Journal of Applied Philosophy, Vol. 24, No. 4, 2007

The Moral Psychology of Conflicts of Interest: Insights from Affective Neuroscience

PAUL THAGARD

ABSTRACT *This paper is an investigation of the moral psychology of decisions that involve a conflict of interest. It draws on the burgeoning field of affective neuroscience, which is the study of the neurobiology of emotional systems in the brain. I show that a recent neurocomputational model of how the brain integrates cognitive and affective information in decision-making can help to answer some important descriptive and normative questions about the moral psychology of conflicts of interest. These questions include: Why are decisions that involve conflicts of interest so common? Why are people so often unaware that they are acting immorally as the result of conflicts of interest? What is the relation of conflicts of interest to other kinds of irrationality, especially self-deception and weakness of will? What psychological, social, and logical steps can be taken to reduce the occurrence of immoral decisions resulting from conflicts of interest? I discuss five strategies for dealing with conflicts of interest: avoidance, optimal reasoning patterns, disclosure, social oversight, and understanding of neuropsychological processes.*

1. Introduction

Conflicts of interest arise when people make decisions biased by their personal goals, neglecting responsibilities to consider the interests of others. Such conflicts are ubiquitous, as in the following examples:

- A government official who has been lavishly entertained by a lobbyist recommends that a lucrative contract go to the company represented by the lobbyist.
- A business executive conceals information that would have a negative impact on a company's stock price in order to get greater value for exercised stock options.
- A medical researcher funded by a pharmaceutical company designs experiments intended to show that a drug is clinically effective.
- A physician neglects to tell a patient about an expensive medical procedure, because ordering would result in a reduced payment to the physician by the patient's HMO.
- A professor strongly recommends admission of a student to graduate school because of a desire to have a sexual relationship with the student.

In each of these cases, the agent has a responsibility to look after broader interests that are neglected because of the agent's own interests. Conflicts of interest arise in many other professions, including law, journalism, accounting, engineering, counselling, and the arts.[1]

This paper is an investigation of the moral psychology of decisions that involve a conflict of interest. Moral psychology, the study of how minds make ethical judgments, has traditionally been pursued in two different ways by philosophers and psychologists. Philosophers have used reflection on ordinary cases of ethical dilemmas to reach

conclusions about the nature of deliberation, practical judgment, free agency, and moral responsibility. Psychologists have used experiments to investigate the cognitive processes employed in the acquisition and application of moral principles. In contrast, my approach in this paper draws on the burgeoning field of affective neuroscience, which is the study of the neurobiology of emotional systems in the brain. In particular, I show how a recent neurocomputational model of how the brain integrates cognitive and affective information in decision-making can help to answer some important descriptive and normative questions about the moral psychology of conflicts of interest.

The questions to which the findings of affective neuroscience are relevant include the following:

1. Why are decisions that involve conflicts of interest so common?
2. Why are people so often unaware that they are acting immorally as the result of conflicts of interest?
3. What is the relation of conflicts of interest to other kinds of irrationality, especially self-deception and weakness of will?
4. What psychological, social, and logical steps can be taken to reduce the occurrence of immoral decisions resulting from conflicts of interest?

2. Conflict of Interest

A vivid instance of conflict of interest occurred recently in Waterloo, Ontario, Canada, a prosperous city of around 100,000 people. In 1999, the city decided to build a large sports facility that would include four ice rinks for hockey and figure skating, a gymnasium for basketball, a field house for indoor soccer, and twelve outdoor soccer fields. City representatives negotiated what they thought was an excellent interest rate for a lease from MFP, an Ontario company with extensive municipal financing experience. On September 25, 2000, the city council authorized the city to sign a lease that they thought committed the city to repay $112 million (Canadian) over 30 years. Months later it was revealed that the true cost of the lease was $227 million, double what the city council thought they were approving. It turned out that the city treasurer, John Ford, had not read the final version of the lease contract that was presented to council, and that he had been given full responsibility for the deal by the city's Chief Administrative Officer, Thomas Stockie. In 2003, the city of Waterloo conducted an official inquiry that assailed MFP for scamming the city[2] and concluded that Stockie and Ford had failed in their responsibilities to the city of Waterloo. Because the major corporate donor for the sports facility was the company Research in Motion (maker of the Blackberry wireless device), the facility is known as RIM Park, but that company was not implicated in the scandal.

The official inquiry judged that Stockie had failed to follow the City Policy Statement with the respect to the disclosure of conflict of interest situations associated with the acceptance of gifts and entertainment from suppliers. According to the policy:

> A conflict of interest is defined as a conflict between an employee's personal interest and his/her responsibility as an employee of the city of Waterloo that interferes with or has the potential to interfere with the performance of his/her position in serving the interests of the City of Waterloo.[3]

Stockie acknowledged that he had attended numerous social events as the guest of MFP's vice president David Robson, including charity events, a Toronto Maple Leafs hockey game, and a golf tournament in Florida. But Stockie denied that there was anything wrong with these occasions because he had been instructed by the Waterloo City Council to build relationships with potential partners and he did nothing in the interests of MFP. Ford had also attended golf events as the guest of MFP but had disclosed this activity to Stockie. He received the final version of the lease from MFP just before the city council met to approve it, but did not check it over because he trusted Robson.

