aps ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Perspectives on Psychological Science
1–8

DOI: 10.1177/17456916231187404
www.psychologicalscience.org/PPS

Sage

Explaining Social Normativity: Introduction to the Discussion Forum on Cecilia Heyes's "Rethinking Norm Psychology"

Daniel Kelly
Department of Philosophy, Purdue University

It is an exciting time to be doing research on norms, and the conversation contained within these pages provides a terrific example of why. The contributions are enormously stimulating, full of interesting insights and probing exchanges. At the center is Cecilia Heyes's wonderful target article "Rethinking Norm Psychology," which has two main aims. The first is to reconsider and criticize what has become an influential account of normative cognition, one that, in her view, is too heavily invested in innate psychological structure. The second is to make a case for her own preferred alternative, one that construes human social normativity as underpinned by a number of different psychological processes that together take the form of a cognitive gadget (Heyes, 2018). In making her negative and positive cases, she also set out several methodological proposals for how the field might better frame hypotheses and use evidence to address theoretical disagreements. The commentaries illuminate Heyes's position, addressing it from a range of perspectives that include the philosophy of archeology; the philosophy of biology; animal cognition; cognitive science; evolutionary neurobiology; cognitive neuroscience; and cultural, developmental, and social psychology. The comments and Heyes's reply to them are careful, thoughtful, and productive, generating light but little, if any, of the distracting kind of heat. One obvious takeaway is that this is a record of a thriving field, full of ideas and energy. It has already achieved a high standard of interdisciplinary work and is capable of conducting the kind of productive and fruitful dialogue that will allow it to continue making important progress.

Lest this comes off as too self-congratulatory, I note that as rich as it is, the issues and disagreements at the center of this exchange address only a subset of those that arise about norms and norm-related phenomena. This is important because as Heyes and several commentators mentioned, norms are ubiquitous. Human affairs are saturated with them, and they regulate enormous swathes of people's lives. It is unsurprising, then, that norms are relevant to the agenda of nearly every discipline that studies human affairs and are just as relevant to anyone outside of academia who wants to influence them. This ubiquity and the motley crew of interested parties it brings along with it raises familiar challenges. Some have to do with the lack of a precise common vocabulary shared by everyone addressing norms and norm-related phenomena. This can lead to a frustratingly common pitfall of interdisciplinary work: researchers and practitioners from different perspectives talking past each other, failing to formulate, let alone make progress on, genuine disagreements at all. Thankfully, I think this kind of problem has been largely (although perhaps not totally) avoided here. Perhaps this benefit is purchased at the cost of not having *every* interested perspective represented—there are no economists, political scientists, or sociologists in this exchange, for example. But this is almost certainly a price worth paying for now. Not every interested party is interested in the same aspects of norm-related phenomena, and not all debates can be had at once. Moreover, within the group represented here, plenty of substantive disagreements remain to be worked out.

Those disagreements arise against the background of a shared interest in the psychological and evolutionary roots of norms. Indeed, I see broad agreement about the utility of what has become known as the cognitive-evolutionary approach to normative cognition (Kelly & Davis, 2018; Kelly & Setman, 2020). Heyes adopted this approach in her target article, in which she used its perspective and drew on its vocabulary to

Corresponding Author:
Daniel Kelly, Department of Philosophy, Purdue University
Email: drkelly@purdue.edu

organize her agenda. Both Heyes's cognitive-gadget account and the account she set up as her foil, Sripada and Stich's (2007) nativist account, agree that normative competence and norm-guided behavior in typical adult humans is underpinned by what is neutrally called a *norm psychology*. The commentaries also took this possibility seriously (although see Sterelny for an expression of doubt). So, let norm psychology be the cluster of traits—a set of specialized behavioral and attentional dispositions; cognitive, affective, and motivational processes; and mental representations—that gives humans the ability to "do norms."

