

Stefan Linnquist  
Biohumanities Project  
University of Queensland  
Forthcoming in Notre Dame Review of Books

If it Feels Good, Believe It. A Review of Paul Thagard's *Hot Thought: Mechanisms and Applications of Emotional Cognition*.

An emerging body of psychological evidence suggests that practical, moral, and even theoretical reasoning is not as cold and unemotional as many cognitive scientists have assumed. In *Descartes's Error* Antonio Damasio describes a brain injury patient, Elliot, who agonizes over the simplest of long term decisions. The problem with Elliot, we are told, is that damage to his prefrontal cortex prevents signals from his limbic system to communicate with his executive decision-making machinery. This leads Elliot and others like him to make impulsive and often self-destructive choices, though in an oddly unemotional manner. Such examples motivate two hypotheses that are central to Paul Thagard's recent book, *Hot Thought*. First, in normally functioning people reason is always influenced by emotion. Second, normative models of rationality should take this (purported) psychological fact into account. Much of *Hot Thought* articulates a formal model of emotionally-laden cognition called HOTCO (short for "hot coherence") which, Thagard argues, is both mechanistically accurate and heuristically valuable as a decision making tool. In what follows I first provide an outline of this skillfully written book. I then offer some challenges to Thagard's two core theses. In particular, I question HOTCO's accuracy either a mechanistic or a computational model of emotional cognition. One problem is that the model seems unable to accommodate the phenomenon of emotional encapsulation. Furthermore, I argue that HOTCO's normative value begs some fundamental questions about which emotions (if any) are conducive to rational thought.

*Hot Thought* is a collection of papers published by Thagard and several co-authors since 2000. Despite a certain amount of repetition in the description of HOTCO, the first half of this book (chapters 1-7) follows a natural progression reflective of a flourishing research program. The opening chapter endorses the model of mechanistic explanation recently developed by Machamer, Darden, and Craver (2000). On this view, a mechanistic explanation of some psychological process identifies the "entities" and

“activities” that reliably generate that process. This position sets the stage for the following four chapters, where Thagard and colleagues present HOTCO as a mechanistic explanation of analogical reasoning, intuitive judgment, and group decision making. Thagard maintains that reasoning in these domains involves a combination of emotion and cognition. For example, your intuitive judgment about whether Sally is trustworthy might be influenced by several “cognitive” considerations, such as Sally’s reputation and your previous interactions with her, as well as several “emotional” factors, such as your gut feeling about Sally or the non-verbal cues she displays. A HOTCO model of this decision would take the form of an artificial neural network. Nodes in the network stand for representational elements like concepts, images or beliefs. Each node is assigned an activation value, which stands for the cognitive significance of the corresponding representational element. For example, Sally’s reputation might receive a high activation value, while one’s gut reaction to Sally might receive low initial activation. However, activation values are then multiplied by a positive or negative valences (between +1 and -1), which stand for the emotional significance attributed to each corresponding element. Reputation might have a low valence rating, for example, while non-verbal cues might be highly valenced. Activation values and valences are then multiplied by the connection weights among nodes. The final activation pattern is interpreted either as mechanistic description of one’s emotionally-influenced judgment about Sally’s trustworthiness, or, if the model is interpreted normatively, as a guide for how such decisions ought to be made.

The following two chapters (6 and 7) depart from HOTCO and offer a more neurologically motivated model in its place. I found these the most thoughtful and interesting sections of the book. Chapter 6 (co-authored with Brandon Wagar) presents a model called GAGE, named after the notorious railroad engineer turned hooligan. Nodes in GAGE represent anatomically defined brain regions like the amygdala, the prefrontal cortex and the hippocampus. Connections among nodes represent known afferent and efferent connections among these brain regions. Strikingly, once GAGE is up and running its behavior displays some of the same functional characteristics, like synchronous firing among clusters of “neurons”, found in emotional neural circuitry. Even more impressive, “lesioning” certain connections in GAGE causes the model to alter its decision making behaviour in roughly the same fashion as brain injury patients

who suffer from same sort of lesion. These chapters illustrate how models like GAGE, which are based on hard anatomical data rather than folk psychological descriptions, are valuable tools for generating and testing hypotheses about the mind/brain relationship.