The Sills inquiry recommended that the city of Waterloo amend its conflict of interest policy statement to specify that a city employee should not accept any gift, favour, or entertainment without written disclosure to a supervisor, and that its maximum value be fixed at $25. Justice Sills clearly thought that Stockie's poor judgment in failing to supervise Ford's conduct of the deal, and Ford's poor judgment in failing to scrutinize the final contract, were partly the result of the friendship they had developed with MFP's Robson based on extensive socializing paid for by MFP. Stockie and Ford were both guilty of conflict of interest, because they neglected their responsibilities to the city of Waterloo as the result of personal interests that arose from their social relationships with David Robson.

No evidence was found that Stockie or Ford had been directly bribed by MFP to agree to the deal that served the interests of MFP over those of the city of Waterloo. A bribe is a kind of blatant conflict of interest that should be obvious to anyone, including the agents who undergo the conflict. In contrast, subtle conflicts of interest such as the ones described here may not be obvious to the involved agents who interpret their situations differently from impartial observers. My concern in this paper is with subtle conflicts of interest that may distort the judgments of decision makers without their awareness of the distortion. Many examples of conflict of interest in government, business, research, medicine, and pedagogy are subtle rather than blatant. I now want to investigate the moral psychology of subtle conflicts of interest.

3. Approaches to Moral Psychology

Moral psychology is the study of how minds make ethical judgments. This study has usually been pursued in three different ways, by philosophical reflection, experimental psychology, and scientifically informed philosophy. After briefly reviewing these approaches to moral psychology, and I will outline a fourth based on affective neuroscience.

Many philosophers interested in ethics and moral reasoning have discussed how minds make ethical judgments, for example Plato, Aristotle, Hobbes, Adam Smith, Hume, and Mill. Before the advent of experimental psychology in the 19th century and the philosophical flight from empirical matters in the 20th, there was no sharp distinction between philosophy and psychology.[4] Philosophical reflection on moral reasoning was unconstrained by the findings of psychological experiments, but could nevertheless draw on important informal observations on how people think about ethical issues. Appropriately, philosophers have been as much concerned with how people *ought* to make ethical judgments as with how they actually do reason.

In contrast, psychologists study how ethical judgments are made using systematic observations and controlled experiments. Influential early investigators included Piaget and Kohlberg, and many subsequent studies of moral reasoning, have been conducted by developmental, clinical, cognitive, and social psychologists.[5] This work is sufficiently diverse that it is impossible to report a consensus about how moral thinking works.

Increasingly, philosophers interested in moral psychology and ethical epistemology have attempted to draw on empirical research.[6] Flanagan argues for a Principle of Minimal Psychological Realism: 'Make sure when constructing a moral theory or projecting a moral ideal that the character, decision processing, and behavior prescribed are possible, or are perceived to be possible, for creatures like us.'[7] This principle is a relatively weak constraint on moral theorizing, but shows a basic way in which empirically based moral psychology can be relevant to philosophical theories in normative ethics.

The approach I want to take to moral psychology is heavily influenced by recent research in affective neuroscience, the study of the neurobiology of emotional systems in the brain. Research in this field has been expanding rapidly over the past decade, and recent studies of brains engaged in moral reasoning have revealed a substantial role for affect in moral judgments.[8] For example, brain imaging studies have been used to argue that moral dilemmas vary systematically in the extent to which they engage emotional processing and that these variations influence moral judgment, and there is a mushrooming literature on the neural mechanisms of moral cognition.[9]

Some philosophical theories in ethics, emotivism and expressivism, have noticed the role of emotions in ethical judgments, but have tended to see it as antithetical to the traditional conception of moral reasoning as rational. But contemporary affective neuroscience emphasizes the interconnections of emotional and cognitive processing, not their independence from or conflict with each other. I will now apply a recent neurocomputational model of how the brain integrates cognition and affect to the moral psychology of conflicts of interest.

4. The GAGE Model of Cognitive/Affective Integration

Brandon Wagar and I developed a computational model, GAGE, of how cognitive information and emotional information are integrated in the brain during decision-making.[10] The model was inspired by the unfortunate case of Phineas Gage, a 19th century railway construction foreman whose brain was severely damaged by an explosion. He recovered enough to go back to work, but had changed from being capable and efficient to being fitful and irresponsible. Antonio Damasio compared Gage to current patients who have also suffered damage to the ventromedial prefrontal cortex (VMPFC).[11] In these patients, cognitive, intellectual, and language skills remain intact, but the ability to reason within social contexts is seriously impaired. According to Damasio, Gage and others with damage to the VMPFC have difficulty making effective decisions because they are unable to produce somatic markers, which are emotional reactions that have become associated through experience with predicted long-term outcomes of responses to situations.

