

## Reply to Dahl: Moral Content is Varied, and Premature Definitions Should Not Constrain It

Roger Giner-Sorolla<sup>a</sup>, Simon Myers<sup>a</sup>, and Joshua Rottman<sup>b</sup>

<sup>a</sup>University of Kent, Canterbury, UK; <sup>b</sup>Franklin & Marshall College, Lancaster, Pennsylvania

Research articles in moral psychology seldom offer a definition of the scope of the word “moral”—that is, a rationale for claiming that the phenomenon under consideration falls into the moral domain. Often, the topic is assumed to be one of moral concern, as when studying life-or-death decisions, or the tension between selfish and cooperative motives. But at other times, the word “moral” seems to be affixed to a phenomenon with a wink and a nod, an implicit understanding that whatever is going on is somehow related to norms, mores, a sense of the “should” that goes beyond the merely pragmatic. To propose a clear psychological definition of morality is no easy task, and Dahl (this issue) is to be commended here for not only doing so, but leaving an explicit paper trail of traits deemed desirable for any such proposal.

However, while a rationale for calling phenomena “moral” would be useful, is it really as vital for the conduct of research as Dahl presumes? We instead argue that the definition of the term “morality” is not always a task of scientific definition similar to defining “cell” or “attitude,” but rather can be seen as a meta-scientific task for organizing research. When morality itself is a construct that figures in theories and hypotheses, this construct usually depends on lay views of whether the moral realm is involved. This insight, in our view, argues for a broader rather than narrower approach to defining morality, which we explain in the final section.

### Do We Need to Classify an Object in the World or a Field of Study?

Like Dahl, we agree that developing specific working definitions of morality will be helpful. Indeed, the process of determining a definition may help drive meta-investigations into discovering which concerns people judge to be obligatory. For example, are some concerns considered obligatory at an early stage in development and across cultural divides, while other concerns are considered obligatory later in development or only in some cultures? These may be very fruitful pursuits.

But we take issue with the assertion that to do any kind of moral psychology well, we must first define morality as a thing in itself. The analogy to defining the categories of

“planet” or “cell” breaks down because most moral psychology is not engaged in determining what actions are actually moral. Rather, morality is treated as an umbrella term that categorizes phenomena, not within a single topic of study, but across many potential topics of study. Whether a particular behavior, judgment or emotion really fits under the umbrella of “morality” has no bearing on whether a theory involving that construct is true, only whether we happen to call it a theory of moral psychology rather than filing it under some other area of psychology. For example, a theory about whether the emotion of disgust influences the tendency to punish non-normative sexual behavior can be confirmed or disconfirmed regardless of whether we classify the punitive tendency, the sexual behavior, or for that matter the emotion of disgust as “moral.”

When might the term “moral” work the same way as a physical scientific term? Only when formulating theories and hypotheses that refer to all things that are considered moral. Thus, one might theorize that disgust makes all “moral” judgments more intense, or that attitudes with a “moral” basis are more resistant to change than attitudes without. However, in this case the utility of a common term for both theories is not self-evident. Maybe both theories need to craft their own definition of what is moral, following the helpful steps proposed by Dahl in each process, but leading to two different outcomes. The disgust hypothesis might find that the effect extends even to conventional violations, as well as to violations of moralized customs such as food or sexual taboos that would be merely conventional under some definitions, including Dahl’s. Meanwhile, the attitude change hypothesis might find use for a definition of morality that is self-imposed by participants—that is, the operationalization of “morally-based attitude” could reasonably rest on people’s rhetorical willingness to call the issue “moral,” or to use adjacent terms such as “right and wrong.” Greater clarity in the definition of individual moral phenomena should not necessarily lead to a single definition of morality across phenomena.

Finding a common definition of what counts as moral, as we see it, is an issue for such roles as encyclopedists, journal editors, and funding bodies who want to classify research, but not as much for researchers themselves. By analogy, whether or not the task of determining what entities are

planets fits under the umbrella category of “astronomy” is not a problem for the scientist researching this. Instead, it concerns which heading is used to classify this research. To ascend one level further and meta-classify these meta-theories, a theory of what should count as morally relevant is an axiomatic meta-ethical theory, but a theory of what people perceive to be morally relevant is a testable theory that falls fully within the capacities and methods of psychology. Such a theory, for instance, might serve as an important grounding for the hypothesis about morally based attitudes in the previous paragraph. When these two accounts of “moral” come apart and why should be of interest to both psychologists and philosophers.

