Conscience: What is Moral Intuition?

Paul Thagard and Tracy Finn
University of Waterloo
pthagard@uwaterloo.ca
Draft 5, August 24, 2007.

Abstract: This paper uses a neural theory of emotional consciousness to develop a novel account of conscience and moral intuition. Emotions are both cognitive appraisals and somatic perceptions, performed simultaneously by interacting brain areas. Conscience is a neural process that generates emotional intuitions combining bodily reactions with cognitive appraisal concerning a special subset of goals. This account explains how moral intuitions can be both cognitive and emotional, and why both moral agreement and disagreement are common. Our theory of conscience is both descriptive and normative, showing how to evaluate the ethical and epistemological validity of intuitions.

Key words: Cognitive appraisal, conscience, consciousness, emotion, intuition, moral judgment, normative, somatic perception,

Conscience is the internal sense of moral goodness or badness of the conduct of oneself or others. The products of conscience are moral intuitions, which are the feelings that some acts are right and others are wrong. This paper offers a new theory of conscience and moral intuition as a particular kind of emotional consciousness, produced by brain processes that combine cognitive appraisal with perception of bodily states. We will discuss the significance of the new theory of conscience for philosophical issues about the role of intuitions in evaluating moral claims.

To evaluate its plausibility, we compare the emotional consciousness theory of moral intuition with previous philosophical and psychological theories of conscience. From a theological perspective, conscience is the God-given ability to distinguish right from wrong. Non-theological accounts of conscience range from the Platonic theory that moral intuitions can be true a priori to the Humean view that they are just emotional reactions based on previous experience. We argue that our emotional consciousness theory provides a better explanation of conscience than these alternatives.

THE EMOTIONAL CONSCIOUSNESS THEORY

Here is an example of conscience experienced by one of the authors of this paper. When Apple announced the iPhone early in 2007, he knew he wanted one, but was taken aback by the high projected price. Then it occurred to him that perhaps he could put the iPhone on his research grant. Immediately, however, he got the bad feeling that this would be wrong, a misuse of government funds for personal purposes. His conscience said not to do it by generating the moral intuition that charging an iPhone to a research grant would be wrong. This reaction was a conscious experience marked by negative emotion tied to a mental representation of the possibility of getting the iPhone in an inappropriate manner.

Introspection supports the claim that moral intuitions generated by conscience are a kind of emotional consciousness. The intuitions that killing and eating people is wrong and that aiding needy people is right are judgments of which we usually have conscious awareness. And their emotional status is evident, as most people feel bad about cannibalism and good about helping. There are obviously other kinds of emotional consciousness besides moral intuitions, as when someone feels happy that the

sun is shining without any moral judgment about it. But if moral intuitions are indeed conscious and emotional, a theory of emotional consciousness should be a good start to understanding them, provided that an account can be given of how moral intuitions differ from other kinds of emotional experiences.

Thagard and Aubie (forthcoming) have proposed a neural theory of emotional consciousness that we will draw on here, although it will not be possible to revisit the full range of neurological and psychological evidence adduced for it. Philosophers and psychologists have long debated the nature of emotion, with two primary views contending for dominance. Since the Greek Stoics, many theorists have argued for the view that emotions are cognitive appraisals concerning the extent to which something aids or hinders our goals (e.g. Nussbaum, 2001; Scherer, Schorr, and Johnstone, 2001). A positive emotion such as happiness is an evaluation that something is contributing to our goals, as when the sun shining enables us to play outside. In contrast, a negative emotion such as sadness or anger is an evaluation that something is preventing the accomplishment of our goals. Since the nineteenth century, the major alternative to the cognitive appraisal theory of emotions has been the somatic perception theory, according to which emotions are perceptions of bodily states (e.g. James, 1884; Damasio, 1994; Prinz, 2004). On this view, emotions are not judgments, but rather internal representations of bodily states. For example, fear is a perception of the bodily changes such as heart rate, respiration, and cortisol levels that respond to external stimuli.

According to Thagard and Aubie (forthcoming), emotions are *both* cognitive appraisals and somatic perceptions, performed simultaneously by interacting brain areas. The interactions are captured by the EMOCON model of emotional consciousness,

shown in figure 1. When a scene is presented to the senses, for example a sunny day or a gruesome corpse, this information is simultaneously conveyed to brain mechanisms for somatic reactions and their perception as well as to mechanisms for appraisal of the significance of the scene. Somatic perception is accomplished by brain areas such as the amygdala and insula that integrate information about bodily states. Cognitive appraisal is performed by brain areas such as the dorsolateral prefrontal cortex and anterior cingulate that evaluate the relevance of a situation for the goals of the perceiver. As shown in figure 1, these two mechanisms interact using many neural connections between the relevant sets of brain areas such as the prefrontal cortex and the amygdala. The EMOCON model of emotional consciousness hypothesizes that positive and negative experiences are generated by ongoing rapid interactions between cognitive appraisals and somatic perceptions. For example, your reaction of disgust and fear to a mangled body results from your brain's integrated representation of your bodily response and your evaluation of a possible threat to your survival goals. Thagard and Aubie (forthcoming) argue that this theory can explain many aspects of emotional consciousness, such as the varying intensity, change, and positive and negative valence of emotions.

From the perspective of the theory of emotional consciousness, conscience is not a special mental faculty, but rather a neural process that involves both cognitive appraisal and somatic perception. The role of bodily reactions explains why moral intuitions are marked by "gut reactions": the disgust most people feel at torture and the joy we feel on a sunny day are indeed visceral. But moral intuition, like emotional consciousness in general, is not *just* visceral, because our overall interpretation of a scene is colored by

cognitive appraisal as well. For arguments that there is more to emotions than somatic perception, see Rolls (2005, pp. 26-29) and Barrett (2006).

There is ample evidence from brain scanning experiments that the brain areas specified by the EMOCON model are involved in moral judgments (see, for example, Greene and Haidt, 2002; Moll, Zahn et al., 2005; Raine and Yang, 2006). But what makes moral intuitions different from other non-moral kinds of emotional reactions? For somatic perception, there seems to be nothing special about the brain processes involved in moral intuition compared to emotional consciousness in general: the relevant areas such as the insula, amygdala, and dopamine system are the same. For example, Moll, de Oliveira-Souza et al. (2005) report that both moral and non-moral disgust recruit remarkably similar brain areas.