The GAGE model presents a detailed neurological mechanism for how the VMPFC interacts with other brain areas to produce effective decision-making. The key

neurological areas are the prefrontal cortex including the ventromedial area, the hippocampus, the amygdala, the nucleus accumbens, and ventral tegmental area. According to Damasio and the GAGE model, effective decision-making requires integration of reasoning with positive and negative emotional reactions based on memories of previous experiences. The hippocampus provides an interpretation of the current context based on stored memories, the nucleus accumbens serves to anticipate pleasurable rewards, and the amygdala processes negative emotional reactions such as fear. The ventral tegmental area is the source of dopamine stimulation of the nucleus accumbens, and the VMPFC provides connections between the more cognitive areas of the prefrontal cortex and more emotional areas such as the amygdala and nucleus accumbens. Figure 1 schematizes the interactions of the different brain areas.

The GAGE model and the large body of research in affective neuroscience that it complements have important consequences for understanding the nature of decisions, including ethical ones. Here are some implications that are most important for understanding conflicts of interest and other phenomena related to deficits in rationality:

1. Human decision-making is not a purely verbal/mathematical process, but requires integration of cognitive and emotional processing.
2. Cognitive-emotional processing requires interconnections among multiple brain areas.
3. The result of decision-making is a feeling about what is the right or wrong thing to do.
4. People have no conscious access to the process of cognitive-affective integration, so that they cannot know why they have particular feelings about what to do.

The view of decision-making presented in the GAGE model cuts across two major divides found in the literatures on emotions and on moral reasoning: cognitive/somatic and cognitive/emotional. The first divide is between cognitive theories of emotions, which view them as judgments that perform an evaluative appraisal of a situation, and somatic theories of emotions, which view them as responses to physiological states. Cognitive theories have been favoured by psychologists such as Keith Oatley and philosophers such as Martha Nussbaum, whereas somatic theories have been favoured

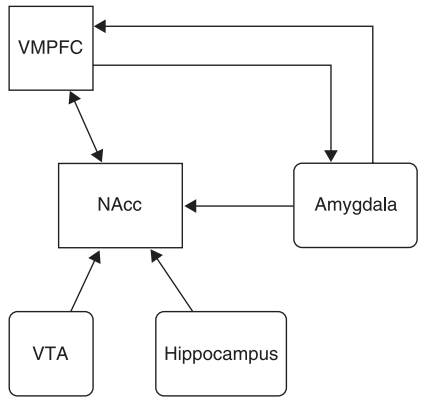

Figure 1. The structure of the GAGE model. VMPFC is the ventromedial prefrontal cortex. NAcc is the nucleus accumbens. VTA is the ventral tegmental area.

by psychologists such as William James and philosophers such as Paul Griffiths.[12] The GAGE model is both cognitive and somatic, as it incorporates cognitive activity in areas such as the prefrontal cortex and hippocampus, and somatically driven activity in the amygdala and nucleus accumbens. The prefrontal cortex performs judgments about current goals and situations, at the same time as the amygdala and other regions summarize bodily inputs. Emotions emerge as the result of both of these kinds of activity, mediated by regions such as the ventromedial prefrontal cortex that provide rich interconnections between cognitive and somatically connected areas. These regions integrate judgments about goals and situations with nonverbal representations of current physiological states. Hence emotions are both cognitive and physiological.

Although GAGE shows how cognitive judgments and somatic reactions can be integrated via the ventromedial prefrontal cortex, it does not itself model either the making of judgments or the occurrence of physiological states. GAGE is by no means the whole story about the brain areas that generate emotional reactions, and research is now underway to develop an expanded neurocomputational model, GAGE II, that will incorporate such areas as the anterior cingulate cortex, the insular cortex, and the dorsolateral prefrontal cortex.

Because GAGE is both cognitive and emotional, it crosses a second divide concerning conflicting views of moral reasoning. Moral thinking is not purely cognitive in the ways that Kantian and utilitarian traditions have supposed, nor is it purely emotional in the ways that emotivist and expressivist views have contended. GAGE posits that *all* thinking, including decision-making and moral reasoning, is inherently both cognitive and emotional.

What empirical evidence is there that the GAGE model gives an accurate account of unconscious decision-making in humans? GAGE does much more than simply explain why Phineas Gage and his contemporary analogs could simultaneously have severely impaired decision-making abilities and relatively unimpaired linguistic and mathematical abilities. GAGE has also been used to simulate the results of two important psychological experiments, concerning the Iowa gambling task and integration of physiological arousal and cognition.[13]

In the Iowa gambling task, participants are given play money and four decks of cards, from which they make a series of selections. Two of the decks yield a large immediate reward and two a small one. However, the two decks that yield a large immediate reward also sometimes produce a large loss, and in the long run the participants are better off if they choose from the two decks that yield only small initial rewards. Normal participants unconsciously learn to choose from the two decks that produce the best overall profit, but patients with VMPFC damage are oblivious to the future consequences of their actions and continue to pick predominantly from the bad decks, because of their lack of covert emotional reactions to good and bad outcomes.[14] Similarly, the full GAGE model is able to simulate the learning of a preference for the good decks, but it prefers the bad decks when the VMPFC part of the model is damaged. This damage prevents stored associations from eliciting a representation of the predicted future outcome of a given response, so that the nucleus accumbens only receives information concerning the amygdala's response to current body states.

