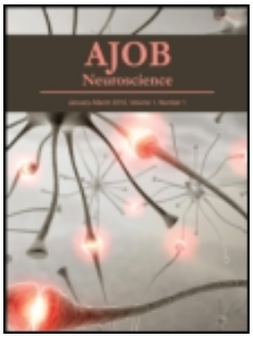

AJOB Neuroscience

ISSN: 2150-7740 (Print) 2150-7759 (Online) Journal homepage: http://www.tandfonline.com/loi/uabn20

Separating Agency From Deeds in Immoral Neurological Patients: A Potential Challenge to the ADC Model

R. Ryan Darby

To cite this article: R. Ryan Darby (2014) Separating Agency From Deeds in Immoral Neurological Patients: A Potential Challenge to the ADC Model, AJOB Neuroscience, 5:4, 25-27, DOI: 10.1080/21507740.2014.951789

To link to this article: https://doi.org/10.1080/21507740.2014.951789

Published online: 02 Oct 2014.

any ethical analysis must be based upon and proceed from fact(s), we maintain that the first (neuroethical) tradition or task should be to sustain the epistemic probity of neuroscience—inclusive of the validity of the tools and techniques utilized to develop neuroscientific (and neuroethical) theories. Then—and only then—can we begin to approach the "neuroscience of ethics" (i.e., individual and group moral and ethical cognition and action) and the "ethics of neuroscience" (i.e., the use of neuroscience in the social sphere) in a genuine and pragmatically rigorous way. We view this as one of the most crucial moral decisions to be made in and about neuroscience—and for neuroethics.

REFERENCES

Avram, M., E. Gutyrchik, Y. Bao, E. Pöppel, M. Reiser, and J. Blautzik. 2013. Neurofunctional correlates of aesthetic and moral judgments. *Neuroscience Letters* 534:128–132.

Avram, M., K. Hennig-Fast, Y. Bao, et al. 2014. Neural correlates of moral judgments in first- and third-person perspectives: Implications for neuroethics and beyond. *BMC Neuroscience* 15:39.

Doya, K., S. Ishii, A. Pouget, and R. P. N. Rao, eds. 2011. *Bayesian brain: Probabilistic approaches to neural coding*. Cambridge, MA: MIT Press.

Dubljević, V., and E. Racine. 2014. The ADC of moral judgment: Opening the black box of moral intuitions with heuristics about agents, deeds, and consequences. *AJOB Neuroscience* 5(4): 3–20.

Giordano, J. 2011a. Neuroethics: Traditions, tasks and values. *Human Prospect* 1(1): 2–8.

Giordano, J. 2011b. Neuroethics—Two interacting traditions as a viable meta-ethics? *AJOB Neuroscience* 3(1): 23–25.

Giordano, J. 2014a. What's neuroethics doing to understand—and maybe affect—morality. *Oxford Medical School Gazette* 64(1): 30–33.

Giordano, J. 2014b. The human prospect(s) of neuroscience and neurotechnology: Domains of influence and the necessity—and questions—of neuroethics. *Human Prospect* 3(3): 2–19.

Giordano, J., and R. Benedikter. 2013. Toward a systems' continuum: On the use of neuroscience and neurotechnology to assess and affect aggression, cognition, and behavior. In *Topics in the neurobiology of aggression: Implications to deterrence*, eds. D. DiEuliis and H. Cabayan, 63–85. Washington, DC: SMA Periodic Publications. Available at: http://nsiteam.com/publications.html

Glimcher, P., and E. Fehr, eds. 2013. *Neuroeconomics: Decision-making and the brain*, 2nd ed. New York, NY: Academic Press.

Levy, N. 2011. Neuroethics: A new way of doing ethics. *AJOB Neuroscience* 3(1): 3–9.

Sugrue, L. P., G. S. Corrado, and W. T. Newsome. 2005. Choosing the greater of two goods: Neural currencies for valuation and decision making. *Nature Neuroscience Reviews* 6:363–375.

Verplaetse, J., J. DeSchrijver, S. Vanneste, and J. Braeckman, eds. 2009. *The moral brain: Essays on the evolutionary and neuroscientific aspects of morality*. Dordrecht, The Netherlands: Springer.

Separating Agency From Deeds in Immoral Neurological Patients: A Potential Challenge to the ADC Model

R. Ryan Darby, Massachusetts General Hospital

Dubljević and Racine (2014) present an interesting and comprehensive attempt to explain moral judgments using the agency–deeds–consequences (ADC) model, highlighting the need for impartiality to particular moral theories when designing experimental moral dilemmas, the consideration of virtue ethics in addition to deontological and utilitarian variables, and the importance of unconscious cognitive processing and heuristics in shaping moral intuitions. However, in neurological patients with deficits in moral decision making, such as those with ventromedial prefrontal cortex (vmPFC) damage or behavioral-variant frontotemporal dementia (bvFTD), the effects of agency and deeds on moral intuitions are not independent of one and other, as suggested by the authors. In these patients, neither intentions of agents nor moral rules regarding certain harm deeds influence moral intuitions. This reflects a more general inconsistency of the agency criteria in their argument, which incorporates judgments of an agent's character and moral "goodness" over time in addition to the agent's current intentions at the time of an immoral action, which may influence moral intuitions differently. Specifically, psychological and patient data demonstrate that intentions are central to moral judgments about deeds, distinct from the influence of agent virtues or action consequences.