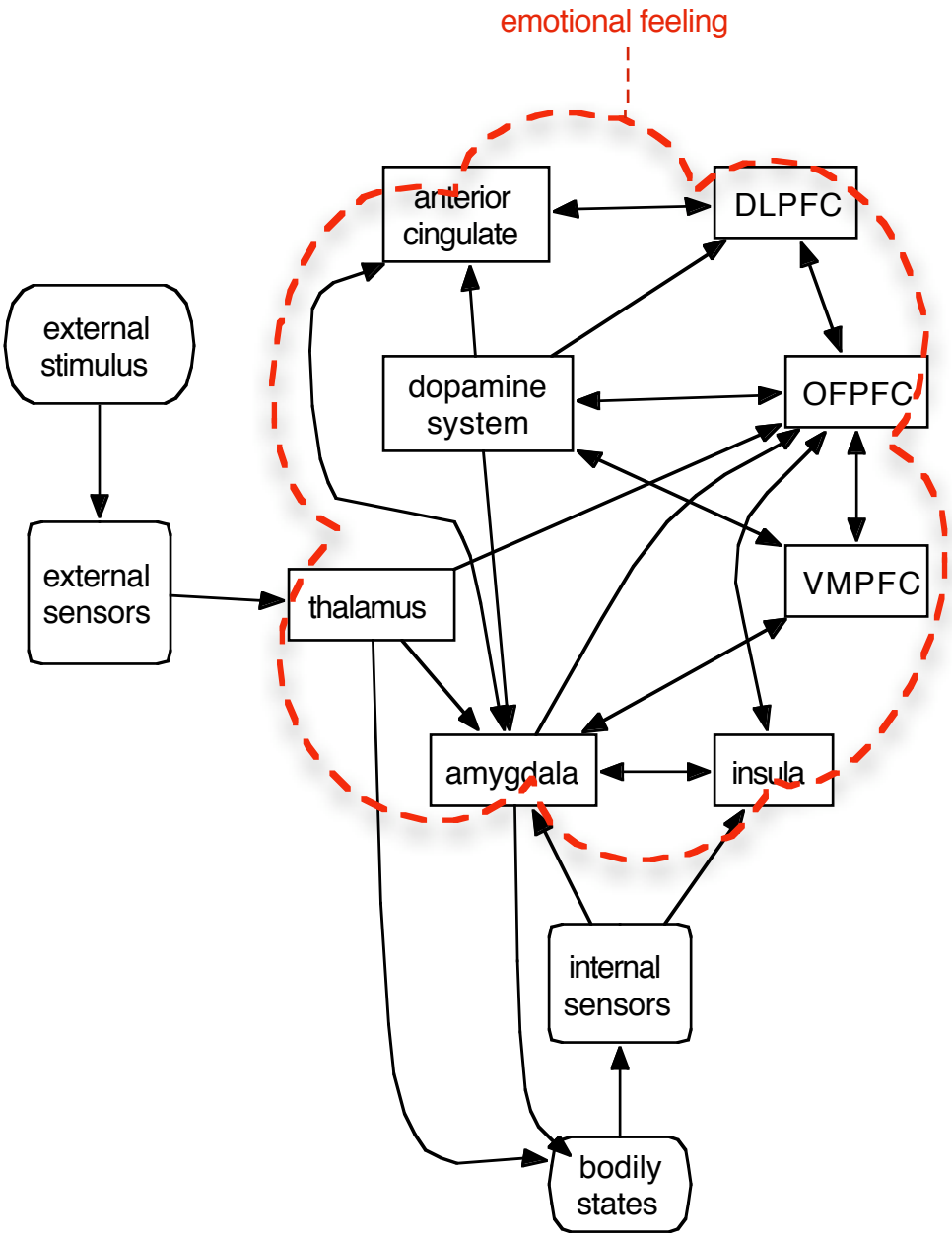

Figure 1. The EMOCON model of emotional consciousness, from Thagard and Aubie (forthcoming). Abbreviations are PFC for prefrontal cortex, DL for dorsolateral, OF for orbitofrontal, and VM for ventromedial. The dotted line represents emotional experience.

Hence the crucial difference must involve cognitive appraisal, with moral intuitions activating a narrower set of goals than general emotional consciousness. From the writings of philosophical and religious writers on ethics, it appears that the relevant moral goals might include following:

- Increase the well-being of people in general.
- Act in accord with abstract moral principles such as fairness and respect for autonomy.
- Satisfy the expectations of social groups such as family.
- Comply with religious standards or other moral code.

Violation of these goals produces negative emotions such as guilt and shame, whereas compliance produces positive emotions such as happiness and relief. Thus when one of us felt bad at the thought of paying for an iPhone with his research grant, it was the result of brain activity that integrated perception of his bodily states with cognitive appraisal that the considered action would violate his goal of using his research funds responsibly.

A highly detailed account of the role of cognitive appraisal in emotions is due to Sander, Grandjean, and Scherer (2005), summarized in table 1. They identify four appraisal objectives concerning the major types of information that an organism needs to adaptively react to a salient event. These objectives (called criteria in table 1) concern relevance for the organism or its social group, implications for goals, coping potential, and significance for social norms and values. For example, the first column of table 1 predicts that an event will tend to generate enjoyment if it is relevant, predictable, very pleasant, intended, highly probable, and one to which the organism can adjust its behavior. However, Sander, Grandjean and Scherer stress that emotional reactions

cannot be reduced to a set of discrete rules, since they involve interactions among the various aspects of cognitive appraisal. Hence there is only a tendency to enjoy an event that has the characteristics just listed, not a rule that says that IF the characteristics apply THEN one has the experience of enjoyment.

Thagard and Aubie (forthcoming) describe a computational model that uses artificial neural networks to provide an overall coherent assessment of an event based on the objectives listed in table 1. It uses 67 units (artificial neurons) to represent the 14 emotional states and many levels of objectives shown in table 1. 262 links between the units represent positive constraints such as how an objective contributes to an emotion and negative constraints such as incompatibilities between emotions. A simple parallel algorithm spreads activation among the units starting with input that describes the characteristics of an event; emerging activation of some of the units and deactivation of others computes a maximally coherent reaction to the event. For example, when the units representing the criteria for enjoyment from table 1 are activated, the unit for enjoyment gets high activation, and the unit for pride is also activated to a lesser extent.