This common background may look trivial, but it is not. The shared focus on the details of and sophisticated hypotheses about the psychology distinctive of norms—rather than on, say, carefully cataloguing the norms present in a particular group, tracking the historical spread of a specific norm through a region, or arguing about whether an entrenched norm is justified—distinguishes these researchers from others who work on norms, such as cultural anthropologists, historians, classical economists, and moral philosophers. Here there is also shared agreement that the range of mental entities that can be appealed to in the investigation of *norm* psychology need not conform to or be constrained by those of *folk* psychology. Rather, hypotheses about norm psychology can draw from the theoretical repository of the cognitive sciences and so might invoke processes and use ideas that are unfamiliar to common sense (for an extreme form of this position, see Richerson and Gavrilets). In addition, the shared evolutionary lens invites researchers to think about the different components of norm psychology in terms of the functions they perform and thus in terms of the common adaptive problems and selection pressures that shaped them. That same lens helps organize and inform investigations into the typical ontogeny of human normative capabilities as they emerge over the course of childhood and into adulthood (Germar and Mojzisch; Taumoepeau) and inquiries into the possibility of social norms in nonhuman animals (Moore and Monsó; Westra and Andrews).

Within this framework, the main substantive issue in the present exchange is as follows: What is the nature and provenance of adult-human norm psychology? This ur-question divides into smaller and more specific ones, many of which are the focal point of different commentaries. Perhaps the most important questions for Heyes's agenda are on the provenance front. How much and which parts of the structure of norm psychology—the information-processing machinery, or *mills* in her terminology—are innate, part of a genetic inheritance? How much and which parts are learned from individual experience or are acquired from others via cultural transmission? Analogous questions can be raised about the content, the *grist* those mills operate on. Questions of this sort might inquire about the provenance of proprietary concepts used by the norm system (e.g., concepts for norm, wrong, obligation, punishment) or about specific norms or types of norms (e.g., harm norms, incest norms, care norms, fairness norms).

Questions about grist are interesting but do not get much attention here. Both Heyes's view and the nativist orthodoxy she set up as her stalking horse accept that norm-psychology grist is mostly culturally acquired, picked up intuitively in the course of social interactions with other community members (for a sophisticated defense of the view that some norms are innate, see Mikhail, 2011; for a pluralist view on which specific norms can be formulated, avowed, and adopted by mature individuals on their own volition, see Kelly, 2022). The most distinguishing feature of Heyes's view, however, is the nuanced answer she gave to the question regarding the origins of the structure of norm psychology, the mills that operate on the grist. Appreciating it requires keeping questions about provenance separate from questions about functional specialization and domain specificity. She held that adult-human minds typically do contain norm-specific psychological machinery but that much more of it is socially learned than is implied by the orthodox nativist account. Call the learned package of specialized machinery posited by this view the *norm gadget*. Heyes held that a large part of norm psychology, especially that responsible for what she called "explicit normativity," is comprised by this norm gadget. But she also held that although the norm gadget is itself made up of processes that are learned and specialized, the *acquisition* of that gadget is underpinned by learning processes that are themselves innate. Moreover, those innate learning processes are *not* specialized, *not* specifically dedicated to norms, norm psychology, or normative behavior. Rather, they are the same domain-general processes (or at least the same types of computational processes, see Vogel and Lockwood) that are used to learn all manner of things. They are relatively unrestricted in the kind of content they can absorb or the type of psychological structure they can help to install in an individual's mind.

As with other cognitive gadgets in Heyes's view (gadgets for literacy, arithmetic, mind reading, etc.), the norm gadget can be thought of as akin to a particular app available for smartphones; it does not come factory installed, but it can be downloaded like other apps on offer at various app stores. The domain-general processes that make this downloading possible *do* come factory installed, and they are the same ones that are used to install any of the apps on offer, regardless of the specifics of an app's proprietary functions,

