BRAIN
A JOURNAL OF NEUROLOGY

SCIENTIFIC COMMENTARIES

The neural basis of morality: not just where, but when

We have known for some time that an intact ventromedial prefrontal cortex (PFC) is essential for normal social and moral functioning, thanks in large part to studies of individuals who have suffered damage to it (Damasio, 1994). In this issue of *Brain*, Taber-Thomas and colleagues report that the specific nature of a person's moral impairment depends not only upon where brain damage occurs, but also when (Taber-Thomas *et al*., 2014).

Their point of departure is a landmark study showing that individuals with damage to ventromedial PFC make abnormal judgements when faced with a specific class of moral dilemma (Koenigs *et al*., 2007). These dilemmas hinge on a trade-off between allowing several people to die, or saving them by personally killing a single victim. For instance, the trolley problem asks whether it is acceptable to save five people from being hit by a runaway trolley by shoving a man in front of it in order to slow it down (such a bizarrely unfortunate railway mishap may seem the stuff of pure fantasy, but try telling that to Phineas Gage).

Koenigs and colleagues (2007) showed that although neurotypical individuals and those with brain damage to regions outside the ventromedial PFC tended to disfavour such direct harm despite its utilitarian benefits, individuals with adult-onset damage to ventromedial PFC are significantly more likely to judge such harms acceptable. Importantly, however, the differences between groups did not extend to another class of moral 'dilemma'. Specifically, ventromedial PFC participants were just as likely as others to condemn self-interested harmful actions (bumping off a stingy boss, stealing a wallet, selling a daughter into the pornography industry, and so forth).

The ventromedial PFC is widely considered to play a role in bringing affect to bear on decision-making processes (Grabenhorst and Rolls, 2011), and recent research emphasizes a key role of affect in moral judgement (Haidt, 2012; Greene, 2013). Koenigs and colleagues therefore concluded that ventromedial PFC damage likely interfered with the typical aversive affective response to harmful actions, thus biasing moral decisions. By contrast, they argued, participants were able to draw on explicit knowledge of the norm against harming in self-interest to deliver typical judgements of the cases that lacked a utilitarian rationale.

Taber-Thomas and colleagues replicated this experiment with the addition of a key new group: eight individuals who suffered damage to ventromedial PFC during childhood. How would early onset of damage affect their patterns of moral judgement?

Several theories propose that an aversion to harmful action is acquired over the course of development, and Blair (2007) argues that psychopathy arises from a failure to acquire this aversion. According to his model, the aversion to harming others is ordinarily acquired through the experience of harmful actions (e.g. pushing or hitting) being paired with the negative affect elicited by victim distress (e.g. via empathy). More recently, a pair of convergent proposals further developed the idea that the negative affect associated with harmful action is often a product of learning, showing how a widely-used family of reinforcement learning models can specifically account for patterns of non-utilitarian moral judgement in moral dilemmas (Crockett, 2013; Cushman, 2013).

Taber-Thomas and colleagues do not find, however, that early-onset ventromedial PFC damage leads to a general increased willingness to harm. The early-onset group showed precisely the same pattern as the adult-onset group in moral dilemmas, such as the trolley problem, that involve a utilitarian rationale for harm. To be sure, both groups were more likely to endorse harm for utilitarian reasons than the control groups—but the timing of the trauma had no apparent influence (neither did judgements appear to be at ceiling—there seems to have been sufficient room for the early onset group to have shown an even greater indifference to harm). Of course, this evidence does not contradict the hypothesis that the aversion to harm is at least partially learned. Rather, it suggests that the ventromedial PFC is not the seat of that learning. In sum, ventromedial PFC appears to play a more pivotal role in the application of an aversion to harm to moral decision-making processes than in the initial acquisition of that aversion.

Timing did matter, however, for judgements of self-interested harmful actions (killing your stingy boss, for instance). Consistent

with Koenigs and colleagues' (2007) report, Taber-Thomas and colleagues show that the adult-onset ventromedial PFC group condemned such behaviours just as strongly as the control groups. Yet the early-onset ventromedial PFC group endorsed these harms at a significantly higher rate—and the rate was highest among those with the very earliest onset times.