Some of the criteria in table 1 are relevant to moral intuition as well as emotional appraisal in general: pleasantness, goal/need relevance, and standards have all been elements of different ethical theories. We will return to these topics in our discussion below of the normativity of the emotional consciousness theory of conscience. For now, it suffices to recognize that moral concerns are an important subset of the criteria expressed in table 1 and incorporated in the EMOCON account of cognitive appraisal.

Criterion	ENJ/HAP	ELA/JOY	DISP/DISG	CON/SCO	SAD/DEJ	DESPAIR	ANX/WOR
Relevance							
Novelty							
Suddenness	Low	High/med	Open	Open	Low	High	Low
Familiarity	Open	Open	Low	Open	Low	Very low	Open
Predictability	Medium	Low	Low	Open	Open	Low	Open
Intrinsic pleasantness	High	Open	Very Low	Open	Open	Open	Open
Goal/need relevance	Medium	High	Low	Low	High	High	Medium
Implication							
Cause: agent	Open	Open	Open	Other	Open	Other/nat	Other/nat
Cause: motive	Intent	Cha/int	Open	Intent	Cha/neg	Cha/neg	Open
Outcome probability	Very high	Very high	Very high	High	Very high	Very high	Medium
Discrepancy from expectation	Consonant	Open	Open	Open	Open	Dissonant	Open
Conduciveness	Conducive	Vcond	Open	Open	Obstruct	Obstruct	Obstruct
Urgency	Very low	Low	Medium	Low	Low	High	Medium
Coping potential							
Control	Open	Open	Open	High	Very low	Very low	Open
Power	Open	Open	Open	Low	Very low	Very low	Low
Adjustment	High	Medium	Open	High	Medium	Very low	Medium
Normative significance							
Internal Standards	Open	Open	Open	Very low	Open	Open	Open
External Standards	Open	Open	Open	Very low	Open	Open	Open
Criterion	FEAR	IRR/COA	RAG/HOA	BOR/IND	SHAME	GUILT	PRIDE
Relevance							
Novelty							
Suddenness	High	Low	High	Very low	Low	Open	Open
Familiarity	Low	Open	Low	High	Open	Open	Open
Predictability	Low	Medium	Low	Very high	Open	Open	Open
Intrinsic pleasantness	Low	Open	Open	Open	Open	Open	Open
Implication							
Cause: agent	Oth/nat	Open	Other	Open	Self	Self	Self
Cause: motive	Open	Int/neg	Intent	Open	Int/neg	Intent	Intent
Outcome probability	High	Very high	Very high	Very high	Very high	Very high	Very high
Discrepancy from expectation	Dissonant	Open	Dissonant	Consonant	Open	Open	Open
Conduciveness	Obstruct	Obstruct	Obstruct	Open	Open	High	High
Urgency	Very high	Medium	High	Low	High	Medium	Low
Coping potential							
Control	Open	High	High	Medium	Open	Open	Open
Power	Very low	Medium	High	Medium	Open	Open	Open
Adjustment	Low	High	High	High	Medium	Medium	High
Normative significance							
Internal Standards	Open	Open	Open	Open	Very low	Very low	Very high
External Standards	Open	Low	Low	Open	Open	Very low	High

Table 1. Predicted appraisal patterns for some major emotions. ENJ/HAP, enjoyment/happiness; ELA/JOY, elation/joy; DISP/DISG, displeasure/disgust; CON/SCO, contempt/score; SAD/DEJ, sadness/dejection; ANX/WOR, anxiety/worry; IRR/COA, irritation/cold anger; RAG/HOA, rage/hot anger; BOR/IND, boredom/indifference; Other/nat, Other or Natural (e.g.weather); Cha/int, Chance or Intent; Cha/neg, Chance or Negative Intent. From Sander, Grandjean, and Scherer (2005), p. 326.

In sum, conscience is a neural process that generates emotional intuitions combining somatic perception (the gut reaction) with cognitive appraisal concerning a special subset of goals.

EVIDENCE FOR THE THEORY

In order to evaluate our new theory of conscience with respect to older ones, we need to specify what should be expected of such a theory from both descriptive and normative perspectives. The first requirement of a theory of conscience is that it should explain how moral intuitions can be both cognitive and emotional. They seem to be cognitive in that they make claims about things in the world, for example that it was wrong for Jeffrey Dahmer to rape, kill, and eat his victims. Moral intuitions also have a large emotional component, as people react to repulsive actions such as Dahmer's with anger and disgust. That moral judgments are generally emotional is evident not only from introspection, but also from studies in experimental psychology and neuroscience (Greene and Haidt, 2002).

Second, a theory of conscience should be able to explain moral agreement: people often have the same moral intuitions in cases such as the Dahmer one. Thus conscience is not totally idiosyncratic, and seems to have less variation across individuals, cultures, and historical periods than, say, fashion sense. Third, a theory of conscience must explain why there is also a great deal of moral disagreement: people have very different moral intuitions concerning topics such as abortion, homosexuality, and capital punishment.

Fourth, it is philosophically important for a theory of conscience to be normative as well as descriptive. In addition to explaining why people have the moral intuitions that they do, a theory of conscience should help to evaluate the ethical and epistemological validity of such intuitions. Are they indubitable as might be claimed by theories that take moral intuitions to be true a priori? Or have they no objective force at all, as is insisted by theories that take ethical statements to be just expressions of emotional attitudes? Thus a theory of conscience should provide insights into the normative status of moral intuitions as well as explaining their cognitive and emotional content and the existence of moral agreement and disagreement.

We view it as a strong merit of our theory of conscience that it is not only consistent with, but helps to explain, a broad range of psychological and neurological evidence. This accomplishment will not impress dualists and anti-naturalists, who think that moral theory should operate in a logical space entirely separate from empirical matters. But even dualists should recognize that a theory of conscience needs to be able to account for the aspects of moral intuitions listed above, including their cognitive/emotional nature, agreement/disagreement aspects, and normativity.

