly pulled the wagon from the room, but unfortunately, 'the bomb was also on the wagon' (Dennett 1984, p. 129). R1 had a fatal flaw: it could not recognize many of the obvious implications of its actions.

R1D1, designed with the lessons learned from R1 in mind, could 'recognize not just the intended implications of its acts, but also the implications about their side-effects . . .' (Dennett 1984). Placed in the same scenario as R1, it set about considering all of the possible implications of the act of pulling the wagon from the room. Just as the robot determined that pulling the wagon from the room would leave the color of the room's walls unaltered, the bomb exploded. R1D1, like its predecessor, had a fatal flaw: the problem wasn't that R1D1 couldn't recognize the implications of its actions, rather, it could recognize *too many* implications of its actions. R1D1 had no feel for relevance or irrelevance, importance or triviality: it simply deduced implications at random, until the bomb exploded.

Determined, the researchers designed R2D1, teaching it not only how to recognize the implications of its actions, but also how to recognize which implications are salient and which aren't, so the robot could in turn ignore irrelevant implications. When placed in the room, the robot sat motionless, calculating implication after implication. Dismayed, the researchers urged R2D1 to act, to which R2D1 responded: 'I am . . . I'm busily ignoring some thousands of implications which I have deemed to be irrelevant' (Dennett 1984, p. 130). The bomb exploded. These examples point to what has become known in artificial intelligence as the 'frame problem.'[1] The frame problem, as Dennett's examples

[1] We have chosen to introduce the frame problem via Dennett's (1984) discussion for three principle reasons: (1) Dennett's discussion is a classic in the literature, (2) Dennett's

illustrate, concerns the question as to how a machine intelligence can be taught to determine the relevant consequences of a given act in a sufficiently efficient manner. This definition of the frame problem, however, can be seen as being too narrow. The frame problem can also be cast in a broader light: how can *any* agent access the relevant knowledge needed to cope with *any* circumstance? In Fodor's words:

> "The frame problem" is a name for one aspect of the question of how to reconcile a local notion of computation with the apparent holism of rational inference; in particular, with the fact that information that is relevant to the optimal solution of an abductive problem can, in principle, come from anywhere in the network of one's prior epistemic commitments. (Fodor 2000, p. 42)

In short, the frame problem can be seen as a cluster of questions, all of which revolve around the question of how an agent determines relevance: (1) how does an agent determine what the relevant objects in its environment are, (2) how does an agent recognize what the relevant implications of any given action are, (3) how does an agent efficiently access what specific pieces of knowledge in a vast knowledge-base are relevant to a given situation? In short, *how does an agent determine relevance?*

Finally, far from being an esoteric problem that plagues only A.I. specialists and cognitive scientists, the frame problem can also be seen as a philosophical problem,

> a new, deep epistemological problem – accessible in principle but unnoticed by generations of philosophers – brought to light by the novel methods of A.I., and still far from being solved. (Dennett 1984, p. 130)[2]

We now explain how nature solved the frame problem for us.

discussion is apt to be of more interest to philosophers than most of the more technical discussions of the frame problem in the artificial intelligence literature, and (3) Dennett's discussion highlights precisely the aspects of the frame problem that are relevant to our arguments. For the reader who desires a more thorough, or more recent, discussion of the frame problem, see Pylyshyn (1987), Ford and Pylyshyn (1996), Shanahan (1997), or Fodor (2000).

[2] In calling the frame problem an 'epistemological' question, perhaps Dennett has something like the following in mind: given all that we know, and given a specific situation, how do we *know* what specific pieces of knowledge in the knowledge base are needed to cope with the situation we are in?

II. Damasio's Somatic Marker hypothesis

A. Damasio's (1994) Somatic Marker hypothesis implies that the emotions play a key role in giving rise to and shaping intelligent behavior. In attempting to account for our intelligent behavior, one can differentiate two distinct questions: (1) how do we decide or determine *what* to reason about, and (2) given the relevant factors we need to take into account, *how* do we reason? The first question involves determining what the content of our reasoning should be, while the second concerns the form of our reasoning, abstracted away from content.

In cognitive science, the common answer to the second question is that we reason via logical inference; this answer is the core tenet of the Computational Theory of Mind. Logical inferences are something that computers can do well, and hence are easily accounted for by the Computational Theory. There are, of course, difficulties with this answer, such as those that Sutherland's (1992) work raises,[3] but it seems a reasonable place to start.[4] It is the first question that proves utterly intractable for the Computational Theory of Mind (as we saw in our discussion of the frame problem), and it is the first question that Damasio's hypothesis might be able to answer.

Imagine that you are facing an important life decision; for example, you are trying to decide what career you wish to pursue (Damasio 1994). Initially, you are faced with a staggeringly large, and hence unmanageable, list of possible choices. Say that each possible career is associated with a certain mental representation; the career of soldier, for example, calls to mind a certain representation involving guns, tanks, and marching.