Dahl does discuss some important counterexamples, particularly the theory of dyadic morality (TDM). It is certainly true that TDM is interested in what falls under the category of morality (Gray et al., 2022; Gray & Wegner, 2011; Schein & Gray, 2018). However, we take the criticism of TDM to be not quite right. The argument presented by the author is that one needs to separate two claims:

1. The definitional claim → Immoral actions are defined as actions that result from perceived harm.
2. And (2) the empirical claim → The class of actions we have previously defined as immoral will always result from perceived harm.

Yes, to test the empirical claim (2) we must first agree on the class of actions we are talking about. But this approach is not quite the approach taken by the defenders of TDM. The empirical claim they test is this:

1. *Whenever an individual determines that an action is immoral, that individual will also perceive harm.*

Dahl (this issue) details criticism of Gray et al. (2012) based on counterexamples where evaluations of immorality would not follow perceptions of harm. In turn Schein and Gray (2018) attempt to address this by giving a wider definition of the perception of harm. Although this new definition successfully grounds TDM, it renders the perception of harm more subjective, shifting the burden of definition from the objective classifier to the subjective judge. Thus, it is possible to see exceptions to Dahl’s argument that all “empirical claims and questions [in moral psychology] are conditional on a definition of morality” (p. 55). Rather, without a specific technical definition, people can simply agree that *at least* the act of ascribing certain actions the property of “moral wrongness” fits into the moral domain. Given this simple agreement, researchers can try to find the necessary antecedents of a person attributing “moral wrongness” to any specified action. There need not be agreement on any specific technical definition in this case.

Researchers can hold their own views about what “wrongness” means and indeed participants in studies may also do so. Importantly, this approach allows us to potentially discover subsets of a population that will only indicate wrongness when they perceive a harm and discover that

individual differences measures may predict this. This can only happen if we don’t enforce a particular definition from the outset, in line with Greene’s (2007) worry about artificially narrowing the scope of enquiry of moral psychology.

Dahl further claims that by “counting as moral anything that people call ‘moral,’ the linguistic approach talks and walks like an approach to defining morality—even if the researchers do not put forth their own substantive definition” (p. \*\*). As such, they claim we need a technical definition to cope with the fact that “the word morality has sundry uses, within and between individuals” (p. 58). However, rather than presenting a problem for psychologists, this may in fact be a useful feature of this approach. Indeed it is true that “[n]o biologist would commit themselves to theorising about whatever it is that people call cells,” (p. 59) but cells have a physical reality outside the mind that morality, along with attitudes, emotions, and schemas, do not. As students of the human mind rather than of whatever mysterious single Platonic forms of morality actually exist, it makes sense for psychologists to concern themselves with whatever it is that people call “moral” and more importantly, when and why they do so. This concern usefully leaves open the possibility of productively studying multiple definitions between individuals, groups, and contexts.

It is possible to illustrate this point through particular examples from metaethics that have argued for singular definitions (or ground-truths). Emotivism (Altham, 1986; Ayer, 1952; Finlay, 2005) claims that when people use moral language they merely express an emotional response, for example, the phrase “murder is wrong” would be the equivalent of saying “boo murder.” Expressive Assertivism (Boisvert, 2008) instead argues that the statement entails a command to not murder (akin to asking someone to shut the window). Neither of these declarations would be truth-apt. In contrast, Moral Realists claim that moral language isn’t just expressions of emotions or commands but rather statements that can be rendered true or false by objective facts about the world (Sayre-McCord, 2021).

So, is it the case that people intend to express truth-apt statements of cognitive content when they use moral language? Rather than establishing a single ground-truth, psychologists are able to be pluralistic. We know that there is considerable variation in use of moral language both between individuals and across different contexts (Goodwin & Darley, 2008, 2012; Yilmaz & Bahçekapili, 2018) and across the lifespan (Beebe & Sackris, 2016, although see Bush & Moss, 2020). By investigating people’s lay views, we can discover that sometimes people are simply declaring their basic emotional reaction when they say something is wrong, while at other times they may intend to say something more complex. As psychologists, without declaring a priori what “morally wrong” should mean, we can find out what people mean under different circumstances and why they do so, treating each meaning as potentially true at any given time in any particular context. Steel (2004), for example, promotes this approach and calls it Pluralistic Reductionism.