If moral intuition is a kind of emotional consciousness as specified by the EMOCON model, it is easy to see how ethical reactions can be both cognitive and emotional. They are inherently emotional because they are carried out by the same neural processes that generate emotional reactions, including the perceptions of bodily states that give emotions – and moral intuitions – their visceral character. A purely somatic theory of emotions would have difficulty accounting for the cognitive content of ethical judgments, but this is not a problem for our emotional consciousness theory that

allows a crucial component of appraisal with respect to goals. This appraisal, carried out by the full range of cognitive processes in the prefrontal cortex, can call on all the representational resources of the brain's most intellectually sophisticated part. Through incorporation of the experimentally established neural interconnections of the prefrontal cortex with viscerally-connected areas such as the insula and amygdala, the EMOCON model shows how moral intuitions can simultaneously and inextricably be both cognitive and emotional.

To explain widespread agreement in moral intuitions among humans, for example in the Dahmer case, we can attend to great similarities in people with respect to both somatic perception and cognitive appraisal. Except in rare cases of damage to brain areas such as the amygdala or the insula, we have evidence from dissection and brain scanning experiments that all humans have the same brain structures and connections, including those shown in the EMOCON model in figure 1. This similarity partly explains why emotions such as happiness, sadness, fear, anger, disgust, and surprise seem to be culturally universal (Ekman, 2003). People worldwide have the same basic emotions because they have the same physiology, including bodily processes such as blood circulation, respiration, and stress hormones, and also the same brain areas including the prefrontal cortex and amygdala.

Physiological similarity alone, however, is not sufficient for a full explanation of moral agreement, because emotional consciousness, according to the EMOCON model is also affected by cognitive appraisal with respect to goals. It seems, however, that there is a great deal of similarity in the goals that enter into people's cognitive appraisal of scenes. People generally experience fear if their survival is threatened or of they are in

danger of being hurt. People generally experience happiness when they can anticipate pleasure or other rewards. Most people are concerned not only with happiness and freedom from pain for themselves, but also for other people, at least those with whom they have emotional connections such as family and friends (Batson, 1991). Hence, from the perspective of conscience as a kind of emotional consciousness, we can explain moral agreement as based on a combination of the similarities of people with respect to physiology and goals.

Nevertheless, despite the general agreement in moral intuitions about cases such as Jeffrey Dahmer, there is also a great deal of disagreement. Here are some examples of moral judgments restricted to particular classes of people:

- People should not eat meat and milk together (Orthodox Judaism).
- Priests should be celibate (Catholicism).
- People should not get blood transfusions (Jehovah's Witnesses).
- People should pray five times per day (Islam).
- People should not kill animals (Hinduism).

Many other examples of religious and cultural differences in ethical beliefs show that there is wide variation in moral intuitions (Prinz, forthcoming).

Even within specific cultural groups, there is variation in moral intuitions. The most striking class of people who deviate from moral norms, estimated to be around 1% of the population, are psychopaths. Hare (1993, p. xi)) describes them as follows:

> Psychopaths are social predators who charm, manipulate, and ruthlessly plow their way through life, leaving a broad trail of broken hearts, shattered expectations, and empty wallets. Completely lacking in

> conscience and in feelings for others, they selfishly take what they want and do as they please, violating social norms and expectations without the slightest sense of guilt or regret.

Whereas members of different religious groups have moral intuitions that others lack, psychopaths stand out in lacking moral intuitions that they should be concerned about the suffering of others. A theory of conscience should be able to explain both kinds of variation.

There is mounting evidence for a neurophysiological explanation of psychopathy that is consistent with our emotional consciousness account of conscience. Blair, Mitchell, and Blair (2005) argue that psychopaths suffer from genetic anomalies that disrupt the functioning of the amygdala, a crucial part of the EMOCON model. They propose that atypical functioning of the amygdala in psychopaths produces impairments in emotional learning. Emotional dysfunction can lead to the learning of antisocial methods of achieving goals, because psychopaths do not receive negative reinforcement from the distress or disapproval of others. Moral socialization fails because psychopaths do not experience negative emotions when they see other people suffering, and they do not understand the efforts by caregivers to make them empathize with victims. Thus we can explain the great divergence between the moral intuitions of psychopaths and most members of society by physiological differences in brain structure, particularly the amygdala.

Damage to other brain areas can also lead to moral intuitions different from those among the general population. Koenigs et al. (2007) presented moral dilemmas to people with damage in the ventromedial prefrontal cortex (VMPFC), part of the

EMOCON model that is known to be involved in emotional judgments. The patients were normal with respect to most kinds of moral dilemmas, but reacted differently to high-conflict personal moral dilemmas that pit aggregate moral welfare against direct harm. For example, most people have the intuition that it is right to save five lives at the expense of one, but would not physically push an individual in front of a train to accomplish this end. But patients with VMPFC damage react unemotionally and make the same utilitarian calculation in the personal case, lacking the highly emotional moral intuition that directly causing harm is especially wrong. We know of no studies of the effects of damage to the insula on moral judgments, but it is interesting that patients with insula lesions sometimes find it much easier to quit smoking, presumably because of a change in physical craving (Naqvi et al., 2007). This finding does not concern moral intuitions, but illustrates the role of the insula in the somatic perception component of emotional consciousness.

Most moral disagreements, however, are not caused by physiological differences, as the vast majority of people have the undamaged brain structures shown in the EMOCON model. Hence religious and cultural differences in moral intuition need to be explained by differences in cognitive appraisal. Cognitive appraisal is affected by at least the following factors:

1. Existence and priority of goals.
2. Beliefs about how to accomplish goals.
3. Other beliefs about people and the world.

Religion and other cultural factors can generate radically different goals that people use to assess situations. For example, a religious upbringing can convince someone that the

most important moral goal is compliance with the code of a particular religion as expressed in a text taken to be sacred, such as the Bible, Talmud, or Koran. People are brought up by parents with different religious backgrounds, so they acquire different moral codes that generate different cognitive appraisals and emotional reactions. Someone brought up to think that the Talmud is the best source of proper behavior will have very different reactions to various possible actions than someone who thinks that the source of moral authority is the Koran, New Testament Bible, or other text. Different religions and cultures also vary with respect to what they think constitutes moral goals, emphasizing to different degrees the importance of prayer, attendance at services, and obedience to moral authorities. Moreover, people can have different factual beliefs tied in with religious teachings, for example about the origins of persons. According to Catholics, people originate at conception, whereas for Hindus they are reincarnated. All of these differences can produce different cognitive appraisals of situations, leading to different emotional reactions and moral intuitions.

