John Doris, *Talking To Our Selves* (Oxford: OUP), 2015

Word Counts: 59 (Abstract), 1969 (Main Text), 197 (References), 2295 (Full Text)

**Why Value Values?**

Samuel Murray, University of Notre Dame

100 Malloy Hall

Notre Dame, IN 46556

(574) 631-6471

smurray8@nd.edu, <nd.academia.edu/SamuelMurray>

(Forthcoming in *Behavioral and Brain Sciences*)

*Abstract*

Doris argues that an agent is responsible for her behavior only if that behavior expresses (a relevant subset of) the agent’s values. This view has problems explaining responsibility for mistakes or episodes of forgetfulness. These problems highlight a conceptual problem with Doris’s theory of responsible agency and give us reasons to prefer an alternative (non-valuational) theory of responsible agency.

Between 2010-2014, U.S. fire departments responded to nearly 166,100 home fires *per year* that involved some sort of cooking equipment. The National Fire Protection Association reports that 49% of these fires were the result of ‘unattended cooking’, where people forgot that they turned on the stove (Ahrens 2016: 39-41). While these house fires are costly, they aren’t nearly as tragic as Forgotten Child Syndrome. On average in the U.S., since 1998, 37 children die *per year* of heatstroke because their parents forgot them in hot cars (Weingarten 2009). In 2016, 39 children died of heatstroke in the U.S. because of parental forgetfulness (Null 2016).

If you think these cases are exceptional, consider the last time that you forgot to: call a friend on her birthday, pick up something from the store, or attend a meeting? Unless you’re a living saint, chances are you forgot one of these things recently.

I mention these cases because they raise problems for Doris’s valuational theory of morally responsible agency. Here, I want to state precisely the phenomenon that these cases capture, the problems that they raise for valuational theories of responsible agency, and alternative (non-valuational) theories that avoid these problems.

First, let me summarize briefly Doris’ valuational theory of responsible agency. Doris claims that in order for you to be responsible for some action or outcome, then you must have exercised agency in so acting or in bringing about that outcome. Further, one exercises agency when one’s behavior is caused (non-deviantly) by one’s values or a suitable subset of one’s values (Doris 2015: 24-26; hereafter, page references refer to Doris 2015 unless otherwise noted). Thus, Doris finds a tight connection between moral responsibility, agency, and value expression. This is a valuational theory of responsible agency, so-called because the theory locates responsible agency in valuative psychological states that figure causally in action production (cf. Frankfurt 1988, Watson 1975, and Bratman 2000). The contrast to a valuational theory is a reasons-responsive, or capacitarian, theory of responsible agency (see Fischer and Ravizza 1998, Wolf 1990, Nelkin 2008, Vargas 2013). Capacitarian theories state that responsible agency is a function of the possession and exercise of agential capacities. I’ll return to capacitarian theories after noting some problems with Doris’ valuational theory of responsible agency.

Consider, again, the cases mentioned at the outset. Leaving the stove on, forgetting the child in the back seat, and missing the meeting are all examples of mistakes or *slips* (cf. Amaya 2013, 2015, 2016). Briefly, an agent slips when she acts intentionally in a way that is contrary to her preferences. Slips can result from a number of different failures. For example, a slip might result from a failure to form an appropriate intention (as in the case of forgetting to call your friend on her birthday). Or, a slip might result from failing to implement an intention at the right time (as in the case of leaving children in hot cars) or failing to maintain an intention for the appropriate duration (as in the case of leaving the stove on too long). The important point is that slips result from mistakes or failures that are also intentional actions *and* that conflict with an agent’s overall balance of preferences (cf. Amaya 2013: 569).

Sometimes when an agent slips, there might be value expression. I wouldn’t turn on the stove unless I loved hot, home-cooked meals. I might also be responsible for the slip (or for the resulting fire damage to the kitchen). But, in the case of responsibility-apt slips, the value expression does not explain the responsibility. I’m not responsible for the kitchen fire because I love home-cooked meals. This point generalizes to other responsibility-apt slips: the valuational theory of agency consistently misidentifies the target of the responsibility attribution in cases of slips. When some agent slips (and is responsible for the slip), the agent is responsible for an omission (e.g., forgetting the birthday, forgetting the stove, failing to check the back seat) and that omission does not express a relevant subset of the agent’s values.