In another famous experiment, participants injected with epinephrine were required to fill out questionnaires in the presence of a person who was either pleasant or

unpleasant. The same dose of epinephrine induced the participants to have different experiences, euphoria or anger, corresponding to the pleasantness of the context. Similarly, GAGE's activation of its nucleus accumbens neurons varied depending on stored associations between the VMPFC and the hippocampus, allowing GAGE to elicit representations consistent with the current context. Thus the GAGE model is able to simulate closely some complex psychological phenomena, thereby explaining the emergence of emotional reactions in terms of cognitive-affective interactions involving multiple brain areas. I now discuss its relevance to conflicts of interest.

5. Application to Conflicts of Interest

This section attempts to answer the following questions posed in the introduction:

1. Why are decisions that involve conflicts of interest so common?
2. Why are people so often unaware that they are acting immorally as the result of conflicts of interest?

Consider cases, such as the RIM Park scandal described in section 2, in which government officials make poor decisions that arise from conflicts of interest. The city treasurer's worst decision was to pass the MFP lease contract to city council without examining it carefully. The chief administrative officer's misconduct was failure to supervise the behaviour of the treasurer on a deal of great importance. In both cases, the conflict of interest arose because of a personal relationship that had developed with the MFP vice president through social contacts that included substantial entertainment benefits.

From the perspective of the GAGE model, the bad decisions of the executives of the city of Waterloo arose from unconscious cognitive/affective processes that involved interactions of multiple brain areas. Because of pleasant social and professional contacts, the Waterloo executives had come to associate positive somatic markers with the MFP vice president and the deal he was proposing. These markers contributed to the positive feelings attached to the decisions. For John Ford, the treasurer, the worst decision was to pass the contract that Robson had just given him on to city council without understanding its financial conditions. Ford may have been driven by positive anticipations of pleasurable occasions with Robson, or possibly even by fear of disappointing his buddy. For Tom Stockie, the chief administrative officer, the bad decision was to allow Ford to operate with little supervision. In the absence of any evidence of direct bribery, it seems that both Ford and Stockie genuinely believed that they were doing the right thing for the city, and they seem to have been unaware that they were in a conflict of interest situation. Like all other ordinary decision makers with no knowledge of neuroscience, they were completely ignorant of the origins of their gut feelings about RIM Park issues.

The general implications of the GAGE model for conflicts of interest are startling. Decision makers who have acquired interests at odds with their official responsibilities have no way of knowing whether their decisions emanate from the biases they have acquired from personal interests instead of from good reasoning that takes their responsibilities into account. People are incapable of knowing whether they are acting appropriately or out of a conflict of interest. Hence morally objectionable factors such as personal relationships can intrude into official decisions without much possibility of

detection. People naturally have personal goals that may conflict with their professional responsibilities, but lack a mental mechanism to detect such divergences. Moreover, they usually lack knowledge of the complex neuropsychological processes, described by the GAGE model, which can easily produce decisions generated by prospects of immediate reward rather than by global calculation. Hence people usually remain unaware that they are acting immorally as the result of a conflict of interest.

It follows that one of the main social tools used to combat conflicts of interest, disclosure, is likely to be of limited use. Authors of research papers in major medical journals are now required to report any financial interests in companies whose products they are discussing. Analogously, officials such as the ones in Waterloo might have been required to report the favours that they received from MFP. But there is no reason to believe that the act of disclosing such interests would influence the unconscious processes of decision-making that allow personal biases to distort reasoning away from professional responsibilities. Indeed, the act of disclosure could even have the opposite effect, of giving the deciders and exaggerated confidence that they are acting morally and are therefore impervious to the effects of conflicts of interest. Evidence that this indeed happens is described in section 6 on normative issues.

6. Related Affective Afflictions

I have argued elsewhere that emotions are a part of good decisions, not just bad ones, contrary to the traditional view that pits emotions against rationality.[15] Nevertheless, there is no question that there are cases where emotional influences distort judgments, as when treasurer Ford's personal allegiances overwhelmed his professional responsibilities. I use the term 'affective afflictions' to refer to a set of phenomena that involve emotional distortion of judgments, including bad decisions based on conflicts of interest. The purpose of this section is to discuss other affective afflictions related to conflicts of interest, including self-deception and weakness of will. Understanding of each of them is increased by attention to the GAGE model and recent developments in affective neuroscience.