Damasio (1994) holds that these representations, or at least some or many of them, will also be associated with what he terms a 'somatic marker.' A somatic marker is a neurophysiological response that, through learning, comes to be associated with a given mental representation. A somatic marker, as a physiological response, will also lead to the visceral experience of an emotion, an emotional quale. These visceral responses help 'edit' the vast list of possible careers. Perhaps you are afraid of driving, in which

[3] That is, many are inept at utilizing probability theory, and often reason fallaciously.

[4] For problems with the notion that we reason via First-Order logic, also see Johnson-Laird's (1983) classic critique; see Gardner (1985) for a clear introductory discussion of this issue.

case the mental representations of many careers (truck driver, cab driver, etc) will be associated with an unpleasant 'gut feeling,' a negative emotional quale that prompts you to discard all careers that involve driving from the list. Further, imagine that money has a very positive somatic marker, in which case those careers that you associate with money will not only be kept on the list, but will have a higher probability of entering your consciousness in the first place, and kept in consciousness longer once there. Eventually, the list is shortened to a manageable length (see Damasio 1994, especially pp. 173–75).[5]

Once the emotions have played their role, the door is opened to the use of rational inference, but without the emotions, rational inference is useless in the face of the bewilderingly large list of possible courses of action facing an agent at any time. So, once you decide that you might want to become, say, a chiropractor, you can then reason that you should go to college, raise the funds for college, take the SATs and so on.

Damasio's (1994) hypothesis can be clarified via appeal to a concrete example, the case of Elliot, one of the patients that inspired Damasio's hypothesis.[6] Elliot had a superior-level IQ and a successful career and was a responsible husband and father. Then, Elliot began to have migraines, and his personality began to change. It was discovered that Elliot had a brain tumor in his frontal lobe.

The brain tumor was successfully removed (and brain tumors rarely reoccur), but some of the damaged tissue in the frontal lobe had to be removed. Elliot suffered no blatantly recognizable complications; for example, his movement, speech, memory, and knowledge-base were intact. However, the changes in Elliot's personality remained.

Formally an excellent employee, Elliot now seemed irresponsible. He needed 'prompting' to go to work (Damasio 1994, p. 36).

[5] We should note that Damasio (1994) uses a terminology that might strike some as odd. *Damasio* uses the word 'emotion' *only* for the neurophysiological aspect of emotion, and he uses the word 'feelings' for emotional qualia. We, however, will not adopt Damasio's perhaps unintuitive terminology. We use 'emotion' as a blanket term for both neurophysiological responses and emotional qualia; when we wish to refer to emotional qualia specifically, we use 'emotional qualia.'

[6] For clarity and vividness of exposition, we continue our introduction of Damasio's theory via a single case study. There are many other cases similar to Elliot in Damasio's work (1994) (see also Damasio (1999)). Further, there is a massive literature on frontal lobe damage, containing discussions of cases similar to Elliot's (see LeDoux (1996), Panksepp (1998)).

Once there, 'he was unable to manage his time properly; he could not be trusted with a schedule' (Damasio 1994, p. 36). He frequently became lost in the minor details of his work, in the trivial aspects of the task at hand, often spending hours brooding over an irrelevancy. For example, while sorting documents, Elliot would begin to read a document with a carefulness that bordered on absurdity, perhaps becoming engrossed for the entire day. As a consequence of his behavior, of his inability 'to perform an appropriate action when it was expected,' Elliot 'could no longer adequately perform goal-oriented activity' (Damasio 1994, pp. 36–37). In short, while the tumor didn't affect Elliot's knowledge-base, it wrecked havoc with his decision-making ability. Elliot was eventually fired, he divorced, remarried and divorced again, was fired from several other jobs, and pursued ill-conceived business ventures, all to the dismay of his friends and family.

Interestingly, Elliot's other symptoms were accompanied by a severe lack of emotional qualia; for Elliot, most experiences lost their affective component. He could speak of his tragedy with an unsettling detachment. He rarely displayed anger or felt pleasure. In short, Elliot's phenomenological experience of the world was by and large devoid of emotion. Intriguingly, Damasio was led to link Elliot's inability to make decisions and effectively cope with his environment with this loss of emotional qualia. The emotions were no longer present to perform their 'editing' work (Damasio 1994, pp. 44–45).

III. The human mind, emotion, and the frame problem

We can draw on these discussions to answer our two questions: (1) what makes the human mind free from the frame problem, and (2) are all of us in fact free from the frame problem?