This is problematic for Doris because he claims that there is a tight connection between responsibility and value expression via exercises of agency: “Responsibility is typically associated with agency” (155) and “archetypal exercises of agency are expressions of the actor’s values” (159). Something has to go here. Either we can have exercises of agency without value expression, or we can have responsibility without exercises of agency.

Doris anticipates this objection somewhat, citing the existence of “candidates for responsible behavior that aren’t exercises of agency,” which candidates are the slips mentioned above (154). Doris responds with the claim that we shouldn’t hold people responsible for slips. And if you don’t like that response, he thinks that a weaker claim is available, namely that “responsibility is *typically associated* with agency” (155; emphasis mine). Thus, even if we concede that people are responsible for some of their slips, we can still maintain that in most cases an agent will be responsible for some bit of behavior only if some subset of the agent’s values causes the behavior in the right sort of way. There are two problems with this. First, we regularly hold people responsible for their slips (if you’re in a significant relationship with someone, then you can see this firsthand; just forget about an anniversary or special day and see how that goes for you). So we should prefer theories of responsible agency that preserve this part of our practices to theories that don’t (this aligns with Doris’ practical conservatism on 158). Second, it’s not clear to me that responsibility-apt slips are rare. After all, how often do people forget things or make mistakes? That’s partly an empirical question, but the answer is not obvious enough to warrant asserting that responsibility is *typically associated* with agency (at least if we understand exercises of agency as expressions of values). This shows that Doris cannot dismiss slips as easily as he supposes.

With respect to slips, capacitarian theories have a clear explanatory advantage over their valuational counterparts. This is because when agents slip (or when agents slip in morally significant ways), there is generally some responsibility-relevant capacity that the agent fails to exercise in some scenario where she could and should have exercised that capacity. For instance, I have recently argued that in some cases of morally significant slips, the agent exhibits a failure of vigilance that makes it appropriate to blame her for those slips (Murray 2017). Insofar as vigilance is a responsibility-relevant capacity, vigilance can figure into a capacitarian theory of responsible agency that captures our intuitive reactions to cases of morally significant slips.

Doris, however, concedes that agency and responsibility may come apart at times. This leads him to endorse pluralism with respect to agency, where diverse psychological processes can support exercises of agency. Thus, in those instances where agency and responsibility come apart (like slips), Doris thinks that non-valuative agential structures can undergird exercises of agency (174-75). Retreating to pluralism, however, does not adequately address the issue. The challenge from slips highlights a *conceptual* problem with valuational theories of morally responsible agency, not just an extensional inadequacy. The conceptual problem is that while responsibility, agency, and value expression typically coincide, they fail to exhibit the kind of explanatory relations that the valuational theory predicts.

If that diagnosis is correct, then we should ask what reason we have to accept a valuational theory. Doris offers two reasons for thinking that exercises of agency just are (or are constituted by) value expressions. The first is that taking exercises of agency to be expressions of values explains why we hold some creatures responsible and not others (24-25). For instance, we hold human beings responsible but not cats, tornadoes, or carpets. Human beings have values and express them in action, whereas the other three do not. But the capacitarian can explain this distinction. Human beings have certain capacities that underwrite certain expectations about how we will behave. Animals, natural disasters, and inanimate objects lack these capacities, so we do not hold them morally responsible for their behavior (Murray 2017: 508-9). The second reason that Doris offers is that creatures that lack values altogether seem strange targets for responsibility ascriptions (25). The capacitarian can say two things here. First, the lack of values might signal a lack of responsibility-relevant capacities constitutive of moral agency (e.g., such an agent might lack prospective memory or cognitive control). Second, the capacitarian might concede that values structure responsibility-relevant capacities while still maintaining the primacy of capacities in settling the appropriateness of certain responsibility ascriptions.

So neither of Doris’s reasons provides rational pressure to accept valuational theories over capacitarian theories, in part because the capacitarian can explain both of the phenomena that Doris cites without making any untoward adjustments to her theory. What is it that, for Doris, tips the balance in favor of the valuational theory?