Consider first weakness of will, which has long been a problem for philosophers who think that humans are inherently rational. Almost everyone has at least one flaw consisting of an inordinate fondness for some physically appealing substance or activity, including food, alcohol, drugs, sex, gambling, and loafing. We all have had occasions when we have found ourselves pursuing one or more of these to the neglect of other goals that we reflectively consider more important, such as health, work, and personal relationships. If our beliefs and desires tell us that we should, for example, exercise rather than eating fattening foods, why do people so often prefer eating over exercising?

The structure of the GAGE model provides an immediate answer. If the decision whether to eat or exercise were made simply by a rational calculation that maximizes expected outcomes, it would primarily be made in the prefrontal cortex, wherein resides the bulk of our linguistic and mathematical abilities. But evidence from functional magnetic resonance imaging suggests that decisions involving immediately available rewards involve parts of the limbic system associated with the midbrain dopamine system.[16] In particular, the nucleus accumbens is well known to be involved in rewards involving food, alcohol, drugs, and sex. Hence the reason that people suffer from

weakness of will is that their decisions are not driven by rational calculation about long term results, but rather by activity in emotionally powerful brain areas such as the nucleus accumbens that respond to immediate stimuli. Without conscious awareness of what is going on in our brains, most people have been in situations where the prefrontal cortex says no, but the nucleus accumbens says yes, yes, yes. Weakness of will is a natural consequence of how our brains are organized to sometimes give priority to motivational rather than cognitive influences.

Sometimes weakness of will can contribute to unethical decisions arising from conflicts of interest. A decision maker might realize that the best thing to do would be to follow professional responsibilities, but simply not be able to resist current or potential pleasurable favours. The Waterloo case, however, does not seem to involve this kind of weakness of will, as Ford and Stockie were not driven by immediate rewards when they neglected the management of the MFP offer. Moreover, they had no sense that they were doing anything wrong, suggesting that their immediate problem was self-deception rather than weakness of will.

Self-deception results from the emotional coherence of beliefs with subjective goals.[17] Self-deception is not a matter of lying to oneself, but rather consists of unexamined acceptance of a belief that could easily be seen to be dubious by an impartial observer or even by the believer operating reflectively. The reason that people frequently succumb to self-deception is that belief acceptance is determined not just by coherence with other beliefs including the relevant evidence, but also by emotional attachments to personal goals such as maintaining self esteem. For example, in Hawthorne's novel, *The Scarlet Letter*, the pastor Dimmesdale manages to convince himself that he is a good minister despite having an adulterous affair with one of his parishioners. He is able to retain the belief that he is good despite overwhelming evidence because it fits emotionally with his goal of achieving redemption. Dimmesdale's hypocritical inference has been simulated using the computational model HOTCO, which shows how emotions can interact with beliefs.[18] In HOTCO, each belief and goal is represented by a single high-level unit in a neural network that is not very neurologically realistic. For example, there is a single unit that represents Dimmesdale's belief 'I deserve redemption' that has an excitatory link to 'I am a good clergyman' and an inhibitory link to 'I have sinned'. These units have both an activation that indicates their degree of acceptance, and a valence that indicated their degree of desirability.

The relationship between HOTCO and GAGE is:

> Each HOTCO unit that stands for a high-level representation can be viewed as corresponding to groups of connected neurons in GAGE, including ones in the prefrontal cortex whose joint spiking activity corresponds to the unit's activation, and ones in the amygdala and nucleus accumbens whose joint spiking activity corresponds to the unit's valence. HOTCO uses activations and valences of units to integrate cognition and emotion, and GAGE uses firing behavior of groups of spiking neurons in different brain areas to accomplish the same task in a more neurologically realistic way.[19]

Thus HOTCO can be viewed as a rough approximation to the more neurologically realistic GAGE model. How can self-deception be understood in terms of GAGE?

Self-deception is puzzling from the perspective of a unitary self, which would seem incapable of both believing something and of potentially realizing that the belief is

unwarranted. From the perspective of the GAGE model, however, the self is far from unitary. Consciousness may appear to be unified, but it gives a very misleading picture of the workings of the mind/brain. Ideally, belief acquisition would be a purely rational process carried out by prefrontal cortex, with acceptance and rejection determined solely by considerations of evidence and overall coherence. But the brain does not have the processing and storage capacity to acquire beliefs of no importance, so we tend to focus belief acquisition on topics that matter to us. Whether a topic matters is intimately tied in with our emotional reactions to various situations, ranging from interest, curiosity, and enthusiasm to anxiety and fear. Hence it is not surprising that there are rich neuronal pathways connecting the prefrontal cortex with intensely emotional brain areas such as the amygdala and nucleus accumbens. There is abundant psychological evidence that people's beliefs are determined in part by their motivations as well as by the available evidence.[20] And there is no central homunculus to check the consistency of one set of beliefs that may be acquired based on evidence and warranted inferences against another set that may have arisen primarily because it fits with one's personal goals. GAGE models decision-making rather than belief acquisition, but it makes it easy to see how self-deception could arise by means of emotion-driven motivated inferences to beliefs that could potentially be recognized by purely neocortical processes as unwarranted.