Thinking about the frame problem in light of Damasio's theory clearly suggests that the emotions play a prominent role in preventing humans from suffering from the frame problem. Recall that the frame problem occurs when a cognitive agent with a presumably vast knowledge-base is placed in any specific situation. In a given situation, an agent will be faced with several questions that have a staggeringly large number of possible answers: (1) what are the more salient features of this situation, (2) what specific pieces of knowledge in my vast storehouse of knowledge are needed to cope with this situation? Given the large number of

possible answers to these questions, how does the agent settle on one appropriate answer?

Now, we have Damasio's theory: when a human is faced with a rather large and unmanageable list, the emotions play a central role in making the list smaller and more manageable. Negative emotional qualia may eliminate some options from the list, for example, while positive emotional qualia may bookmark other options as desirable and hence as deserving of attention.

The implication is obvious: humans, constantly faced with a large number of possible options, can quickly settle on a handful of options because of the editing work performed by the emotions, and as a result, humans are by and large free from the frame problem. Emotional qualia play a key role in determining relevance.

To clarify with an example: imagine a human placed in a room with a ticking time bomb. There are many other objects in the room, such as tables, chairs, a painting on the wall, ceiling tiles, floor tiles, paint on the walls and so on. So, what should the human pay attention to? Quickly, the more banal objects in the room are forgotten as the agent, overcome with a fear quale, focuses on the bomb. The initially large enumeration of possible objects to pay attention to is now rather short. Now, what should the agent's goals be? The painting is crooked, so perhaps the agent should attempt to straighten it? But, such commonplace possible goals don't even arise as options for the agent, who is already focused upon the bomb. The agent quickly decides upon a goal: flee the room. Now, in the agent's database of knowledge there are perhaps thousands of pieces of knowledge that are potentially relevant to the room: the agent, for example, knows that the painting on the wall is a Picasso knockoff. But, already focusing on the bomb, and having already decided upon the need to flee, the agent swiftly accesses one specific piece of information from this database: doors are how one gets out of a room. The agent leaves the room, the whole episode perhaps taking less than 15 seconds. The human has performed significantly better than Dennett's imaginary robots in their similar hypothetical situation, largely because of the terror the human felt upon seeing the bomb.

Our second question is 'are *all* of us in fact free from the frame problem, or do some humans suffer from the frame problem?' Our answer is that Damasio's case studies, such as the one involving Elliot discussed above, suggest that some of us do in fact suffer

from the frame problem. The presence of certain mechanisms associated with the emotions are what prevent us from suffering from the frame problem, so if an agent's mechanisms are defective or damaged, it stands to reason that the agent would suffer from the frame problem. Looking at the case of Elliot, this is precisely what we see. Elliot's emotional centers were damaged from a brain tumor, which presumably prevented the emotions from effectively performing their role in cognition; as a result, Elliot displayed several aspects of the frame problem, including his inability to focus on the relevant aspects of the task at hand, which in turn led to his inability to achieve goals.[7]

IV. Implications

We conclude by discussing two implications of our argument.

First, recall that the frame problem is one of the most serious difficulties facing the Computational Theory of Mind. When this notion is combined with the insight that in the absence of certain mechanisms, we suffer from the frame problem, one seems to obtain further evidence for the Computational Theory of Mind. If not for the emotions, we suffer from the biggest obstacle facing the Computational Theory of Mind, which seems to suggest that the Computational Theory of Mind is on the right track, but is simply grossly incomplete. Thus, paradoxically, Damasio's research into 'Descartes' Error' shows how workers in artificial intelligence, Descartes' heirs, have succeeded in correctly modeling key components of mentality. Again, if the Computational Theory were largely right, then one would expect that it would be possible for people to suffer from the frame problem. That people do is thus very good inductive evidence for the Computational Theory.

We should not allow extreme cases like Elliot to lead us to overidealize ourselves. How many of us experience a creeping paralysis of reason when we are hungry and in a very good restaurant? Everything looks good and we can't decide. Like Elliot, we

[7] One can plausibly speculate that autism might be another example where humans suffer from the frame problem. It is well known that autism effects the emotions; for example, often, autistic people don't display the normal affection a child has for his or her parents. Further, autistic people often display behaviors reminiscent of the frame problem, such as the focusing of attention on irrelevant aspects of the environment.

may fixate on minutiae. When the waiter comes we often order something randomly because we just can't decide. Here perhaps hunger interferes with the manner in which our affective states help us cut down the phase space of possible meals. More plausible examples might involve just having the blues. When depressed we may score well on written tests measuring practical rationality yet still be unable to make rational decisions about our lives.

If we are correct, then human beings manifest the irrationality of artificial intelligence systems when the affective states associated with emotions are off. For then, following Damasio, the somatic markers will not function properly and the frame problem will arise.