Thus the emotional-consciousness theory of conscience can explain individual and cultural differences in moral intuitions as arising sometimes from physiological differences, as in the case of psychopaths, but more often from goal-based differences in cognitive appraisal. This understanding of why there is much disagreement in people's moral intuitions, as well as the earlier discussion of why there is also much agreement, provides the basis for a discussion of the normative significance of conscience.

PHILOSOPHICAL IMPLICATIONS: NORMATIVITY

The central normative question about moral intuitions is: Does having a moral intuition that an act is right or wrong provide any degree of justification for the judgment

that the act really is right or wrong? Extreme answers to this question include the affirmative one that moral intuitions can reflect judgments that are true a priori, and the negative one that moral intuitions are just emotional reflections of untutored prejudices and therefore have no evidential force. One of the advantages of the emotional consciousness theory of conscience is that it avoids both of these extremes and can provide some guidance about when moral intuitions deserve to be taken seriously.

If emotions were just somatic perceptions, then emotional consciousness would be a poor contributor to the justification of moral judgments. The fact that you have a physical reaction to something provides no evidence at all that the reaction is morally appropriate. However, the EMOCON model contends that emotions are an integrated combination of somatic perceptions and cognitive appraisals, and the latter can be evaluated with respect to the extent to which they reflect considerations that are relevant to justifying moral judgments. The normative question is transformed into: Does the coherence-based assessment that generates different emotional reactions provide any reason to believe that the existence of a moral intuition can contribute to the justification of a judgment of right and wrong?

To answer this question, we refer back to the model of cognitive appraisal described earlier. Some of the criteria listed in table 1 are clearly very relevant to normative ethical judgments. For consequentialist ethics that judge the morality of an event based on its results, the most relevant criteria include pleasantness, goal relevance, outcome probability, and control. For deontological ethics that judge the morality of an event based on its compliance with ethical principles, the most relevant criteria are agency, motive, and internal and external standards. Hence the contribution of cognitive

appraisal to emotional reactions can include many of the kinds of considerations that leading ethical theories advocate.

Moreover, an emotional appraisal does not have to be limited to only consequentialist or deontological criteria, but can simultaneously take into account both consequences and principles. In medical ethics, for example, people commonly have to balance concerns about the well-being of patients and their families with principles such as autonomy and privacy. According to the emotional consciousness theory of conscience, this balancing is a largely unconscious process that takes place by parallel computing in the brain's neural networks, taking into account both goals and standards.

Potentially, therefore, an emotional intuition can involve a highly sophisticated, multi-factorial appraisal of an event that incorporates elements of both consequence-based and principles-based ethical theories. An emotional intuition that performs such an integration is not just an unreflective bodily reaction, but rather an integration of legitimate contributors to ethical judgments with somatic perceptions. Hence conscience can indeed sometimes be an indicator of the legitimacy of a moral judgment. Feeling bad about putting a personal item on a research grant can indeed be an indicator that this act would be wrong, if the feeling derives in part from an appraisal of the consequences of the act and its compliance with internal and external standards.

The problem, of course, is that it is very difficult for people to identify the actual criteria that contribute to their cognitive appraisals. We have no conscious access to the neural processes that the EMOCON model describes as underlying our emotional reactions, and hence have no way of knowing what cognitive and somatic factors went into a feeling that an act is right or wrong. Although an emotional intuition may in fact

be based in part on appraisals that incorporate legitimate ethical assessments about consequences and principles, we have no way of knowing in any particular case whether the appraisals we make are based on these or other considerations.

Rousseau (1960, pp. 61-62) described the difficulty of knowing the sources of our moral behavior:

> They [reflections] have taught me one great maxim of morality, the only one perhaps which is of practical use: to avoid situations which place our duties in opposition to our interests, and show us where another man's loss spells profits to us. For I am sure that, in such situations, however sincere and virtuous the motives we start with, sooner or later and unconsciously we weaken, and become wicked and unjust in practice, although still remaining good and just in our hearts.

Thagard (forthcoming) argues that understanding the unconscious neural processes underlying moral judgments makes it easy to see why conflicts of interest are such a large problem in government, business, and other enterprises. When your conscience tells you something, you simply have no way of knowing why it is telling you that. You may think that you are making the judgment that is best with respect to all relevant consequences and principles, but it may be that you are unconsciously responding to some aspect of self interest that is particularly salient in the current situation.

Although you cannot absolutely trust your conscience in a particular case, it may be possible to acquire inductive evidence that your moral intuitions have a good chance of being morally legitimate. For decision making, Thagard (2006) recommends a procedure called *informed intuition*. This procedure recognizes that the formal methods

of mathematical decision theory are rarely applicable in real life, because of lack of knowledge about relevant probabilities and utilities. But the procedure also acknowledges that intuitions about what to do may be based on prejudices and misinformation. Informed intuition operates as follows (Thagard, 2006, p. 22):

> 1. Set up the decision problem carefully. This requires identifying the goals to be accomplished by your decision and specifying the broad range of possible actions that might accomplish those goals.
>
> 2. Reflect on the importance of the different goals. Such reflection will be more emotional and intuitive than just putting a numerical weight on them, but should help you to be more aware of what you care about in the current decision situation. Identify goals whose importance may be exaggerated because of emotional distortions.
>
> 3. Examine beliefs about the extent to which various actions would facilitate the different goals. Are these beliefs based on good evidence? If not, revise them.
>
> 4. Make your intuitive judgment about the best action to perform, monitoring your emotional reaction to different options. Run your decision past other people to see if it seems reasonable to them.

Similarly, a moral intuition need not come out of the blue, but can be based on some conscious reflection on the relevant consequences or principles. Hence, if your current moral intuition is based on ethical information, then there is an increased likelihood that your moral intuition yields a defensible moral judgment. Or, equally legitimately, a moral intuition for a particular act may be automatic because its context is analogous to

previous situations in which you have exercised informed intuition. In cases like these, experience and analogy can enable your conscience to be a good guide to action.