There are other affective afflictions that affective neuroscience and the GAGE model can illuminate, including rationalization (which is akin to self-deception) and empathy gaps, which occur when people in one emotional state fail to predict their behaviours in other emotional states.[21] Rationalization seems relevant to understanding conflicts of interest because compromised decision makers may construct self-serving explanations of why they acted as they did, insisting that they were just doing their jobs, or even that they thought they were acting in accord with their professional obligations. Perhaps empathy gaps are also relevant, whenever decision makers fail to understand how the emotional states that they were in while being influenced by non-professional interests would produce decisions other than ones that would arise in their more objective emotional states. However, rather than delve into the neuroscience of rationalization and empathy gaps, I turn to a discussion of the normative implications of affective neuroscience for managing conflicts of interest.

7. Normative Issues

This paper has largely been concerned with descriptive questions about why people are so prone to conflicts of interest, but its intent is also normative. Why are conflicts of interest wrong and what can be done to overcome them? The first question is easy to answer from the perspective of familiar ethical theories. A Kantian would say that the people who make a decision under a conflict of interest are acting wrongly because they are in violation of their ethical duties. For example, the officials of the City of Waterloo who were taken in by MFP were clearly in violation of their professional duties to the city and its people. A consequentialist would say that the people who make a decision under a conflict of interest are acting wrongly because their decision has negative consequences for many of the people affected, and positive consequences only for a few who benefit from the decision. For example, tens of thousands of

citizens of Waterloo are having to pay much higher city taxes in order to cover the excessive cost of RIM Park, whereas the only people who benefited financially were a few managers and owners at MFP. Other reasons for moral suspicions about conflict of interest are that it renders judgments less reliable than normal, and that it undermines public trust in institutions.[22] There is thus no doubt that acting under the influence of a conflict of interest is often morally wrong.

So what measures can be taken to eliminate or reduce the occurrence of bad decisions deriving from conflicts of interest? I will consider five strategies: pure reason, disclosure, complete avoidance, social oversight, and neuropsychological information. By pure reason I mean the normative strategy that when people make decisions they should ensure that they are governed only by optimal reasoning patterns such as mathematical decision analysis, Kantian and consequentialist moral evaluation, and appropriate canons of deductive and inductive inference. The problem with this strategy is that it assumes that people can disconnect their reasoning apparatus from the emotional machinery described by the GAGE model. Although people can certainly improve their reasoning in some case by using optimal patterns, it is psychologically unrealistic to expect them to disengage their emotional systems while making important decisions. You cannot simply turn off your amygdala and nucleus accumbens, nor would you want to.

In medical circles, a currently popular way of dealing with conflicts of interest is to have the relevant parties disclose them. For example, authors of articles in major medical journals now need to report their sources of funding. A recent experimental study of disclosing conflicts of interest reveals that it can have perverse effects.[23] A simplified situation in which individuals' professional responsibilities diverged from their personal interests revealed two undesirable effects of disclosure. Disclosure can actually increase the bias in advice because it can lead advisors to feel morally licensed to exaggerate their advice even further. Moreover, when advisors' conflicts of interests are honestly disclosed, people generally do not discount their advice as much as they should. There is clearly no guarantee that merely disclosing a conflict of interest compensates for or counterbalances the emotional pull inherent in the conflict. Indeed, disclosure may have the negative effect of giving decision makers false assurance that they are immune from the effects of the conflict. Given the unconscious nature of decision-making according to the GAGE model, one can never have full confidence that one is acting with such immunity. Moreover, those to whom the conflict is disclosed will likely not realize the extent to which a conflicting interest may continue to influence decisions even after disclosure. Hence disclosure is by no means a panacea to problems of conflict of interest.

Chugh, Banaji, and Bazerman argue that the only effective way to resolve conflicts of interest is for people with conflicts to completely remove themselves from relevant decisions.[24] Certainly the most effective way of dealing with conflicts of interest is to eliminate them entirely. The City of Waterloo has a new policy preventing employees from receiving any gifts, entertainments, or favours worth more than $25. Undoubtedly this is a good policy for public servants, but it is not always feasible in other domains. Much valuable medical research depends on funding by pharmaceutical companies, and there does not appear to be either the means or the political support to have it all taken over by publicly funded institutions, contrary to the suggestion that clinical trials be completely sequestered from industry.[25] Many political campaigns

depend on large amounts of volunteer labour and a network of financial donors, and it is hard to see how candidates for office could be anything but favourably disposed to those who help them. University professors are supposed to treat all their students equally, but personal interactions are such that they will inevitably like some more than others. Hence complete avoidance of conflict of interest, although a useful ideal, will often fail in practice.

Another strategy for dealing with conflicts of interest is social oversight, based on the principles that supervisors or peers of an agent may be able to identify decision errors to which the agent is oblivious. Such oversight is certainly useful, but does not help much if the supervisors or peers have conflicts of interest of their own, which may be as mild as a desire to get along with or be liked by the agent. Moreover, there are situations where decision makers such as medical researchers act largely on their own without much supervision or peer scrutiny. There may be useful interactions between the strategies of disclosure and social oversight, in that disclosure may prompt increased oversight, and awareness of oversight may provoke disclosure.