This order of explanation is a radical inversion of that governing standard cognitive science. While standard cognitive science seeks to computationally model human cognitive abilities, we seek psychological models of computational defects. Standard cognitive science has followed Chomsky's injunction to derive 'competence' rather than 'performance' models. This involves seeking to get computers to match or surpass human cognitive successes. Such an approach has, for example in computational linguistics, shed impressive light on our cognitive abilities. We hope that our discussion of the frame problem shows that computational failures shed light on human disabilities. It is not too utopian to hope that some day cognitive science will in this manner be able to shed bright light on the whole panoply of psychological disorders, perhaps fundamentally altering the way we individuate, diagnose, and treat varieties of mental illness.

On a less utopian note, our claims seem to call into question the now commonplace distinction between the 'easy' and 'hard' problems of consciousness (see Chalmers 1995). The easy problems supposedly concern cognitive abilities. For example, what procedures or algorithms are involved in visual information processing, understanding of language, making and executing plans? One must add to this list of easy questions, 'how do we determine relevance?' But, Damasio's work makes clear that this 'easy' problem cannot be resolved without appeal to the phenomenology of emotion, as emotional qualia are what allow properly functioning adults to focus on the relevant possibilities when solving problems. Questions concerning phenomenal consciousness, however, are supposedly the 'hard questions;' that is, in order to solve an easy question, one must appeal to phenomenology, or

hard issues. In short, in this case at least, the distinction between hard and easy questions is hopelessly blurred; Chalmers' distinction is in peril.

Please note that we are not in any way claiming to have taken a step towards actually solving the hard problem. Rather, we are simply calling into question Chalmers' much discussed and influential distinction between 'easy' and 'hard' problems.

The force of this result is directly proportional to how hard the hard problem really is. If one thinks that the hard problem is easily solvable or dissolvable, then one will not be troubled by the fact that the easy problem requires a solution to the hard problem.[8] On the other hand, if one thinks that there are no completely satisfying solutions to the hard problem (see Seager 1999; McGinn 1989), then our result is clearly of greater magnitude.

Department of Philosophy and Linguistics Program
106 Coates Hall
Louisiana State University
Baton Rouge, LA 70803

jasonmegill@hotmail.com
jlm3am@Virginia.edu
jcogbul@lsu.edu
joncogburn@yahoo.com

References

Chalmers, D. (1995). 'Facing Up to the Problem of Consciousness', *The Journal of Consciousness Studies* 2 (3), pp. 200–219.

Damasio, A. R. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain.* (New York: Grosset/Putnam).

—— (1999, 2000). *The Feeling of What Happens: Body and Emotion in the Making of Consciousness.* (New York: Harcourt Brace).

[8] If one believes that the *affective* component of emotion (or emotional qualia) are epiphenomena (or play no causal role in shaping the character of our cognition (so, for example, the emotions could be completely unconscious and still play whatever role they do in shaping cognition)), then one will fall into the category of those who think the hard problem is easily dissolvable. This stance, however, does not undermine our claim that a solution to the easy problem requires a solution to the hard problem. That is, on this tack, one is still offering a solution to the hard problem in order to solve the easy problem, a solution that simply amounts to the claim that the hard problem can be easily dissolved. Further, it does not undermine our claim that the computational theory of mind gets something right about cognition.

Dennett, D. (1984). 'Cognitive Wheels: The Frame Problem of Artificial Intelligence,' in Hookway, ed., *Mind, Machines and Evolution.* (Cambridge: Thyssen Grove Volume). (Also in Pylyshyn 1987).
Fodor, J. A. (2000). *The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology.* (Cambridge, MASS: MIT Press).
Ford, K. and Pylyshyn, Z. (eds.). (1996). *The Robot's Dilemma Revisited.* (Norwood, NJ: Ablex).
Gardner, H. (1985). *The Mind's New Science: A History of the Cognitive Revolution.* (New York: Basic Books).
Johnson-Laird, P. N. (1983). *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness.* (Cambridge MASS: Harvard University Press).
Kim, J. (1996). *Philosophy of Mind.* (Boulder, CO: Harper Collins).
LeDoux, J. E. (1996). *The Emotional Brain.* (New York: Simon & Schuster).
McGinn, C. (1989). 'Can we solve the mind-body problem?,' *Mind* 98, pp. 349–366.
Panksepp, J. (1998). *Affective Neuroscience: The foundations of human and animal emotions.* (New York: Oxford University Press).
Pylyshyn, Z. (ed.). (1987). *The Robot's Dilemma.* (Norwood, NJ: Ablex).
Seager, W. (1999). *Theories of Consciousness: An Introduction and Assessment.* (London: Routledge).
Shanahan, M. P. (1997). *Solving the Frame Problem.* (Cambridge, MASS: MIT Press).
Sutherland, S. (1992). *Irrationality: The Enemy Within.* (London: Constable).