The real value in the valuational theory is that it answers the skeptical challenge that Doris poses. Roughly, Doris argues that empirical evidence suggests that people are not sufficiently reasons-responsive in the way that capacitarians suppose (cf. 51-52, 171). The empirical evidence does not cast similar doubt on the role that valuing plays in our lives. In particular, the phenomenon of choice blindness provides evidence that people are not responsive to reasons. If people were responsive to reasons, then we would expect to find reversed statements in choice blindness experiments detected at higher rates (139). And given the fact that choice blindness can occur even in morally significant contexts (Hall, Johansson, Strandberg 2012), the capacitarian theorist is at a significant disadvantage. *This*, I think, is the real reason why Doris plugs for a valuational theory of responsible agency.

While choice blindness studies provide some evidence against capacitarian theories of responsible agency, the evidence is not decisive—especially given the replication crisis in psychology (cf. 44-49; I mention the crisis as someone that accepts Doris’s methodological commitment to empirically informed theorizing, so I don’t mean to be dismissive of scientific data). Also, the survey questions that Hall *et al.* use to study choice blindness are somewhat complicated. Would the effect remain if the questions were simple moral statements like ‘Torture is wrong’? Surely people’s judgments with respect to those statements would be stable. So why don’t people notice switching in Hall *et al.*’s experiments? It could be that people don’t really understand the statements. Or, it might be that people just aren’t paying enough attention to what they’re saying and doing. Surely *something* must account for the difference between seemingly stable judgments with respect to simple statements like ‘Torture is wrong’ and the emergence of choice blindness with respect to complex statements like ‘Large scale governmental surveillance of e-mail and Internet traffic ought to be forbidden as a means to combat international crime and terrorism’ (example from Hall *et al.* 2012). Here’s another, more speculative, interpretation. Perhaps with complex moral statements about the permissibility of government surveillance, people are ambivalent. They can think of reasons that support mutually exclusive positions. The initial judgment accords with the reasons that the subject finds salient at that moment. When the experimenters reverse the answer and ask for justification, perhaps the switched answer makes salient the considerations in favor of the competing position. If that interpretation is correct, then it explains why complex statements, but not simple statements (supposedly), generate choice blindness.

In any case, the capacitarian can interpret the choice blindness studies within the confines of her theory. And if she can furnish an adequate explanation of choice blindness, then we’re left without a reason to prefer a valuational theory of responsible agency to a capacitarian theory. And given the conceptual inadequacy of valuational theories (highlighted by cases of slips), I think there is significant pressure to reject valuational theories for capacitarian ones.

**Acknowledgements**

Thanks to Laura Wells and Maria Altepeter for discussion of various parts of this commentary. Special thanks to Santiago Amaya for written comments on an earlier draft.

**References**

Ahrens, M. (2016). *Home fires involving cooking equipment* (Quincy, MA: NFPA Fire Analysis and Research).

Amaya, S. (2013). “Slips,” *Noûs* 47:3, 559-576.

Amaya, S. (2015). “The Argument from Slips,” In A. Buckareff, C. Moya, and S. Rosell (eds.), *Agency, Freedom, and Moral Responsibility* (New York: Palgrave Macmillan), 13-29.

Amaya, S. (2016). “Slip-Proof Actions,” In R. Altshuler and M.J. Sigrist (eds.), *Time and the Philosophy of Action* (London: Routledge), 21-36.

Bratman, M. (2000). “Valuing and the Will,” *Philosophical Perspectives* 14, 249-265.

Doris, J. (2015). *Talking to Our Selves* (Oxford: Oxford University Press).

Fischer, J.M. and M. Ravizza (1998). *Responsibility and Control* (Cambridge: Cambridge University Press).

Frankfurt, H. (1988). *The Importance of What We Care About* (Cambridge: Cambridge University Press).

Murray, S. (2017). “Responsibility and Vigilance,” *Philosophical Studies* 174:2, 507-527.

Nelkin, D.K. (2008). “Responsibility and rational abilities: Defending an asymmetrical view,” *Pacific Philosophical Quarterly* 89:4, 497-515.

Null, J. (2016). “Heatstroke Deaths of Children in Vehicles.” From http://noheatstroke.org

Vargas, M. (2013). *Building Better Beings* (Oxford: Oxford University Press).

Watson, G. (1975). “Free Agency,” *Journal of Philosophy* 72:8, 205-220.

Weingarten, G. (2009). “Fatal Distraction,” *The Washington Post*, 03/08/09. From http://www.washingtonpost.com/wpdyn/content/article/2009/02/27/AR2009022701549.html

Wolf, S. (1990). *Freedom within reason* (Oxford: Oxford University Press).