The final strategy for reducing bad decisions resulting from conflict of interest is simply to make people more aware of the moral neuropsychology of decision-making. Perhaps if people knew more about how cognition and affect are intimately connected and how the connections are inaccessible to conscious introspection, they would be much less confident about the basis and validity of their decisions. They should be able to generate questions such as 'Am I doing this for the people I represent, or for the lobbyist who took me out to a great dinner last week?' They could carefully use the optimal reasoning patterns mentioned above, while at the same time watching for deviations spurred by their conflicts of interest. People also need to be educated concerning the prevalence of motivated inference, so that they can watch for cases where their conclusions derive more from their personal goals than from the available evidence, keeping in mind that even the friendship of a lobbyist can have distorting effects on their judgments. Disclosing their interests to others and subjecting themselves to social oversight may also help to keep people on target to keep their decisions in line with optimal patterns. Thus when complete avoidance of conflicts of interest is not possible, the strategies of optimal reasoning patterns, disclosure, social oversight, and understanding of neuropsychological processes may combine to reduce the prevalence of immoral decisions deriving from conflicts of interest.

My approach to conflict of interest in this paper has been very different from the usual discussions of this topic in applied ethics.[26] Philosophers and professional practitioners have provided insightful discussions of the nature of conflict of interest, what is wrong with it, and strategies for dealing with it. But they have largely ignored questions about the moral psychology of conflicts of interest that are crucial for understanding why people so easily fall into them. I have argued that affective neuroscience as exemplified by the GAGE model of cognitive-affective integration has the resources to address such questions. In particular, the model makes it clear why people often have little conscious awareness of the extent to which their decisions are influenced by conflicts of interest. Moreover, an approach based on neural and behavioural moral psychology can supplement purely philosophical discussions of the efficacy of different strategies by providing psychological reasons why frequently advocated solutions to problems of conflicts of interest may not work very well. In particular, mere disclosure of conflicts of interest may be of little benefit in reducing distorted reasoning.

Folk psychology, which assumes that people's actions derive from their conscious beliefs and actions, makes affective afflictions such as self-deception and weakness of will highly paradoxical: how could people possibly be so irrational? But these afflictions and the distorted decisions that can arise from conflicts of interest, are naturally understood from the perspective of affective neuroscience, which views our decisions and judgments as arising from the unconscious interplay of numerous brain areas, including ones that encode emotions. For both the descriptive task of understanding how conflicts of interest influence thinking and the normative task of devising strategies to counter this influence, philosophical reflection can gain much from attention to the operations of the brain.[27]

Paul Thagard, Philosophy Department, University of Waterloo, Waterloo, ON N2L 3G1, Canada. pthagard@uwaterloo.ca

NOTES

1 M. Davis & A. Stark, A. (eds.), *Conflict of Interest in the Professions* (Oxford: Oxford University Press, 2001). A. Stark, *Conflict of Interest in American Public Life* (Cambridge, MA: Harvard University Press, 2000). D. A. Moore, D. Cain, G. F. Loewenstein & M. Bazerman (eds.), *Conflicts of Interest: Challenges and Solutions in Business, Law, and Public Policy* (Cambridge: Cambridge University Press, 2005).

2 R. C. Sills, *Final Report of the City of Waterloo Judicial Inquiry* (2003) Retrieved December 17, 2004, from http://www.waterlooinquiry.ca/final_report.asp

3 *Ibid.*, p. 163.

4 E. S. Reed, *From Soul to Mind: The Emergence of Psychology from Erasmus Darwin to William James* (New Haven: Yale University Press, 1997).

5 J. Piaget, *The Moral Judgment of the Child* (New York: Free Press, 1948). L. Kohlberg, *The Psychology of Moral Development: The Nature and Validity of Moral Stages* (San Francisco: Harper & Row, 1984). N. Eisenberg, 'Emotion, regulation, and moral development', *Annual Review of Psychology* 51 (2000): 665. J. Haidt, 'The emotional dog and its rational tail: A social intuitionist approach to moral judgment', *Psychological Review* 108 (2001): 814. M. L. Hoffman, *Empathy and Moral Development: Implications for Caring and Justice* (Cambridge: Cambridge University Press, 2000).

6 R. Campbell & B. Hunter (eds.), *Moral Epistemology Naturalized* (Calgary: University of Calgary Press, 2000). J. M. Doris & S. P. Stich, 'As a matter of fact: Empirical perspectives on ethics' in F. Jackson & M. Smith (eds.) *Oxford Handbook of Contemporary Analytic Philosophy* (Oxford: Oxford University Press, forthcoming). O. Flanagan, *Varieties of Moral Personality* (Cambridge, MA: Harvard University Press, 1991). M. L. Johnson, *Moral Imagination: Implications of Cognitive Science for Ethics* (Chicago: University of Chicago Press, 1993). L. May, M. Friedman & A. Clark (eds.), *Mind and Morals: Essays on Ethics and Cognitive Science* (Cambridge, MA: MIT Press, 1996).