These data force us to re-evaluate the moral competency of individuals with adult-onset ventromedial PFC damage. Recall Koenigs and colleagues' explanation for why these individuals are able to condemn self-interested harm: because they have intact knowledge of the explicit norms against murder, theft and prostitution or, more generally, against harming others for one's own benefit. This account does not square easily with the finding that early-onset ventromedial PFC damage leads to abnormal judgement of self-interested harm. Although it is possible that an intact ventromedial PFC is necessary to acquire explicit knowledge of moral norms, neither current models of ventromedial PFC function nor our understanding of the neural correlates of declarative memory provide much basis for such a conclusion. Rather, it seems likely that the early-onset group is perfectly aware of these explicit moral norms. It would be interesting to test this hypothesis directly.

What, then, explains the difference between the early- and adult-onset groups? Given the presumed role of ventromedial PFC in processing affective contributions to decision-making, a more parsimonious explanation is that the adult-onset group successfully condemns self-interested behaviours due to an affective response. In order for this explanation to work, the ability of an affective response to prohibit self-interested behaviours must require intact ventromedial PFC function early in development (explaining why it is impaired in the early-onset group), but not require intact ventromedial PFC function at the time of moral judgement (explaining why it is preserved in the adult-onset group). In other words, ventromedial PFC would play a role in getting young people to feel that it is fundamentally wrong to harm others just to benefit oneself, but this feeling would subsequently influence moral judgements in a manner largely independent of ventromedial PFC.

This raises a key question, and highlights the value of Taber-Thomas and colleagues' findings as a prod for further research. This question orbits around a peculiar juxtaposition of inferences. On the one hand, it appears that the aversion to harm (as expressed in utilitarian dilemmas) requires an intact ventromedial PFC in adulthood, but is not modulated by early-onset damage. On the other hand, it appears that the aversion to self-interested behaviour (as expressed in non-utilitarian dilemmas) does not require an intact ventromedial PFC in adulthood, but is modulated by early-onset damage. Why should this be?

One approach to this question focuses on the content of the moral values in question. Naively, we might have assumed that value of non-harm and value of non-selfishness would be birds of a feather, acquired developmentally and applied online in similar ways. Taber-Thomas and colleagues' data require, instead, that we think of these moral values as being of fundamentally distinct kinds. Understanding the nature of this distinction is a key challenge for future research. For instance, might it depend on distinct roles of empathy (for harm) versus social reprobation (for self-interested acts) as the basis for learning moral values?

Another way of approaching the question focuses on the computational role of the ventromedial PFC. Much current research focuses on the role that the ventromedial PFC plays online, as an arbiter of diverse affective commitments during decision-making. Relatively less research focuses on the role of the ventromedial PFC in establishing values during development. A major contribution of Taber-Thomas and colleagues' report is to focus our attention on understanding how the computations performed by ventromedial PFC come to support the acquisition of feelings or values encoded within other neural structures. In the end, then, a solution must address not only the 'where' of moral affect (ventromedial PFC), and the when (early in development), but also: 'with whom'?

Fiery Cushman
Brown University, Department of Cognitive, Linguistic and Psychological Sciences, Providence, RI, USA

Correspondence to: Fiery Cushman
E-mail: fiery_cushman@brown.edu

doi:10.1093/brain/awu049

References

Blair R. The amygdala and ventromedial prefrontal cortex in morality and psychopathy. Trends Cogn Sci 2007; 11: 387–92.
Crockett MJ. Models of morality. Trends Cogn Sci 2013; 17: 363–6.
Cushman FA. Action, outcome, and value: a dual-system framework for morality. Pers Soc Psychol Rev 2013; 17: 273–92.
Damasio A. Descartes' error. Boston, MA: Norton; 1994.
Grabenhorst F, Rolls ET. Value, pleasure and choice in the ventral prefrontal cortex. Trends Cogn Sci 2011; 15: 56–67.
Greene J. Moral tribes: emotion, reason and the gap between us and them. New York: Penguin Press; 2013.
Haidt J. The righteous mind. New York: Pantheon; 2012.
Koenigs M, Young L, Adolphs R, Tranel D, Cushman FA, Hauser M, et al. Damage to the prefrontal cortex increases utilitarian moral judgements. Nature 2007; 446: 908–11.
Taber-Thomas B, Asp E, Koenigs M, Sutterer M, Anderson S, Tranel D. Arrested development: early prefrontal lesions impair the maturation of moral judgment. Brain 2014; 137: 1254–61.