programs, and subject matter. Individual human minds, the analogy goes, do not have an innate norm psychology, but they do have the general capability to acquire gadgets, any of the distinct, more or less functionally integrated packages of specialized and domain-specific mental processes that are available in a culture. In Heyes's view, the norm gadget is one of those packages, available in every known contemporary culture (every extant app store, in the analogy) and perhaps most of those stretching far back into the history and evolution of the human species itself (Pain; also Boyd, 2017; Henrich, 2015). This view also adds a twist to the "evolutionary" part of the cognitive-evolutionary approach to norms by claiming that the norm gadget did, in fact, evolve—it was not designed by a single inventor, nor is it constructed anew by individuals on their own—even though it is not a product of *natural* selection. Rather, the view sees a much bigger role for culture. The norm gadget is a package of abilities that is not only culturally inherited but that was itself developed and pieced together by cultural-evolutionary processes acting over the course of generations, slowly refined by cultural-selective pressures to perform the set of functions it does.

Although many offered qualifications, amendments, slight tweaks, or skepticism about details, none of the commentaries directly challenged the heart of this story about evolutionary origins, and none took the opportunity to defend a more nativist account—although I mention the main point that such a defense might be built around below. Most of the action in the comments is over questions not about provenance but nature: What is the *character* of the component parts that comprise adult-human norm psychology, wherever they come from? How do they interact with each other to produce normative judgment and behavior? Are the processes involved in *learning* norms—individual norms or even the norm gadget itself—and *doing* norms completely distinct (Theirault)? Or, alternatively, might the operation of the norm gadget be best understood as a virtual machine that is realized in or implemented by more fundamental computational structures found in the brain, structures that perform the kinds of domain-general learning functions that Heyes's acquisition story appeals to? Which of the processes in the norm gadget are implicit, and which are explicit? Or, might the implicit–explicit distinction and the dual-process framework often used to understand it obscure more than they clarify (Germar and Mojzisch)? What are the subcomponents of norm psychology, and what are the functions of those subcomponents? Are they fundamentally cognitive, affective, or some mixture of both (Birch)? How does a norm, once acquired, become intrinsically motivating, and how is that motivation produced (Theirault; also see Kelly, 2020)? Does norm psychology involve proprietary forms of mental representation, perhaps representations of rules (Westra and Andrews), or shared intentions (Schmidt, Vaish, and Rakoczy), or a concept of normalcy (Knobe)? What is the relationship of normative cognition to language (Moore and Monsó)? To mentalizing and discursive reasoning (Taumoepeau)? To the psychology of social identity and status (Kish Bar-On and Lamm)? To particular emotions such as anger, disgust, guilt, or increasingly, outrage (Brady and Crockett)?

Many of the comments also remark, head on or more obliquely, on methodology as well, considering what forms of evidence—developmental, cross-cultural, historical, interspecies—are most useful in investigating different aspects of norm psychology. These mostly cosign onto Heyes's poverty–wealth schema, an analytic device she constructed by developing the logic of classical poverty-of-the-stimulus arguments, and that is useful for determining when one is and is not justified in positing that some psychological trait is innate. Many also endorsed her call to use that schema in the service of what she called "contrastive hypothesis testing." This is a method of developing experiments and marshaling evidence that explicitly aims to pit different theoretical accounts against each other. Heyes put it forward as an alternative to merely confirmatory hypothesis testing that gathers and views data through the lens of a single theory, asking only whether the data it finds are compatible with it (although see discussion, pp. 132–137).

Heyes responded to each of these comments, organizing her replies around three issues: the explanatory target, her model, and the evidence. Rather than try to adjudicate each of those exchanges directly, I leave it to the reader to dig into the details, free of any prior and potentially skewing editorial opinion. Instead, I try to fill in some tacit but relevant context. I also call attention to places in which the discussion helps to sharpen a question and raises a new issue, sheds new light on the significance of a known one, or just unsettles existing (or at least my own) views. Hopefully, these will turn out to also be topics that deserve and will reward careful attention from others going forward. Even better, perhaps calling attention to them will help careful readers see in the exchanges something that I missed.

