Rather than presenting a fully developed virtue theory for moral enhancement, Fabiano’s (2021) “Virtue Theory for Moral Enhancement” seeks traction for a virtues approach by linking virtues to five “frameworks”: the Big Five model of personality (BFM), Peterson’s model of Character Strengths (CS), Social Good Cognitive-Affective Processing Systems theory (CAPS), Social Value Orientation (SVO), and Kohlberg’s model of moral development (KMD). Each framework is subjectively evaluated for its ability to meet a number of desiderata. These evaluations are quantified on a scale from 0 (none) to 3 (high). Desiderata considered to be more important are weighted double. Summing these quantified judgments led Fabiano (2021) to conclude that SVO “is significantly superior to the other frameworks being considered” (99).

Evaluating the ability of each framework to meet each desideratum obviously requires a thorough understanding of both the frameworks and the desiderata. When psychologists have judges engage in this kind of rating task, each judge is required to be familiar with the targets of judgment and is trained in applying the rating criteria on which the targets are judged. The reliability of the judgment process is assessed by comparing the ratings of the judges. In the current case, there is only one judge, Fabiano, so the reliability of his judgments is unknown. Although we can safely assume that he understands his own desiderata used as criteria to judge the frameworks, an examination of his judgments of the five frameworks raises questions about his understanding of them. Space limitations prevent a complete accounting of possible misunderstandings, so this commentary presents some examples.

The desideratum that Fabiano (2021) calls virtue mapping refers to the degree to which the traits in a framework are either virtues themselves or at least correlate with virtues. The Big Five model of personality traits (BFM) was assigned a virtue mapping score of zero. Contrary to this judgment, however, virtues have been empirically mapped onto the BFM (Cawley, Martin, and Johnson 2000). A complete mapping can be found in the article; here I simply list a few examples. Empathy and compassion correlate with BFM Agreeableness and Openness to Experience. Discipline and scrupulousness correlate with BFM Conscientiousness. Resourcefulness and sagacity correlate with BFM Conscientiousness and negatively with BFM Neuroticism. And forgiveness and patience correlate with BFM Agreeableness and negatively with BFM Neuroticism. Fabiano’s (2021) judgment of zero compliance of the BFM to virtue mapping is clearly incorrect.

The desideratum that Fabiano (2021) calls correct balance refers to the optimum level of a virtue-related behavior. This optimum is said to vary, depending on the presence of other virtue-related behaviors and social context. According to Fabiano (2021), Chris Peterson’s model of character strengths (CS) completely fails to consider issues of correct balance. However, a chapter on character strengths and virtues (Peterson 2006, chap. 6) and a closely related chapter on values (Peterson 2006, chap. 7) show great sensitivity to issues of balance. For example, chapter 7 begins, “One of the questions raised about our work on character strengths is whether these strengths can be used for the wrong purposes. Although a character strength is by definition morally valued, a despot can be an effective leader, a suicide bomber can be brave, and a bitingly sarcastic person can be humorous … We need other considerations to deem a person good, and these include the larger goal to which his strengths of character are directed” (Peterson 2006, 165). Peterson (2006) also present an empirically derived circumplex model that explicitly maps trade-offs among virtues.
These trade-offs are similar to the trade-offs found in Schwartz’s empirically derived circumplex model of values (Peterson 2006). In his discussion of combinations of virtue, Peterson (2006) confesses that one of his early goals—finding ways to increase all virtues for all people—was unrealistic. To attempt such a task would ignore natural combinations of virtues and violate preservation of identity, which is another of Fabiano’s (2021) desiderata. The ratings of 0 for correct balance and 1 for preservation of identity do not accurately describe Peterson’s (2006) CS model.

One problem with Fabiano’s (2021) presentation of Social Good Cognitive Affective Processing Systems (CAPS) is that it is uninformed by the voluminous criticism of the ideas of CAPS creator Walter Mischel. It seems that the ratings of CAPS on desiderata compliance are based more on a derivative book by Nancy Snow than the primary, foundational literature on the CAPS model. As Snow indicates in her book, Mischel’s CAPS model arose as part of the situationist argument that behavioral stability is due to stable situations rather than to stable characteristics of people. However, after a decade of debate, Mischel conceded that stable situational influences require a degree of stability of actors’ dispositions toward those situations. Mischel argued that his cognitive-affective person variables were superior to traditional personality traits for understanding stability and situational specificity. However, a close comparison of the CAPS model with traditional personality traits reveals that CAPS variables seem to be merely different names for traditional personality traits, which already have an implicit range of situational applicability (Johnson 1999).