The second half of *Hot Thought* (chapters 8 – 16) outlines several practical applications of HOTCO. These chapters are more fractured than in the first half of the book, and the arguments are less convincing. There is a chapter on “Why wasn’t OJ convicted?” in which a modified version of HOTCO simulates irrational decision making in the courtroom. Here, the model appears to have been tweaked to generate the desired verdict of not guilty by simply amping-up the influence of valence on judgment. This is followed by an interesting chapter on the nature of reasonable doubt, which will appeal to specialists working on this topic. The following few chapters present a rather disappointing discussion of the emotional nature of scientific reasoning. While I agree with Thagard that scientific reasoning is often influenced by emotion, his methodology for exploring this idea is extremely flat footed. For example, to determine which emotions were instrumental in Watson and Crick’s discovery of the structure of DNA, Thagard conducts a survey of James Watson’s autobiography *The Double Helix*, counting the number of times emotion labels are used. Thagard reports that Watson uses “happiness” and “interest” more frequently than “sadness” or “beauty” – but what does this reveal about the role of emotions in science? To begin with, the discovery of the double helix is an exceptional moment in the history of biology. Even if Thagard could determine which emotions guided Watson and Crick’s discovery, it is doubtful that this result would generalize. More problematically, Thagard relies exclusively on Watson’s first person reports to determine which emotions were experienced and when. Such reports are notoriously inaccurate to begin with, let alone when they are being recounted a full decade after they were allegedly experienced. Finally, Watson’s version of the events surrounding the discovery of the double helix has been roundly criticized by Francis Crick, Linus Pauling and other eminent biologists, earning Watson the dubious title of “Honest Jim” (Judson, 2001). An autobiography by Donald Rumsfeld wouldn’t be a very reliable source of information about the emotions involved in nation building, either.

The final chapters of *Hot Thought* include an interesting discussion of self deception, where Thagard operationalizes this concept in terms of the HOTCO model.

This is followed by a chapter on the role of emotion in religious belief, a very interesting topic that Thagard barely touches upon. Much of the religion chapter is spent criticizing (unconvincingly) recent evolutionary accounts of religion. This gives the misleading impression that evolutionary accounts of religious belief systems are in conflict with proximate explanations of the mechanisms involved in religious belief formation. In fact, the two approaches seem perfectly compatible with one another.

There are surely many other practical applications of HOTCO than the examples that Thagard explores in the latter half of his book, and the explanatory purchase of HOTCO does not depend their plausibility. It is therefore worth considering whether HOTCO provides an accurate mechanistic model of emotionally influenced cognition. As Thagard notes, HOTCO has no capacity to model the influences of discrete emotions like anger, fear or jealousy on cognition. Nor does HOTCO represent anatomically defined brain regions. Thus, one might argue that HOTCO is not a *mechanistic* model of emotional cognition, because the relevant “entities” and “activities” featuring in the head do not appear in the model.

Thagard defends HOTCO as a mechanistic model by arguing that it is compatible with GAGE, which is more anatomically explicit: “each HOTCO unit that stands for a high-level representation can be viewed as corresponding to groups of connected neurons in GAGE”, Thagard maintains, and “HOTCO uses activation and valences of units to integrate cognition and emotion, and GAGE uses firing behaviour of groups of spiking neurons in different brain regions to accomplish the same task” (69). Thagard’s argument is that these two models are both mechanistic accounts of the same brain circuits, only at different levels of abstraction. By analogy, one could imagine two mechanistic models of a pulley system, one which depicts every last thread and wheel, another which represents pulleys more schematically with lines and circles.

However, this reply seems a little off the mark. In HOTCO, the relevant “entities” and “activities” are representational states and the semantic relationships among them (like explanatory relevance or analogical similarity). In GAGE, the “entities” are clusters of neurons and the “activities” are spiking frequencies. Whereas HOTCO models are constructed from the “top down” by reflecting on conceptual relationships, GAGE models are constructed from the “bottom up” by studying neuroanatomy. Why expect

that the folk psychological entities and activities will map cleanly onto the neurological ones? It would be quite surprising if the representational elements that are more closely related semantically end up being more tightly clustered together in the head. Yet, this appears to be exactly what Thagard's compatibility claim requires.