So you can sometimes trust your conscience, but what about the consciences of other people? How should you react to the moral intuitions of people who disagree with you? Moral disputes should not degenerate to exclamations that "my intuitions are stronger than your intuitions". The EMOCON model suggests that there are various situations in which it can be legitimate to discount the moral intuitions of others.

1. Physiological defects. If someone has malfunctions in crucial brain areas such as the amygdala, insula, or dorsolateral prefrontal cortex, then their emotional consciousness and hence their moral intuitions are likely to be disrupted. To take an extreme example, we have no reason to respect the moral judgments of psychopaths who have defects in their amygdala-based social learning mechanisms.

2. Lack of moral education. Consider rare cases of people who have been raised without any training by parents or other caregivers in judgments of right and wrong. There is no reason to believe that the conscience of such people would be a reliable guide to morality. For example, people raised by amoral psychopaths would be unlikely to have acquired good moral intuitions, even if their emotional learning mechanisms are not defective.

3. Biased moral education. One of the contributors to cognitive appraisal in emotions is fit with internal and external standards. We have reason to discount the moral intuitions of people that we know have acquired illegitimate standards, for example racism and egoism, the view that one should act only out of self-interest.

4. Situational distortions. As we saw in the case of conflicts of interest, it sometimes happens that our decisions and our moral intuitions are distorted or swamped by salient stimuli that lead to neglect of relevant ethical considerations. For example, someone may be experiencing temptation to have an illicit affair that contradicts the person's own standards, but nevertheless may have the intuition that the affair is the right thing to do. In this case the intuition is governed more by lust than by the interplay of assessment of consequences and principles that marks a full cognitive appraisal. We are justified in rejecting the moral force of the intuitions of others if we can identify causes of the intuitions such as intense temptation. In such cases, people may be acting against their own consciences, as often happens in cases of weakness of will, which results when subcortical motivations such as food and sex dominate conscious cortical reflections about what ought to be desired (Thagard, forthcoming).

We should, however, take seriously the moral intuitions of people who do not suffer from neural defects, absent or biased moral education, or situational distortions. Their intuitions may be as well-based on experience as our own, especially if their backgrounds have ensured that the intuitions are highly informed. In cases of disagreement, people should attempt to move toward consensus by the usual means of verbal communication, such as exchanging facts and arguments. But if we are right that the moral intuitions are inherently emotional, then the kind of communication needed to produce a moral consensus may need to be emotional as well as discursive. Thagard and Kroon (2006, reprinted in Thagard, 2006) propose a theory of emotional consensus that describes how agreement can arise from a combination of verbal and non-verbal communication, including empathy and emotional contagion. It would be interesting to

explore how these mechanisms of emotional consensus might be used to increase agreement in moral intuitions.

In sum, we have shown how our naturalistic, neural account of conscience can be used to address normative questions about the legitimacy of moral intuitions of ourselves and others. Conscience can inform us about what are our moral goals, as well as about good ways to meet these goals. But we must be wary of moral intuitions, of ourselves as well as others, that may derive from defects in physiology, deficiencies in moral education, or situational distortions such as weakness of will.

ALTERNATIVE THEORIES

In philosophy as in science, a theory does not have to be perfect to be worthy of adoption, just better than the alternatives. We will now argue that our emotional consciousness theory of conscience is superior to alternative accounts based on theology, a priori reasoning, and moral grammar. Space does not permit detailed discussion of these alternatives, but we will outline how they compare with the emotional consciousness theory with respect to the requirements discussed earlier: cognition/emotion, agreement/disagreement, and normativity.

Theological Theories of Conscience

According to theological theories, conscience is a God-given ability to distinguish right from wrong. Langston (2006) states that conscience was not explicitly discussed by Plato and Aristotle, but was an important topic for medieval thinkers such as Thomas Aquinas. Many religious people believe that right and wrong are determined by God's prescriptions, and that conscience is a divine gift for telling whether an act is in line with these prescriptions.

The most obvious problem with theological theories of conscience is that they make the strong metaphysical assumptions that that there is a God who provided humans with a faculty of conscience, presumably operating in a non-material soul. Thagard (2000) argues that theism and dualism hang together in a highly coherent explanatory package, but that this package is less coherent with observed facts than a set of scientific alternatives provided by physics, biology, and psychology. If belief in God and the soul are not justified by inference to the best explanation, then theological theories of conscience are dubious. Without good evidence for the existence of God and the soul, there is no reason to believe in them or in any theological theory of conscience.

Even without this major metaphysical problem, theological theories face many difficulties. They provide no information about how conscience actually works, since the operations of the soul are inherently mysterious. Hence they cannot explain why moral intuitions have both cognitive and emotional aspects: introspection alone, without scientific experiments and theories based on them, tells us little about how cognition and emotion operate.

In principle, theological theories of conscience could explain agreement in people's moral intuitions by assuming that God provided us all with a similar faculty for identifying right and wrong. But disagreement is a much greater problem for such theories, as it is difficult to see from the theological perspective why people have ended up with such a wide variety of conflicting intuitions on topics such as abortion and capital punishment. People's disagreements often correlate with disagreements about religious principles, as Catholics, Protestants, Muslims, Hindus, Buddhists, Jews, atheists and others reach highly divergent views about the nature and commandments of God and

about the validity of attendant moral judgments. All this is a big puzzle if conscience is a God-given faculty, but makes perfect sense if conscience is a kind of emotional consciousness shaped by the varied experiences of brains raised in different cultural circumstances.

The issue of normativity could be addressed from a theological perspective by arguing that moral intuitions are a legitimate source of justified moral judgments when they duly convey information from God about his moral prescriptions. Philosophers have long argued, however, that theology cannot be the basis for morality, as it is always legitimate to ask whether the actions and will of God are right or wrong. For example, one can challenge the ethics of the Old Testament God who told Abraham to kill his son Isaac. Given the dubiousness of the existence of God and the lack of evidence of a divine influence on people's moral intuitions, we have little reason to expect theological theories of conscience to tell us much about the normative status of moral intuitions.

Moral Intuitions as A Priori

Ethical intuitionism, the view that people can have immediate awareness of objective moral truths, is a distinguished position in ethical theory (e.g. Audi, 2004, Huemer, 2005; Moore, 1903; Ross, 1930; Stratton-Lake, 2002). We cannot provide a full discussion, but need to highlight how the emotional consciousness theory of moral intuition provides a major challenge to ethical intuitionism.