Normative Phenomena and Their Explanations

Many of the commentaries followed Heyes in framing their points in terms of *the* explanatory target. I remain suspicious of this seemingly innocent locution and worry that it masks important complications. Taking a

step back, it is always a compelling challenge to keep straight all the candidate explanatory relations in discussions such as these, in part because there are so many of them about. When there is an evolutionary framework in force as there is here, both proximate explanations and ultimate evolutionary explanations are often in play. Although ideally consistent with each other, these do fundamentally different explanatory work and so benefit from being kept analytically distinct. Likewise, when questions about typical developmental trajectory arise, there will also be ontogenetic explanations in the mix. Those, the specific kinds of evidence they appeal to, and the particular ways in which they bear on different phenomena all need to be handled carefully and on their own terms as well. The same can be said for comparative questions about interspecies behavior and the associated stock of phylogenetic explanations and evidence. Evidence of within-species, intergroup variation between human groups may reflect differences in culture but may not (see Kelly & De Block 2022, especially Section 2.2).

Even in relatively straightforward cases, unconcerned with evolution, ontogeny, or variation, explaining behavior typically involves appeal to circumstances and external stimuli together with reference to processes internal to the actor whose behavior is being explained—but accounts of what goes on inside the head of the actor can be pitched at different levels of explanation as well (as Vogel and Lockwood recognized in their discussion and its invocation of Marr's influential tripartite distinction). In all these cases, candidate explanations that appear to be in competition with each other may be complementary. They may be aiming to account for different aspects of the complicated phenomenon that is under consideration, so there may be several related but subtly different explanatory targets. Likewise, theorists might be offering answers to different "how" or "why" questions (e.g., Tinbergen, 1963) or just describing the same thing at different levels of granularity in different vocabularies. When things work out especially well, these different descriptions and explanations can mutually inform and constrain each other, culminating a more complete and satisfying overall account.

This kind of maximally informative perspective may eventually be achieved for norm psychology, and although the field is surely getting closer, I do not think it is in striking distance quite yet. The challenges are exacerbated by several factors specific to norms. First, the terms "normative" and "normativity" have come to be applied to a much broader range of phenomena than those at issue in this conversation. There is a primitive kind of normativity found in entities that can engage in basic forms of self-directed behavioral regulation by doing simple course *correcting*, using processes that can be modeled as negative feedback loops and seen even in artifacts such as thermostats and heat-seeking missiles. There is the normativity found in evolved traits that can function *properly*, performing those tasks whose effects were selected for, or that can *malfunction* and fail to produce those selected effects (Griffiths, 1993; Millikan, 1989; Neander, 1991). There is the normativity of individual intentional agents, whose behavior makes sense only if they pursue their desires, given their beliefs, in a way that is *rational* and who are otherwise acting *irrationally*. None of these are the kind of normativity that Heyes and most of her commentators are intent on.

Perhaps this is obvious, but I find it helpful to approach the central phenomenon slowly and carefully. The form of normativity at interest here is what can be distinguished as *social normativity*. Its primary manifestation is not in the functioning of individual traits or the behavior of individual agents but, rather, at the level of *groups*. Social normativity requires at least two agents but typically involves the collective behavior of entire communities. Somehow or other, individual humans manage to interact with each other in ways that produce what Westra and Andrews usefully picked out as *normative regularities*. These are a species of real pattern (Dennett, 1991). Like other real patterns, a normative regularity is an objective phenomenon, one that genuinely exists in the world independently of any observer (even if it can be seen or known only by entities that have the requisite conceptual and epistemic capabilities). And like those under consideration in Dennett's (1991) introduction of the idea of real patterns, normative regularities are also patterns of *behavior*. However, Dennett's initial discussion was focused on patterns found in the behavior of individual intentional agents. Normative regularities, in contrast, are realized not in the sequence of actions exhibited by a single individual but, rather, at the level of groups, where they emerge out of the *interactions* between the agents that make up the group.