## The Scope of Research on Moral Psychology

By raising these points, we do not mean that that the quest for technical working definition(s) is not helpful; it is still very likely to be. Yet, we disagree with Dahl on his claim that while “having a definition might render the inquiry too narrow, lacking a definition might render the inquiry too broad” (p. 56). We think the opposite may instead be true. Within the field of moral psychology, lacking consensus on clear working definitions, the scope of some proposed definitions including Dahl’s may currently be too narrow.

Descriptively, Dahl’s definition encompasses the typical focus amongst developmental psychologists studying morality, who have largely relied on similar definitions for the past half-century. Social Domain Theory has dominated the field since Turiel’s landmark contributions demonstrating that children do not progress through discrete stages of moral development but rather appreciate different kinds of normative concerns by early childhood (e.g., Killen & Smetana, 2015; Turiel, 1983). This framework has produced many important findings that have considerably advanced our understanding of moral cognition. However, it has been limited. By focusing narrowly on “obligatory concerns with others’ welfare, rights, fairness, and justice,” developmental psychologists have missed out on some of the important phenomena that inform people’s evaluations in the moral domain.

This is not unique to developmental psychology. Indeed, the vast majority of moral psychology has focussed on attributions of wrongness and blame, leaving character-based judgements, partner choice, moral praise and supererogation as relatively overlooked. Recent calls by social psychologists, who typically have not relied on the restrictive definition used by Social Domain Theorists, emphasize the need to actually use the possibility of broader definitions of morality in the active choice of research topics (e.g. Anderson et al., 2020; Landy & Uhlmann, 2018; Martin et al., 2019; Pizarro & Tannenbaum, 2012).

Consider a person who has eaten a taboo food or who has had sex with their sibling. These actions are typically considered immoral by lay people for reasons that go beyond concerns with autonomy or justice (e.g., Gutierrez & Giner-Sorolla, 2007). In addition to having moral concerns about others’ bodily conduct, people possess a range of other disparate and potentially arbitrary concerns that factor into their moral considerations. If we adopt Dahl’s definition of morality, we may be in danger of overlooking many of the factors that people consider when assessing others’ moral character. For example, even though caring for animals is considered substantially less obligatory than caring for humans (Caviola et al., 2019), we tend to think that moral character is more heavily implicated by harming animals than by harming people (Tannenbaum et al., 2011).

It is important to note that researchers can disagree, for example, about whether these kinds of character judgments (or other mental or behavioral tendencies, like selecting cooperative partners) fit under the umbrella of morality. However, these disagreements have no bearing on the truth

of theories about character attributions, partner choice, or other phenomena that moral psychology researchers might choose to study. Once again, they will mainly affect meta-scientific decisions such as who funds such research and where it is published. And importantly, if people who self-identify as moral psychologists had not included these phenomena in their research programme, we might have missed out on a rich set of tendencies that deeply inform how people make judgments about obligatory concerns regarding welfare and justice.

Developing a definitional framework of what morality entails will help to direct us down a shared path of investigation, but it runs the risk of fundamentally distorting our observations in a way that misses critical features (Rozin, 2001). By narrowly restricting the study of moral psychology in the way that Dahl suggests, developmentalists might also miss out on important processes of socialization that may be uniquely critical for acquiring more arbitrary moral concerns (Rottman & Young, 2015). And moral psychologists more generally might miss out on evaluations that are central to everyday moral cognition. We advocate taking a more bottom-up, data-driven approach to understanding the scope of moral psychology (rather than imposing a top-down definition). This will allow researchers to capture certain thick moral concepts that are unique to particular cultures and provide important insights into, for example, how others evaluate what makes a “good person” (in addition to evaluating morally right actions).