We note first that ethical intuitionists have been largely silent about how human minds manage to have a priori knowledge of ethical truths. They seem to implicitly assume a dualistic view in which disembodied souls grasp non-natural objects, as in Plato's theory that people can have objective knowledge of moral concepts by grasping

the forms. But there is lack of evidence for souls and for the non-natural objects that they are supposed to grasp. Hence ethical intuitionism does not provide much of an explanation for the cognitive and emotional aspects of moral intuitions. Ethical intuitionists who are not dualists are faced with the problem of explaining how a material brain can arrive at knowledge that is a priori.

Intuitionism can explain moral agreement by hypothesizing that people can somehow grasp the same concepts or truths, but the common process by which they reach similar conclusions remains mysterious. Intuitionism is even less successful in explaining disagreement: if there are objective moral truths of which people can become intuitively aware, why do so few people manage to grasp the same truths? Stratton-Lake (2002) argues that moral propositions can be self-evident even if they are not obvious to every person, because of the difficulty of understanding them. But moral claims, unlike mathematical propositions, are usually not that difficult to understand; and many people who do understand claims proposed by some as self-evident have no difficulty rejecting them (e.g. abortion is wrong).

Stratton-Lake (2002) addresses the problem of disagreement in his discussion of Ross' intuitionism. The charge with which he is concerned comes from Korsgaard (1996), that all an intuitionist can do in the case of disagreement is to assert dogmatically that a proposition is true. Stratton-Lake (2002) argues that Ross does not hold that "moral propositions are brute givens that cannot be questioned (p. 114)." Instead, they become self-evident after we have reflected on them, and have reached what Ross calls "sufficient mental maturity." In other words, once we have given sufficient attention and reflection to a particular moral principle, we come to accept it as self-

evident. Thus, disagreement can occur, and continue to occur until we have sufficiently reflected on the self-evident principles. As noted earlier, an intuitionist need not hold that self-evident principles are easily grasped or obviously true.

Nevertheless, there remains the possibility of unresolved moral disagreement. Ross seems to take it for granted that once one reflects on moral propositions, disagreement will dissolve, but dissolution may not occur. Moreover, the process by which one comes to disagree with another's moral view remains unexplained on this account. Why does one agent think that a proposition is self-evident while another does not? An intuitionist can reply that perhaps one agent is tapping into certain self-evident principles, whereas another taps into different ones. This difference can account for how two agents may disagree about whether a moral proposition or set of principles is self-evident.

However, this account does not give a full explanation of why disagreement continues to occur about the same issue. Consider again the issue of abortion. Often, agents have opposing, not just differing, views on the moral status of abortion. How could two agents discussing the same moral issue come to believe such drastically different viewpoints? Intuitionists may contend that that one agent is unable to grasp the proposition, or to believe the truth of it, but why does this occur? If neither of the agents seems to hold a moral belief based on faulty information or biases, the cause of such disagreement is unknown. This sort of disagreement is difficult to explain if moral propositions are a priori, since there is no obvious reason why two agents cannot grasp the same propositions. Hence intuitionists cannot dodge the difficulty of explaining

disagreement, whereas we showed that the emotional consciousness theory of moral intuition can robustly explain both agreement and disagreement.

If ethical intuitionism were a viable theory, it could contribute substantially to the normativity problem of evaluating the validity of intuitions, marking at least some of them as yielding judgments that have a kind of objective, a priori truth. Unfortunately, intuitionism has no way of identifying which intuitions have the desired objective character, which is a major problem when intuitions so often differ. When one person is adamant that abortion is wrong and another person has an equally strong intuition that abortion is right, it is impossible to say which one has immediate awareness of some moral truth. In contrast, we argued that the emotional consciousness theory of conscience can help illuminate the conditions under which moral intuitions tend to yield objective judgments: when they are based on experience rather than on neural defects, poor education, or situational distortions.

Moral Universal Grammar

A new, more psychological account of the origins of conscience is provided by the theory of moral universal grammar, according to which all humans are born with innate ethical principles (Hauser, 2005; Mikhail, 2007). These rules are like the innate linguistic rules that Noam Chomsky claims is the basis for all human grammar. According to Chomsky, the diversity of human languages is built on a common set of innate grammatical principles that contain parameters that may be set in different ways to yield the grammars of different languages. Similarly, according to the theory of moral universal grammar, the diversity of moral views is built on a common set of innate

principles that contain parameters that allow for cultural diversity. Thus moral universal grammar explains agreement and disagreement in ethical views in much the same way that linguistic theory explains commonality and divergence in human languages.

As we remarked earlier, our emotional consciousness theory of conscience is in principle consistent with the view that moral judgments are partly innate, but we find this innateness hypothesis implausible. First, no one has identified anything like a brain module for moral reasoning. Whereas there are human brain areas that seem dedicated to linguistic processing (e.g. Broca's and Wernicke's areas), there are no areas that are known to play a special role in ethical judgments. The brain areas in the EMOCON model, such as the prefrontal cortex, the anterior cingulate, and the amygdala, are all involved in many kinds of human thinking. There is no apparent module for emotional consciousness, which arises from the interaction of multiple brain areas, all of which have diverse functions. Similarly, there is no apparent module for moral intuition, which also requires involvements of many brain areas known to be required for a wide range of cognitive and emotional processing. To paraphrase Bertrand Russell, postulating modules has all the advantages of theft over honest toil, and should not be done unless there is substantial neurological and psychological evidence for them.

Second, it is not at all obvious that there are innate moral principles whose parametric variations can explain moral disagreements. Haidt (2007; Haidt and Bjorklund, forthcoming) contends that cross-cultural studies suggest that humans have five major moral concerns: suffering, reciprocity, hierarchy, purity, and group boundaries. But he and the proponents of moral universal grammar have not shown how such concerns constitute the principles and parameters that are supposed to underlie

ethical thinking. Moreover, they have not accounted for the kind of individual and group moral diversity that we explained using the emotional consciousness theory of conscience. For example, some individuals (psychopaths) show little concern for human suffering, as do some cultures (Prinz, forthcoming). For example, slavery and the suffering it causes were widely prevalent until the nineteenth century.