A second challenge for norm psychology arises here. Assume normative regularities are in fact the primary phenomenon of interest, the main type of stylized fact that norm psychologists aspire to account for. This would be reasonable. On the one hand, a normative regularity is the right *kind* of thing to be an explanatory target for a psychological theory: It is behavioral, and it is the *job* of psychological theories—whatever kinds of mental entities they appeal to and whatever the provenance of those entities—to explain behaviors and patterns of behavior. On the other hand, a normative regularity is a mildly nonstandard explanatory target for a psychological theory. A normative regularity is *not* a chunk of behavior that can be found in or exhibited

by a single individual—it is, again, a piece of collective behavior, realized at the level of groups—but a norm-psychological explanation of a normative regularity will appeal to the processes, states, and other mental entities that *are* found in the minds of individuals.

Of course, this kind of mereological challenge is not unique to theories of norm psychology; something similar arises in other areas of social psychology as well. But this extra leap in "levels" between collective explanandum and individual explanans opens up space between the main phenomenon to be explained and the types of entities that can be appealed to in order to explain it. This, in turn, creates more wiggle room for debate and disagreement. For example, it may contribute to a terminological ambiguity. The word "norm" itself is sometimes used to refer to the pattern in collective behavior, where it can be read as shorthand for what I have been here calling a "normative regularity." In other instances, however, the same term, "norm," is used as the name of a posited mental entity, often a representation of a rule, appealed to in psychological explanation and implicated in producing the behavior of an individual who participates in a normative regularity. Of course, on pain of a notorious kind of Molièrean vacuity, whatever a norm is, it cannot be both the explainer and the explained.

That extra space may exacerbate another familiar kind of challenge that I think remains unresolved in this discussion. All parties here take seriously the idea that adult humans have a norm psychology, understood as a more or less functionally integrated package of domain-specific mental processes that underpin their ability to "do norms." Those scare quotes can now be removed, and the original placeholder of "do norms" can be replaced with "participate in normative regularities": Norm psychology is the specialized subsystem of a human mind that allows the individual to participate in normative regularities. This is an improvement, but questions remain. One is how to distinguish norm psychology proper from other, distinct subsystems found in human minds, especially those that often work alongside and in conjunction with norm psychology but are not parts of the package of functionally dedicated *processes* that are specifically *for* social normativity. For example, participating in normative regularities typically involves perceptual cognition (visual, auditory, etc.). It would be unreasonable, however, to conclude from this that perceptual processing is a part of norm psychology. The field still lacks agreement about what falls within the boundary of norm psychology and which parts of the mind are merely adjacent to it. Call this the "boundary problem" (e.g., Cummins, 2000).

A similar problem arises for so-called moral psychology and the moral domain (see Sinnott-Armstrong & Wheatley, 2012, 2014; Stich, 2018; Young & Dungan, 2012), but I think normative psychologists have better resources to address their version of the problem than moral psychologists do theirs. One strategy for solving the boundary problem is by appealing to the character of normative regularities themselves. Most would agree that these are but one of many kinds of patterns identifiable at the level of groups, different from other collective regularities such as conventions, customs, traditions, and fads. What is distinctive about normative regularities, the line of thought might go, is the special way they are stabilized and sustained by the individuals who participate in them (and perhaps also the ways they are created, changed, and abolished, à la Germar and Mojzisch). So, there is a correspondingly special subset of behaviors that individuals engage in, and any given collective-level pattern qualifies as a normative regularity, rather than, say, a fad or a convention, only if it is produced by individuals engaging in the behaviors in that special subset. Norm psychology would then include the set of mental processes dedicated to producing the behaviors in that special subset and exclude all else.