Unlike personality traits, however, the theoretical CAPS variables do not have widely accepted methods for reliable, valid measurement. As a consequence, the model lacks empirical adequacy and practical applicability. CAPS should not have received medium compliance with those two desiderata. Rather than representing a paradigm breakthrough, the CAPS model is little more than Hume’s belief-desire model of practical reasoning dressed up in the language of behaviorism (Alston 1975).

I have little to say about the Social Values Orientation (SVO), which Fabiano (2021) concludes “is significantly superior to the other frameworks being considered” based on his assessment of desiderata compliance. Whether or not SVO complies with the desiderata, I would point out a serious shortcoming of SVO as a general framework for virtue enhancement, namely, that SVO is limited to one moral issue: self-enhancement versus other-enhancement (Murphy and Ackermann 2014). Granted, this is a very important issue, and it is one that is considered in circumplex models of virtues and values (Peterson 2006). But how can a one-issue model serve as a general framework for the enhancement of all virtues?

Kohlberg’s model of moral development (KMD) ranked as the second most adequate framework due to its high ratings on virtue mapping and empirical adequacy and medium ratings on preservation of identity and practical considerations. However, Cawley, Martin, and Johnson (1999) empirically demonstrated that the “psychology of virtue is more closely related to personality than to Kohlbergian moral development” (1007). This is unsurprising given that KMD is founded on an ethic of duty rather than virtue and that Kohlberg dismissed the latter as a “bag of virtues” approach (Cawley, Martin, and Johnson 1999, 998). Furthermore, KMD has been heavily criticized for gender and political biases as well as its failure to actually predict moral behavior (Hogan and Emler 1978; Hogan, Johnson, and Emler 1978; Kurtines and Greif 1974).

A final comment concerns a desideratum Fabiano (2021) calls practical robustness to moral uncertainty. Fabiano (2021) believes that the frameworks that he judged to comply with this desideratum (BFM, CS, and SVO) are unlikely to prescribe actions that will lead to catastrophes. Surely this is desirable, but is it possible? Our moral traits evolved to solve conflicts in dyads and small groups, but today we interact with others at great distances in a system of complex, global relations. It is unclear whether our traditional, evolved virtues are sufficient for optimal behavior in the modern world or whether we can ever predict how an intervention will affect the future of our complex societal systems (von Hayek 1968).

REFERENCES


Fabiano makes a case for a virtue-theoretical account of moral enhancement projects in order to safeguard against some common objections like that of “moral uncertainty.” Fabiano’s account is compelling in certain ways and, given its deep familiarity with the potential problems with virtue-based moral enhancement, it is the most promising account to date. However, I see two significant worries for virtue-based moral enhancement frameworks. The first is the problem of ambiguous desiderata—what exactly are we aiming for when we morally enhance an agent? Is it actually the inculcation of virtue in the agent? Or is it merely some solution toward alleviating a bad state of affairs in the world? The second worry is one that Fabiano recognizes, namely that of conflating virtue with some behavioral trait or tendency, i.e., cooperation, impartiality, etc. Cooperation, for example, is not a virtue. One can cooperate really well in unjust projects like genocide. So, there is a problem of how to appropriately direct agents’ actions and behaviors in ways that are consistent with virtue given that the targets of moral enhancement are phenomena like behavioral traits and tendencies which are morally neutral in themselves and can be wielded toward both just and unjust ends.

THE PROBLEM OF AMBIGUOUS DESIDERATA

There are two ways to answer this question: agent-based (virtue cultivation) and outcomes-based (problem solving). If the goal is to simply solve some collective problem, then an outcomes-based strategy will do the trick. Outcomes-based moral enhancement instrumentalizes agential traits for the aim of problem solving. For example, imagine that a society finds itself on the brink of financial and infrastructural ruin given significant free riding in the form of tax evasion. Moral enhancement can help solve this problem by increasing and/or decreasing certain relevant, agential tendencies of the individuals in that population to minimize local free riding. Notice, this does not actually require that any of the individuals in the population become better people in the strong sense that virtue-theorists emphasize. One can cooperate toward a good goal but act for the wrong reason i.e. a university admissions officer who admits ethnic minorities into the university but only because he wants to impress his colleagues despite having strong, racist tendencies and sentiments “deep down.” To merely accord with the virtue by doing some good action (avoiding free riding, cooperating, avoiding racist behavior) is not sufficient for that agent performing the action to be virtuous. Actions done from virtue are only “morally worthy” if they are both good and done for the right reasons (Arpaly 2002; Markovits 2010). It’s insufficient to merely procedurally get individuals to contribute toward some good goal that their actions be properly classified as virtuous. It’s unclear...