Reflections on Anderson’s “Psychology contra Morality”

Jennifer Daigle

I want to begin by thanking Professor Anderson for sharing her paper (“Psychology contra Morality”) with us, which forms but one—and the first—piece of a much larger project on “the relation between psychological and moral explanation in the modern era” (p. 1). Many thanks, Professor Anderson.

In this paper, Anderson sets out to do at least two things: first, “to establish the ways in which psychology has posed a challenge to morality”; and second, to criticize “much of the work in cognitive psychology” for failing to “adequately capture core elements of human experience” (p. 3).

In discharging the first of these tasks, Anderson draws on research in cognitive and social psychology, which seems to support a set of three theses, what I’ll call the *undue influence, intuitionist,* and *self-justification theses*. On the undue influence thesis, “much of our thinking takes place automatically, conditioned by forms of bias, or by situational factors that ‘prime’ our response” (p. 7). Anderson connects this claim with a dual-process theory of cognition, which divides the mind into two systems: the fast, intuitive, and automatic System 1, and the slow, methodical, and effortful System 2. Here, it’s important to emphasize that the biases or situational factors affecting System 1 are supposed to be those whose influence is in some way undesirable or undue. Taking up the dual process model, the intuitionist thesis locates the source of moral judgment in System 1 operations, stating that “*moral judgments themselves* are typically made automatically and intuitively, and…moral reasoning is itself therefore necessarily post-hoc” (p. 7). Finally, the self-justification thesis posits unwitting self-justification as a “driving force” whose aim is the reduction or prevention of cognitive dissonance (p. 8).

Anderson’s critique of the psychological literature draws attention to what she calls “the more existential or meaning-laden realms of life” (p. 18), and is meant to show the crudity or shortsightedness of these theses. First, in response to the undue influence thesis, Anderson argues that at least certain of the influences discussed in the literature are not undue after all, and that seeing them as such is symptomatic of the misguided imposition of “a financial and calculative framework on human experience” (pp. 18-19). Next, Anderson objects to the intuitionist thesis by arguing that its “claim of belatedness is misleading,” citing the “existential significance” of moral reasoning (p. 17). Anderson concludes her thoughts on this literature by arguing that, because the activity of self-justification is premised on some level of awareness of wrongdoing, it’s less threatening to morality than initially supposed (pp. 19-21).

I now want to raise some questions and concerns, starting with how these psychological theses are supposed to challenge *morality*. Presumably, they are thought to do so by undermining to some degree our conception of ourselves as agents capable of autonomous, self-transparent, deliberative agency. But here one must wonder whether and, if so, to what extent morality depends on our being capacitated in these ways. Far from undermining morality, empirical work might just show that we are less good than can support the high esteem in which we hold ourselves.

If, on the other hand, Anderson is concerned, not with *morality*, but *moral explanation* (which is, after all, an element she explicitly situates within the topic of her larger project), then we need some conception of moral explanation and its explanandum, such that the psychological theses in question might plausibly be said to challenge this explanatory form. Consideration of these theses would lead us to conclude that the target phenomena are instances of *human agency*, whether overt or mental acts. Further, we might define moral explanation as an account of human agency that implicates the exercise of moral capacities. This would clarify the challenge these psychological theses are supposed to pose to moral explanation, though the obvious next question is whether the considerations Anderson adduces in response to these theses curb this challenge. (It’s not clear that they do.)

Now turning to these considerations, we see that Anderson emphasizes experience, as well as what she calls existential significance. But what’s *existential significance*? In introducing the concept, Anderson implies that something’s existential significance to somebody corresponds to the amount of effort he expends in doing that thing, writing: “The very fact that so much effort is expended in rumination and moral reflection…argues for its existential significance” (p. 17). Yet this seems unlikely. A person might stay for many years at a thankless job involving long hours and backbreaking labor, without once missing a day, though we’d be disinclined to say that the job has existential significance to him, if it’s perceived to be unimportant. This thought suggests that existential significance corresponds to felt importance or value, thus rendering it entirely dependent on individuals’ attitudes or beliefs. Yet this also seems unlikely: for suppose that I came to believe that staying home all day, eating popcorn shrimp and playing Candy Crush, is of utmost value, but that this belief was in no way reflected in my actions. Under such circumstances, it’s unlikely that these activities have existential significance to me. We might suggest, then, that something’s existential significance depends on its perceived value and the extent to which that perception is action-guiding.

Assuming that this is basically right, we are still left to wonder: what’s the significance of existential significance and experience? Anderson argues that, if we reflect on the nature of these things, we’ll see that they resist being translated into a “calculative” framework (p. 19). This is supposed to show that many apparently flawed choices aren’t so flawed after all. For example, cold-pressor studies have discovered a widespread tendency to prefer a longer period of more pain to a shorter period of less pain, so long as the ending moments of the longer period include pain of less intensity. It’s been thought that this phenomenon represents an error to be chalked up to a System 1 bias, since the addition of more bad can only make something worse. On Anderson’s view, however, this simply isn’t so, because the experiences in question aren’t reducible to quantifiable pain or pleasure units.

But notice that Anderson’s response trades on an ambiguity between reducing the *experiences* in question to relative amounts of pleasure and pain, and reducing the *relative value of these experiences* to relative amounts of pleasure and pain. We should be wary: for the mere fact that an experience has aspects that cannot be captured by such analyses might have no bearing on the legitimacy of the use of such analyses in evaluating the rationality of certain preferences. The preference under scrutiny is a prime example. After all, even if one were *disinclined* to think that [i] all pleasures are created equal, or that [ii] such pleasures are the ultimate arbiters of value, or that [iii] all values are commensurable (all disinclinations with which I’m sympathetic), I fail to see how such considerations cut against this particular analysis. For, in the case at hand, what’s being compared are relative amounts of the same stuff (namely, pain resulting from cold-water immersion), so that neither experience involves a distinctive value absent from the other.

This is to say that it’s unclear to me how an appeal to the irreducibility of experience would secure a general defense of excluding such quantitative considerations from evaluation. It’s not to say that such considerations are always possible or decisive. But even in cases where such considerations aren’t possible, I’d still caution against giving all that much weight to existential significance, as we’ve understood it. For although it follows from something’s being existentially significant to us that we value it and structure our activity around it, it doesn’t follow that we should, or that our doing so is what’s overall best. If, say, a deeply cherished relationship turns out to be overall harmful and unfulfilling, then one would do well to remove oneself from it; it might even be irrational not to. Here, the weight of the relationship’s existential significance would reside exclusively in the deep emotional pain and struggle for change that naturally accompany such shifts.

No doubt more can (and in a few moments will) be said about the topic and arguments of this paper and the questions they inspire. Their significance runs deep, potentially influencing how we should think about and evaluate our choices and actions, even how we should see and interact with the important people in our lives. It’s no surprise that these few remarks merely scratch the surface.