Described at this level of abstraction, this seems to capture the strategy broadly in play. (Compare Bicchieri, 2016, who started with mental states and then distinguished different kinds of group-level patterns directly by appeal to the kinds of mental states that stabilize them.) There are common themes concerning what is in the special subset and so which behaviors fall within the scope of norm psychology. Heyes would point to compliance, enforcement, and commentary as comprising the special subset of individual behaviors, at least for explicit normativity; Stich and Sripada would point to compliance and punishment; Westra and Andrews (2022), the first to introduce the terminology of normative regularities, explicitly characterized them in terms of behavioral conformity and social maintenance. Norm psychology, on each account, is the package of mental processes functionally specialized to produce those key behaviors.

Viewed from this perspective, this discussion sharpens many questions. How might theorists develop arguments for the exclusion or, more interestingly, the inclusion, of other key behaviors in this subset? For example, future work can continue to explore the connections—psychological, developmental, evolutionary—between human social normativity and the psychology of status and social identity (Kish Bar-On and Lamm; also see Smaldino, 2019). One might make the case that in humans, the two coevolved with each other, and the most fundamental kinds of normative regularities do *not* involve everyone within a community or group but are, rather, identity- and role-specific. Perhaps impartial or

impersonal norms that apply to everyone in a group regardless of rank or station are a special case, an exception to the rule. Likewise, the relationship between natural language and linguistic capacities, on the one hand, and norm psychology, on the other, seems like an avenue of inquiry that will benefit from the clearer picture of social normativity and norm psychology that is emerging here (Taumoepeau; cf. Moya & Henrich, 2016). Perhaps the early evolutionary benefits of norm psychology were enhanced by the ability to provide verbal instruction (Pain; also see Birch, 2021) or to publicly reason about norms—about their scope; content; justification; whether a particular action is in compliance or is excusable, and if so, what types of excuses are acceptable, when, in what form; or when punishment is appropriate, what form it should take, who can apply it—so much so that human's increased reliance on norms drove the evolution of natural language, and the two became deeply intertwined (Lamm, 2014; Mercier & Sperber, 2017). Further exploration of these possibilities may provide a fresh perspective on questions about the innateness of both norm psychology and linguistic capacities.

Many of the commentaries noted that as useful as the distinction between implicit and explicit cognition is, it is a heuristic at best, and the two terms may be best understood as picking out poles on a spectrum. A large and fascinating middle ground remains to be explored. Indeed, what happens in that middle ground is likely to be intriguingly dynamic with respect to what is implicit and what is explicit. Many comments pointed to ways that this dynamism might be manifesting in normative cognition and behavior, all of which are worthy of more attention (Birch, Brody, and Crockett; Germar and Mojzisch; Schmidt et al.; Taumoepeau). One idea not explicitly articulated here but that the discussion brought to my mind is an analogy to what Godfrey Smith (2016, pp. 92–95) in another context (he was considering the evolution of subjective experience) called a "transformative view." The general idea is that simple forms of a psychological property or capacity are transformed and complexified by the later addition of more psychological sophistication. Here such a view would hold that simple forms of normative cognition, or more primitive forms of the core psychological processes that comprise norm psychology, were reshaped by their interaction with some additional, more complex piece of psychological machinery—but those core processes were not brought into being by the addition. A proponent could make the case that the core features of norm psychology preceded late arising things, such as natural language; status, hierarchy, and social identity; mind reading and mentalizing, and so on, but were transfigured by them when they came along. This may make the boundary problem more difficult to solve, but keeping the possibility in mind may help make progress on it. The proponent could also develop the view to shed light on proximate functioning. Perhaps the typical operational dynamics that determine which norm-psychological processes are more implicit and which are more explicit in any given episode of normative behavior have been shaped by these kinds of transformative phase shifts.