## A More Inclusive Definition of Morality

If we would be asked to supply our own candidate definition of morality for taxonomic purposes, one possibility might be to emphasize the functionality of moral norms (Giner-Sorolla, 2012). Moral systems work to reduce focus on the individual interest here and now, and to boost consideration of other people, or even the future perspective of an individual. When studied at the group level, an act is relatively more moral if it promotes the interest of an appropriate higher level of social organization over a lower one (e.g., the interests of the whole nation over the interests of a single family or community). Supporting this focus is a sense of moral rules being normatively more important than other concerns, from which we derive the observation that they are often seen as obligatory.

Consider an example from Spain in the last century, of a man—a member of a radical left party—who refuses to take off his hat as the Holy Eucharist passes by in a procession. If this is simply seen as an eccentric, unconventional action, it is not likely to offend the function of social cohesion. But if flouting the rules of respect is seen as an affront to God, to the literal Body of Christ, and to the social institution of the Church, then it is hard to deny the moral nature of this act. This definitional judgment would coincide with lay standards affirming that, to a believer, disrespecting the Host is a moral offense. The opprobrium leveled at the hat-wearer could be seen as fulfilling the function of a moral

rule, affirming a larger social organization (Catholic Spain) at the expense of individual freedom or the beliefs of a minority political party. Psychologically it is moral, regardless of whether we, from our standpoint of values, think it objectively to be moral.

The functional definition has the advantage of flexibility. It covers many phenomena heretofore grouped under the moral umbrella: obviously the topic of pro- and antisocial behavior, but also the following prominent and well-researched topics:

- Dilemmas in decision making (how exactly harm to some should balance against benefit to others),
- Moral condemnation (taking individual effort and risk to punish and decry wrongdoers),
- Reactions toward taboo violations in food, sexuality, and religious-facing behavior (individual liberty vs. collective moral norms)
- Attitudes having a moral component (connected to larger-scale and high-importance social concerns that outweigh pragmatism)
- Moral character (tracking individuals for their potential to support higher-order projects)

Still, despite the surface aptness of this taxonomic definition for grouping the research areas, the working definitions of morality used in each research area can differ. For some, dealing with specific types of decisions and actions that we classify as moral (such as dilemmas), the definition of morality is not necessary to formulate theories. For others, such as the moral labeling of attitudes and character, it would be important to know whether participants themselves classify the phenomena as “moral” by subjective criteria. We think that this is an adequate and desirable state of organization in moral psychology. That is, we need not impose an unbending definition on a variety of individual phenomena, nor should we avoid drawing on lay perceptions of the moral realm as a way to operationalize some theories of moral psychology.