Hence we are skeptical of the existence of a moral universal grammar that would provide an alternative theory of conscience to our emotional consciousness account, according to which moral intuitions largely derive from learning processes that are both cognitive and emotional. The analogy between moral intuitions and linguistic intuitions is too weak to support the hypothesis that conscience is innate.

Our review of alternatives to the emotional consciousness theory of conscience has been unavoidably brief, but should serve to indicate ways in which alternative theories appear weaker. We have not discussed the process of reflective equilibrium, which views objective moral judgments as arising from an ongoing adjustment of intuitions and principles. Our account of moral intuition is broadly compatible with the reflective equilibrium metaphor, but we prefer an approach to ethical coherence that specifies in much greater detail how principles, judgments, actions, goals, evidence, hypotheses, and analogies can combine to maximize coherence (Thagard, 2000, ch. 5). Achieving reflective equilibrium only provides a justification of a system of principles and judgments if the system is more coherent than alternatives with respect to explanatory, deductive, conceptual, analogical, and deliberative constraints.

CONCLUSION

A theory of conscience should be able to account for the cognitive/emotional character of moral intuitions, the existence of both ethical agreement and disagreement, and the normative status of ethical judgments. With respect to these requirements, the emotional consciousness theory appears to be superior to the main alternatives based on theology, a priori knowledge, and moral universal grammar. Of course, much remains to be done to explain in much more detail how the kinds of neural interactions in the EMOCON model generate diverse moral intuitions. We have not yet conducted simulations that show how the interactions between brain areas involved in various cognitive and emotional processes produce particular moral judgments. But we have outlined a descriptively adequate and normatively suggestive theory of conscience that should provide a basis for further naturalistic investigations of moral intuitions.

Acknowledgements: We are grateful to Lorraine Besser-Jones and Patricia Marino for comments on an earlier draft. This research is supported by the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

Audi, R. (2004). *The good in the right: A theory of intuition and intrinsic value*. Princeton: Princeton University Press.

Barrett, L. F. (2006). Are emotions natural kinds? *Perspectives on psychological science, 1*, 28-58.

Batson, C. D. (1991). *The altruism question*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Blair, J., Mitchell, D. R., & Blair, K. (2005). *The psychopath: Emotion and the brain*. Malden, MA: Blackwell.

Damasio, A. R. (1994). *Descartes' error*. New York: G. P. Putnam's Sons.

Ekman, P. (2003). *Emotions revealed: Recognizing faces and feelings to improve communication and emotional life*. New York: Henry Holt.

Greene, J., & Haidt, J. (2002). How (and where) does moral judgment work? *Trends in Cognitive Sciences, 6*, 517-523.

Haidt, J. (2007). The new synthesis in moral psychology. *Science, 316*(5827), 998-1002.

Haidt, J., & Bjorklund , F. (forthcoming). Social intuitionists answer six questions about moral psychology. In W. Sinnott-Armstrong (Ed.), *Moral psychology, vol. 3*. Cambridge, MA: MIT Press.

Hare, R. D. (1993). *Without conscience: The disturbing world of the psychopaths among us*. New York: Pocket Books.

Hauser, M. D. (2006). *Moral minds: How nature designed our universal sense of right and wrong* (1st ed.). New York: Ecco.

Huemer, M. (2005). *Ethical intuitionism*. Houndmills, Basingstoke: Palgrave Macmillan.

James, W. (1884). What is an emotion? *Mind, 9*, 188-205.

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., et al. (2007). Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature, 446*(7138), 908-911.

Langston, D. (2006). *Medieval theories of conscience*. Retrieved May 31, 2007, from http://plato.stanford.edu/archives/sum2006/entries/conscience-medieval/

Mikhail, J. (2007). Universal moral grammar: Theory, evidence and the future. *Trends in Cognitive Sciences, 11*(4), 143-152.

Moll, J., Zahn, R., de Oliveira-Souza, R., Krueger, F., & Grafman, J. (2005). The neural basis of human moral cognition. *Nature Reviews Neuroscience, 6*(10), 799-809.

Moore, G. E. (2003). *Principia ethica*. Cambridge: Cambridge University Press.

Naqvi, N. H., Rudrauf, D., Damasio, H., & Bechara, A. (2007). Damage to the insula disrupts addiction to cigarette smoking. *Science, 315*(5811), 531-534.

Nussbaum, M. (2001). *Upheavals of thought*. Cambridge: Cambridge University Press.

Prinz, J. (2004). *Gut reactions: A perceptual theory of emotion*. Oxford: Oxford University Press.

Prinz, J. (forthcoming). Against moral nativism. In D. Murphy & M. Bishop (Eds.), *Stephen Stich and his critics*. Oxford: Blackwell.

Raine, A., & Yang, Y. (2006). Neural foundations to moral reasoning and antisocial behavior. *Social cognitive and affective neuroscience, 1*, 203-213.

Rolls, E. R. (2005). *Emotion explained*. Oxford: Oxford University Press.

Ross, W. D. (1930). *The right and the good*. Oxford: Clarendon Press.

Rousseau, J. (1960). *The confessions* (J. M. Cohen, Trans.). London: Penguin.
Sander, D., Grandjean, D., & Scherer, K. R. (2005). A systems approach to appraisal mechanisms in emotion. *Neural Networks, 18*, 317-352.
Scherer, K. R., Schorr, A., & Johnstone, T. (2001). *Appraisal processes in emotion*. New York: Oxford University Press.
Stratton-Lake, P. (Ed.). (2002). *Ethical intuitionism: Re-evaluations*. Oxford: Oxford University Press.
Thagard, P. (2000). *Coherence in thought and action*. Cambridge, MA: MIT Press.
Thagard, P. (2006). *Hot thought: Mechanisms and applications of emotional cognition*. Cambridge, MA: MIT Press.
Thagard, P. (forthcoming). The moral psychology of conflicts of interest: Insights from affective neuroscience. *Journal of Applied Philosophy*.
Thagard, P., & Aubie, B. (forthcoming). Emotional consciousness: A neural model of how cognitive appraisal and somatic perception interact to produce qualitative experience. *Consciousness and Cognition*.
Thagard, P., & Kroon, F. W. (2006). Emotional consensus in group decision making. *Mind & Society, 5*, 1-20.