Perhaps Thagard would be better off abandoning the idea that HOTCO offers a mechanistic model of emotional cognition (at least, not in the strong sense of mechanism that he endorses in Chapter 1) and view the model instead as a purely computational description of this process (*sensu* Marr). But even at this level there is room for skepticism. My primary concern about the computational accuracy of HOTCO is that it fails to account for emotional encapsulation. It is widely recognized that some emotional appraisals are insensitive to "higher" cognitive judgments. If you are snake phobic, for example, seeing one of these creatures slithering nearby triggers a strong fear response that no amount of rationalization can override. HOTCO seems incapable of modeling this phenomenon. Suppose we construct a HOTCO-style model where the elements "harmless snake," "preoccupied snake" and "never bit a human" are assigned positive valences while the element "there's a snake now" is assigned negative valence. And assume that these elements are all linked to a "remain perfectly calm" node. Presumably, as activation spreads through this network the "remain perfectly calm" unit will take on a high, positive activation value. This outcome, however, might be a poor reflection of your response to a snake in the wild. The general problem here is that, in the model, valence is transmitted among any two elements that bear an evidential relationship to one another; whereas, in reality, emotional appraisal is often encapsulated from relevant pieces of information.

Let's turn to HOTCO's normative value. Should we use HOTCO, for example, the next time we need to decide whether someone is trustworthy? An obvious problem is that emotions can be just as irrational as they are at times informative. Perhaps you don't trust Sally because she looks vaguely similar to someone you never liked in high school. Or perhaps you are misinterpreting her shyness for arrogance and that sort of thing really puts you off. In HOTCO models, such irrational responses carry significant weight in influencing decisions. This has always been a good reason for valuing models

that minimize the influence of emotion on cognition, however psychologically difficult they might be to emulate.

Thagard is well aware of these dangers, and in chapter 2 he offers a general argument for why any rational decision-making formula should include an appeal to emotional intuitions. All human decisions, Thagard assumes, are invariably influenced by emotion. If the influence of emotion on judgment is unavoidable, the argument continues, then a model that ignores this fact is more likely to lead one astray than one which makes the influence of emotion on cognition explicit. Thagard's hope is that emotions which tend to distort good judgment (he calls them "emotional skewers") can be isolated and avoided, while emotions that promote good judgment can be identified and cultivated.

One problem with this argument is that the relevant psychological evidence is highly controversial. The philosopher Philip Gerrans has recently developed a critique of Damasio's somatic marker hypothesis. The impulsive behaviours displayed by brain injury patients are better explained, Gerrans argues, as the result of an inability to form images of future events rather than as an emotional deficit *per se* (Gerrans, forthcoming). Another problem with the argument is that it equivocates between two different interpretations of the claim that reason is always influenced by emotion. On one reading, the claim simply states that emotional mechanisms are constantly interacting with cognitive mechanisms. This does not imply, of course, that emotions are constantly distorting reason. By analogy, the fact that an automobile engine is constantly encountering friction from the road does not mean that the car will be distorted from its path. Perhaps the reasoning machinery just needs to work a little harder to overcome the influences of emotion. The other, stronger reading states that the influences of emotion systematically distort reason, such that one can never determine the "direction" of reason (to continue with the analogy) without taking emotion into account. None of the psychological evidence that Thagard cites supports this stronger reading, and, frankly, I find it implausible to think that emotional influences cannot be dampened down to the point where their influence on reason is negligible. But suppose I'm wrong. Perhaps models of rationality that do not take emotional influences into account are impossible to emulate for psychological reasons, and therefore we require models that take emotional

influences into account. The question then becomes: how do we distinguish the emotions that promote rational judgment from “emotional skewers”. People are often unaware of what they are feeling at a given moment. In fact, people often confabulate their emotional responses in ways that promote their own ends (for example, in agreeing with one’s boss that so and so’s behaviour was “shocking” or “disgusting” in an effort to curry favor). Suppose that we are simply unable to determine with any reliability which emotions are conducive to certain rational ends and which ones are skewers. What then? Thagard assumes that a psychologically realistic model is preferable to one that ignores the influence of emotion on reason altogether. But what about a model that gets it wrong? What’s worse, following a model that ignores emotional influences on reason and is thereby difficult to emulate, or following one that identifies the wrong emotions as conducive to sound judgment?

Stefan Linnquist

Biohumanities Project,  
University of Queensland.

#### **Works cited**

Gerrans, P. (forthcoming) “Mental time travel and myopia for the future”, *Synthese*.

Judson, H. F. (2001) “Honest Jim: The Sequel”, *Nature* 413: 775-776.

Machamer, P., Darden, L., and Craver, C.F. (2000) Thinking about mechanisms. *Philosophy of Science* 67: 1-25.