Address correspondence to R. Ryan Darby, Massachusetts General Hospital, Brigham and Women's Hospital, Department of Neurology, 25 Watson St., Cambridge, MA 02139, USA. E-mail: darby.ryan@gmail.com

Damage to the vmPFC leads to impaired decision making and judgments without clinically identifiable impairments in memory, executive functions, or general reasoning. Impaired judgments in these patients are detrimental, leading to inability to work, impaired personal relationships, and in some cases moral and legal transgressions. Compared with brain-damaged controls and normal subjects, vmPFC-damaged patients are more likely to approve of violations of deontological rules in favor of superior consequentialist outcomes, such as in the footbridge dilemma (Koenigs et al. 2007). Additionally, these patients are more likely to judge intended but unsuccessful harms, such as attempted poisoning, as morally permissible, suggesting that the intentional states of agents do not influence their moral judgments (Young et al. 2010). While Dubljević and Racine claim the vmPFC might be a "convergence point" for the three approaches to moral judgments and that damage to this area would impair integration of ADC intuitions while leaving each individual intuition intact, evidence from patients would suggest that consequentialist judgments are not affected after damage to this area whereas agency and deed judgments are affected. The counterexample provided regarding retained altruistic punishment in vmPFC patients (Krabjich et al. 2009) showed that patients would reject unfair offers in the ultimatum game, which Dubljević and Racine claim is evidence for retained deontological/deed-based moral intuitions. However, choosing to punish those who have wronged oneself is not a moral decision per se, and could instead be interpreted as preserved selfish motivations. The same study showed that vmPFC patients were less generous and less trustworthy, which would be in accordance with preserved egocentric motivations and decreased consideration of deontological moral "rules." Studies in this patient population therefore highlight the interdependence of agency and deeds in moral intuitions separate from evaluations of consequences.

bvFTD is a disorder characterized by loss of empathy, socially inappropriate behavior, lack of insight, and, often, moral and legal transgressions, without loss of other cognitive functions early in the disease. Anatomically, bvFTD predominantly involves atrophy in the anterior insula, anterior temporal lobes, and ventromedial and orbitofrontal lobes. In one study, 57% of patients with bvFTD had committed a sociopathic act, compared with only 7% of patients with Alzheimer's dementia. This included theft, violence, sexual misconduct, and social norm violations (Mendez, Anderson, and Shapira 2005). Patients with bvFTD are more likely to approve of deontological rule violations in favor of consequentialist decisions in moral dilemmas, despite intact knowledge of moral rules (Mendez et al. 2005). bvFTD patients also show impairments in theory-of-mind tasks (Lough et al. 2006) and have difficulty forming representations of the intentions and beliefs of others. Therefore, similar to vmPFC-damaged patients, bvFTD patients show impairments in determining intentions in agents, as well as diminished sensitivity to deontological rule violations in moral judgments, compared with brain-damaged and normal controls.

Thus, two distinct groups of neurological patients characterized by immoral behavior show impairments in using intentional states of others in moral judgments, and are more willing to violate deontological moral rules, with relatively normal use of consequences to form moral judgments. This suggests that the intentions of agents are central to normal moral judgments about deeds, and that these two concepts cannot be separated, as suggested by the ADC model. That is, the moral judgment regarding a deed is determined by its consequences and the intentions of the agent performing the action, leaving little to justify the inclusion of a separate category of deed independent of these two factors.

However, agency in the context of virtue ethics might instead be characterized as the collected behavior, character traits, and intentions of an agent over time, separate from the intentional state of an agent in that moment. While there is some circumstantial evidence that patients do not utilize character heuristics either when making moral judgments (at least in determining the value of victims; see Thomas, Croft, and Tranel 2011), the contribution of virtue ethics to abnormal decision making has not been tested directly. If vmPFC or bvFTD patients take an agent's prior virtuous qualities into account when making moral judgments, this would lend credibility to a distinction between agency and deeds. However, a meta-analysis of functional magnetic resonance imaging (fMRI) studies in normal subjects in social cognition studies concluded that short-term intentional states are more associated with activation in the temporoparietal junction (TPJ), whereas more durable intentional beliefs/traits (more akin to virtues) were associated with activation in the vmPFC (Van Overwalle 2009). Therefore, a preservation of the effects of agent virtues and moral intuitions in these patient groups seems unlikely.