In Place of a Conclusion: Future Directions

What might the core features of normative cognition be, then? The two best candidates are conformity and enforcement. The relationship between conformity and social normativity remains perplexing. For example, a simple form of conformity is associated with imitation. The human capacity for imitation, which allows people to easily bring their own behaviors into conformity with those they observe, is implicated in much of this discussion. Perhaps a more complex form is picked out by the term "compliance," which appears to imply not just behavioral mimicry but also conformity to a rule (or something like it), perhaps of the sort required to participate in normative regularities. It remains unclear whether something extra is required, psychologically, for a behavior to count as an instance of compliance rather than mere imitative conformity. If so, what might that something extra be, and how could experiments test for its presence in children and other animals? Comparatively, it would be especially interesting to know whether the ability to produce normative regularities could evolve even in the *absence* of imitation. The possibility that there is an evolutionary route to social normativity that need not go through imitation is suggested by Powell's (2023) argument that normative regularities, although absent in most other species, including people's closest relatives and other species capable of cultural learning, are in fact found in comparatively cognitively simple social insects, such as ants.

Researchers exploring the relationship between social normativity and conformity will also do a great service by helping the field better understand the relationship between social normativity and normalcy—and just normalcy and normativity in general. Conceptually, it is not difficult to distinguish between something—the length of a spear, a style of clothing, a type of behavior—being *common* or *uncommon*, strictly statistically normal or rare, on the one hand, and something being *right* or *wrong*, as it should be or should not be according to some normative standard, on the other. The categories obviously cross-cut: Prescriptively correct behaviors may be rare; rule violations may be common. Psychologically, however, normal and normative are intimately intertwined, and the roots of the interconnection may run deep. In the literature on

norms, this manifests in many places. An influential distinction between descriptive norms and injunctive norms verbally marks the conceptual difference (e.g., Schultz et al., 2007), but a wealth of evidence suggests that human minds do not easily keep track of it. Rather, people intuitively slide from normal to normative: When people perceive a type of behavior as common, they go on to conceive of and treat it as if it were *correct* (Hoehl et al., 2019; Kenward, 2012; Kenward et al., 2011; Keupp et al., 2013, 2015; Lindström et al. 2017; Roberts et al., 2017, 2018; Schmidt et al., 2016; Tworek & Cimpian, 2016). This conflation of common with correct, however, does not seem to be unique to social normativity or norm psychology. Knobe presented evidence that it occurs in categorization in several different domains; something similar also appears in the use of indexicals (Lemeire, 2023), and it may be the manifestation of more fundamental operational principles of minds more generally (Theriault; also see Clark, 2016). Perhaps the roots of social normativity and the capacity for participating in normative regularities that are found in humans have their origins here, origins that are then transformed and complexified as more cognitive sophistication is layered on top of them by evolution. And perhaps a better understanding of the relationship between the psychology of normalcy and social normativity can give a better purchase on how to determine whether that layering is done by cultural or biological evolution.

Finally, research into the character and provenance of norm psychology will be enhanced by turning up the focus on the nuances of, broadly speaking, enforcement. The attentional, motivational, and behavioral aspects of enforcement are discussed under many names ("punishment," "policing," "sanctioning," "social maintenance"). Continued progress will be made by sorting out the most important distinctions marked by those terms and settling on a common vocabulary for talking about them across disciplinary lines. In addition, many dimensions and rich distinctions have been mapped out in adjacent literatures by game theorists, computer modelers, behavioral economists, ethnographers, and anthropologists—about second-party punishment, third-party punishment, third-party costly punishment, altruistic punishment. Experimental and theoretical norm psychologists build on and continue to explore the significance of these distinctions. Indeed, the psychology dedicated specifically to enforcement may be one of those subcomponents of adult human norm psychology that has been most transformed by the addition of further cognitive sophistication. The origins of enforcement may lay in simple negative responses to novelty, which are then ramped up, directed, or otherwise transfigured by the dynamics generated when they interact with processes layered on top of them. There is already ample reason to think that in human communities, punishment practices are themselves pervasively structured by norms and reputation (Boyd, 2017; Boyd & Mathew, 2015; Mathew et al., 2013). They may lie further back still (Flack et al., 2006).

Dyed-in-the-wool defenders of a nativist view of norm psychology would, I suspect, look to these types of considerations to mount a defense against Heyes's critique. They might marshal evidence of a propensity for enforcement that cannot be accounted for as a cognitive gadget. Nativists might argue that children's learning environments are marked by poverty rather than wealth when it comes to examples of enforcement. But researchers have to wait until that case has been made to assess it. The discussion here gives plenty to think about already.