7 Flanagan op. cit., p. 32.

8 A. R. Damasio, *Descartes' Error* (New York: G. P. Putnam's Sons, 1994). R. J. Davidson, K. R. Scherer & H. H. Goldsmith (eds.), *Handbook of Affective Sciences* (New York: Oxford University Press, 2003). J. LeDoux, *The Emotional Brain* (New York: Simon and Schuster, 1996). J. Panksepp, *Affective Neuroscience: The Foundations of Human and Animal Emotions* (Oxford: Oxford University Press, 1998). E. T. Rolls, *The Brain and Emotion* (Oxford: Oxford University Press, 1999). J. Greene & J. Haidt, 'How (and where) does moral judgment work?', *Trends in Cognitive Sciences* 6 (2002): 517.

9 J. D. Greene, R. B. Sommerville, L. E. Nystrom, J. M. Darley & J. D. Cohen, 'An fMRI investigation of emotional engagement in moral judgment', *Science* 293 (2001): 2105–2108. W. D. Casebeer & P. S. Churchland, 'The neural mechanisms of moral cognition: A multiple-aspect approach', *Biology and Philosophy* 18 (2003): 169.

10 B. M. Wagar & P. Thagard, 'Spiking Phineas Gage: a neurocomputational theory of cognitive-affective integration in decision making', *Psychological Review* 111 (2004): 67–79.

11 Damasio op. cit.

12 P. E. Griffiths, *What Emotions Really Are: The Problem of Psychological Categories* (Chicago: University of Chicago Press, 1997). W. James, 'What is an emotion?', *Mind* 9 (1884): 188. M. Nussbaum, *Upheavals of Thought* (Cambridge: Cambridge University Press, 2001). K. Oatley, *Best Laid Schemes: The Psychology of Emotions* (Cambridge: Cambridge University Press, 1992).
13 A. Bechara, H. Damasio, D. Tranel & A. R. Damasio, 'Deciding advantageously before knowing the advantageous strategy', *Science* 275 (1997): 1293. S. Schacter & J. Singer, 'Cognitive, social, and physiological determinants of emotional state', *Psychological Review* 69 (1962): 379.
14 Bechara *et al.* op. cit.
15 P. Thagard, 'How to make decisions: coherence, emotion, and practical inference' in E. Millgram (ed.) *Varieties of Practical Inference* (Cambridge, MA: MIT Press, 2001).
16 S. M. McClure, D. I. Laibson, G. F. Loewenstein & J. D. Cohen, 'Separate neural systems value immediate and delayed monetary rewards', *Science* 306 (2004): 503.
17 B. Sahdra & P. Thagard, 'Self-deception and emotional coherence', *Minds and Machines* 15 (2003): 213.
18 *Ibid.*; see also P. Thagard, *Coherence in Thought and Action* (Cambridge, MA: MIT Press, 2000). P. Thagard, 'Why wasn't O. J. convicted? Emotional coherence in legal inference', *Cognition and Emotion* 17 (2003): 361.
19 P. Thagard, & F. W. Kroon, 'Emotional consensus in group decision making', *Mind and Society* 5 (2006): 1–20.
20 Z. Kunda, 'The case for motivated inference', *Psychological Bulletin* 108 (1990): 480–498. Z. Kunda, *Social Cognition* (Cambridge, MA: MIT Press, 1999).
21 G. F. Loewenstein, 'Hot-cold empathy gaps and medical decision making', *Health Psychology* 24 (2005): S49–S56.
22 M. Davis, 'Conflict of interest' in R. Chadwick (ed.) *Encylopedia of Applied Ethics*, Vol. 1 (Orlando, FL: Academic Press, 1998).
23 D. M. Cain, G. F. Loewenstein & D. A. Moore, 'The dirt on coming clean: Perverse effects of disclosing conflicts of interest', *Journal of Legal Studies* 34 (2005): 1.
24 D. Chugh, M. R. Banaji & M. H. Bazerman, 'Bounded ethicality as a psychological barrier to recognizing conflicts of interest' in D. A. Moore, D. Cain, G. F. Loewenstein & M. Bazerman (eds.) *Conflicts of Interest: Challenges and Solutions in Business, Law, Medicine, and Public Policy* (Cambridge: Cambridge University Press, 2005).
25 J. Schafer, 'Biomedical conflicts of interest: A defence of the sequestration thesis — learning from the cases of Nancy Olivieri and David Healy', *Journal of Medical Ethics* 30 (2004): 8.
26 See references in notes 1 and 22.
27 This paper originated in discussions with Ziva Kunda, and is dedicated to her memory. This research was supported by the Natural Sciences and Engineering Research Council of Canada. Thanks to Abninder Litt and an anonymous referee for helpful comments on an earlier draft.