An alternative hypothesis is that virtues, like intentional states of agents, are used to determine the predicted future outcomes of harmful behavior. The moral "goodness" of a person can be thought of as a heuristic for future cooperation in a social setting, with moral blameworthiness as a surrogate for prediction of future antisocial behavior. This prediction would be based on estimates of the probability of recurrence (estimation of current and prior intentional states), as well as the intensity of the moral transgressions (e.g., deeds). Virtues are incorporated by giving additional validity or doubt to estimations regarding current intentional states and future risk of transgressions and therefore would be directly incorporated into moral intuitions. In this sense, our moral intuitions are capturing judgments about the agent as a whole, with considerations to past, present, and predicted future behavior, not merely judgments regarding the morality of a single action. This is in contrast to consequentialist judgments, which focus on outcomes of the immediate action and are not dependent on vmPFC structures (while consequentialist moral theorists might incorporate predictions

of future behavior into their moral calculus, this would be a distinct, separate process from the intuitions described earlier).

In conclusion, the ADC model offers an exciting opportunity to test the interactions between agency, deeds, and consequences not only in normal subjects but also in neurological patients where dissociation of the influence of each factor in different patient groups can add validity to psychological theories regarding moral judgments. Existing evidence from patients argues against a separation of agency from deeds and suggests instead that deontological rules inherently involve determinations of intentional states of agents. Character virtues might contribute, along with current intentions of the agent and the severity of the deed, toward the heuristic of "moral goodness" and blameworthiness of the agent based on predictions of future social behavior, separate from consequentialist intuitions, which focus on the action and subsequent causative effects but not on the agents themselves. Further empirical data from both patients and normal subjects is needed to determine the validity of such an approach.

REFERENCES

Dubljević, V., and E. Racine. 2014. The ADC of moral judgment: Opening the black box of moral intuitions with heuristics about agents, deeds, and consequences. *AJOB Neuroscience* 5(4): 3–20.

Koenigs, M., L. Young, R. Adolphs, et al. 2007. Damage to the prefrontal cortex increases utilitarian moral judgments. *Nature* 446: 908–911.

Krajbich, I., R. Adolphs, D. Tranel, N. L. Denburg, and C. F. Camerer. 2009. Economic games quantify diminished sense of guilt in patients with damage to the prefrontal cortex. *Journal of Neuroscience* 29(7): 2188–2192.

Lough, S. L., C. M. Kipps, C. Treise, P. Watson, J. R. Blair, and J. R. Hodges. 2006. Social reasoning, emotion and empathy in frontotemporal dementia, *Neuropsychologia* 44(6): 950–958.

Mendez, M. F., E. Anderson, and J. S. Shapira. 2005. An investigation of moral judgment in frontotemporal dementia. *Cognitive and Behavioral Neurology* 18(4): 193–197.

Mendez, M. F., A. K. Chen, J. S. Shapira, and B. L. Miller. 2005. Acquired sociopathy and frontotemporal dementia. *Dementia and Geriatrics Cognitive Disorders* 20(2–3): 99–104.

Thomas, B. C., K. E. Croft, and D. Tranel. 2011. Harming kin to save strangers: Further evidence for abnormally utilitarian moral judgments after ventromedial prefrontal damage. *Journal of Cognitive Neuroscience* 23(9): 2186–2196.

Van Overwalle, F. 2009. Social cognition and the brain: A meta-analysis. *Human Brain Mapping* 30(3): 829–858.

Young, L., A. Bechara, D. Tranel, H. Damasio, M. Hauser, and A. Damasio. 2010. Damage to ventromedial prefrontal cortex impairs judgment of harmful intent. *Neuron* 65(6): 845–851.

Clarifications of the Descriptive/Normative Distinction in Dubljević and Racine's "The ADC of Moral Judgment"

Dale Murray, University of Wisconsin–Baraboo/Sauk County and University of Wisconsin–Richland

Dubljević and Racine (2014) make a good case that traditional moral theories still have some role to play in understanding how we actually think about descriptive moral claims. However, I don't share their enthusiasm as to how great a degree that role is. In the use of their pragmatic, pluralist, integrative model of moral decision making inspired by Dewey, they recognize that one might worry about their approach falling into a deep-seated moral relativism. Yet I don't think that their response to that charge adequately addresses the full force of the concern. Second, the authors also comprehend that there is an issue with using normative theories to figure out descriptive moral accounts, as they straightforwardly point to this type of objection and try to respond to it. Again, I am not sure they completely cast doubt aside. None of this is meant to undermine the core structure of Dubljević and Racine's intriguing view. This critique is written in the spirit of presenting challenges that will encourage further refinement of their approach.

Dubljević and Racine note that advances in neuroscience have put into question the role of traditional moral theories in explaining how we come to moral decisions. They are correct that moral philosophers have argued in favor of certain approaches to what we should do. For

Address correspondence to Dale Murray, Associate Professor, Department of Philosophy, University of Wisconsin–Baraboo/Sauk County, 1006 Connie Rd., Baraboo, WI 53913, USA. E-mail: dale.murray@uwc.edu