## References

- Altham, J. E. J. (1986). The legacy of Emotivism. In G. F. Macdonald & C. Wright (Eds.), *Fact, science and morality: Essays on A.J. Ayer's Language, Truth and Logic* (pp. 275–288). Basil Blackwell.
- Anderson, R. A., Ruisch, B. C., & Pizarro, D. A. (2020). Differentiating between different forms of moral obligations [Review of *Differentiating between different forms of moral obligations*]. *The Behavioral and Brain Sciences*, 43, e57. <https://doi.org/10.1017/S0140525X19002589>
- Ayer, A. J. (1952). *Language, truth, and logic*. Courier Corporation.
- Beebe, J. R., & Sackris, D. (2016). Moral objectivism across the lifespan. *Philosophical Psychology*, 29(6), 912–929. <https://doi.org/10.1080/09515089.2016.1174843>
- Boisvert, D. R. (2008). Expressive-assertivism. *Pacific Philosophical Quarterly*, 89(2), 169–203. <https://doi.org/10.1111/j.1468-0114.2008.00315.x>
- Bush, L. S., & Moss, D. (2020). Misunderstanding metaethics: Difficulties measuring folk objectivism and relativism. *Diametros*, 17(64), 6–21. <https://doi.org/10.33392/diam.1495>
- Caviola, L., Everett, J. A., & Faber, N. S. (2019). The moral standing of animals: Towards a psychology of speciesism. *Journal of Personality and Social Psychology*, 116(6), 1011–1029. <https://doi.org/10.1037/pspp0000182>
- Finlay, S. (2005). Emotive theory of ethics. In *Encyclopedia of philosophy*. Macmillan.
- Giner-Sorolla, R. (2012). *Judging passions: Moral emotions in persons and groups*. Psychology Press.
- Goodwin, G. P., & Darley, J. M. (2008). The psychology of meta-ethics: exploring objectivism. *Cognition*, 106(3), 1339–1366. <https://doi.org/10.1016/j.cognition.2007.06.007>
- Goodwin, G. P., & Darley, J. M. (2012). Why are some moral beliefs perceived to be more objective than others? *Journal of Experimental Social Psychology*, 48(1), 250–256. <https://doi.org/10.1016/j.jesp.2011.08.006>
- Gray, K., & Wegner, D. M. (2011). Morality takes two: Dyadic morality and mind perception. In Mikulincer, M., & Shaver, P. R., (Eds.), *The social psychology of morality: Exploring the causes of good and evil* (pp. 109–127). American Psychological Association.
- Gray, K., MacCormack, J. K., Henry, T., Banks, E., Schein, C., Armstrong-Carter, E., Abrams, S., & Muscatell, K. A. (2022). The affective harm account (AHA) of moral judgment: Reconciling cognition and affect, dyadic morality and disgust, harm and purity. *Journal of Personality and Social Psychology*, 123(6), 1199–1222. <https://doi.org/10.1037/pspa0000310>
- Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological Inquiry*, 23(2), 101–124. <https://doi.org/10.1080/1047840X.2012.651387>
- Greene, J. D. (2007). The biology of morality: Neuroscientists respond to Killen and Smetana. *Human Development*, Letters to the Editor. <http://www.karger.com/hde/>
- Gutierrez, R., & Giner-Sorolla, R. (2007). Anger, disgust, and presumption of harm as reactions to taboo-breaking behaviors. *Emotion (Washington, D.C.)*, 7(4), 853–868. <https://doi.org/10.1037/1528-3542.7.4.853>
- Killen, M., & Smetana, J. (2015). Origins and development of morality. In M. Lamb & C. Garcia-Coll (Eds.), *Handbook of child psychology (Vol. 3). Social and emotional development* (7th ed., pp. 701–749). Wiley-Blackwell.
- Landy, J. F., & Uhlmann, E. L. (2018). Morality is personal. In K. Gray & J. Graham (Eds.), *Atlas of moral psychology* (pp. 121–133). Guilford.
- Martin, J., Young, L., & McAuliffe, K. (2019). *The psychology of partner choice*. Retrieved from <https://psyarxiv.com/weqhz/>.
- Pizarro, D. A., & Tannenbaum, D. (2012). Bringing character back: How the motivation to evaluate character influences judgments of moral blame. In M. Mikulincer & P. R. Shaver (Eds.), *The social psychology of morality: Exploring the causes of good and evil* (pp. 91–108). American Psychological Association.
- Rottman, J., & Young, L. (2015). Mechanisms of moral development. In J. Decety & T. Wheatley (Eds.), *The moral brain: A multidisciplinary perspective* (pp. 123–142). MIT Press.
- Rozin, P. (2001). Social psychology and science: Some lessons from Solomon Asch. *Personality and Social Psychology Review*, 5(1), 2–14. [https://doi.org/10.1207/S15327957PSPR0501\\_1](https://doi.org/10.1207/S15327957PSPR0501_1)
- Sayre-McCord, G. (2021). Moral realism. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/sum2021/entries/moral-realism/>
- Schein, C., & Gray, K. (2018). The Theory of dyadic morality: Reinventing moral judgment by redefining harm. *Personality and Social Psychology Review: An Official Journal of the Society for Personality and Social Psychology, Inc.*, 22(1), 32–70. <https://doi.org/10.1177/1088868317698288>
- Steel, D. (2004). Can a reductionist be a pluralist? *Biology & Philosophy*, 19(1), 55–73. <https://doi.org/10.1023/B:BIPH.0000013263.35121.5a>
- Tannenbaum, D., Uhlmann, E. L., & Diermeier, D. (2011). Moral signals, public outrage, and immaterial harms. *Journal of Experimental*

- Social Psychology*, 47(6), 1249–1254. <https://doi.org/10.1016/j.jesp.2011.05.010>
- Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge University Press.
- Yilmaz, O., & Bahçekapili, H. G. (2018). Meta-ethics and the mortality: Mortality salience leads people to adopt a less subjectivist morality. *Cognition*, 179, 171–177. <https://doi.org/10.1016/j.cognition.2018.06.014>