Transparency

Action Editor: Daniel Kelly
Editor: Interim Editorial Panel
Declaration of Conflicting Interests
The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

ORCID iD

Daniel Kelly https://orcid.org/0000-0003-2235-1963

References

Bicchieri, C. (2016). *Norms in the wild*. Oxford University Press.

Birch, J. (2021). Toolmaking and the evolution of normative cognition. *Biology & Philosophy*, *36*(1), Article 4. https://doi.org/10.1007/s10539-020-09777-9

Boyd, R. (2017). *A different kind of animal: How culture transformed our species*. Princeton University Press.

Boyd, R., & Mathew, S. (2015). Third-party monitoring and sanctions aid the evolution of language. *Evolution and Human Behavior*, *36*(6), 475–479.

Clark, A. (2016). *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford University Press.

Cummins, R. (2000). 'How does it work?' vs. 'What are the laws?' Two conceptions of psychological explanation. In F. Keil & R. Wilson (Eds.), *Explanation and cognition* (pp. 117–145). The MIT Press.

Dennett, D. C. (1991). Real patterns. *The Journal of Philosophy*, *88*(1), 27–51.

Flack, J. C., Girvan, M., Waal, F. B., & Krakauer, D. C. (2006). Policing stabilizes construction of social niches in primates. *Nature*, *439*, 426–429.

Godfrey-Smith, P. (2016). *Other minds: The octopus, the sea, and the deep origins of consciousness*. Farrar & Giroux.

Griffiths, P. E. (1993). Functional analysis and proper functions. *The British Journal for the Philosophy of Science*, *44*(3), 409–422.

Henrich, J. (2015). *The secret of our success: How culture is driving human evolution, domesticating our species, and making us smarter*. Princeton University Press.

Heyes, C. (2018). *Cognitive gadgets: The cultural evolution of thinking*. Harvard University Press.

Hoehl, S., Keupp, S., Schleihauf, H., Mcguigan, N., Buttelmann, D., & Whiten, A. (2019). 'Over-imitation': A review and appraisal of a decade of research. *Developmental Review*, *51*, 90–108.

Kelly, D. (2020, June). Internalized norms and intrinsic motivation: Are normative motivations psychologically primitive. *Emotion Review*, 36–45.

Kelly, D. (2022). Two ways to adopt a norm: On the (moral?) psychology of internalization and avowal. In M. Vargas & J. Doris (Eds.), *The Oxford handbook of moral psychology* (pp. 285–309). Oxford University Press.

Kelly, D., & Davis, T. (2018). Social norms and human normative psychology. *Social Philosophy & Policy*, *35*(1), 54–76.

Kelly, D., & De Block, A. (2022). Culture and cognitive science. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of philosophy*. https://plato.stanford.edu/archives/sum2022/entries/culture-cogsci/

Kelly, D., & Setman, S. (2020). The psychology of normative cognition. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of philosophy*. https://plato.stanford.edu/archives/fall2020/entries/psychology-normative-cognition/

Kenward, B. (2012). Over-imitating preschoolers believe unnecessary actions are normative and enforce their performance by a third party. *Journal of Experimental Child Psychology*, *112*, 195–207.

Kenward, B., Karlsson, M., & Persson, J. (2011). Over-imitation is better explained by norm learning than by distorted causal learning. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1709), 1239–1246.

Keupp, S., Behne, T., & Rakoczy, H. (2013). Why do children overimitate? Normativity is crucial. *Journal of Experimental Child Psychology*, *116*(2), 392–406.

Keupp, S., Behne, T., Zachow, J., Kasbohm, A., & Rakoczy, H. (2015). Over-imitation is not automatic: Context sensitivity in children's overimitation and action interpretation of causally irrelevant actions. *Journal of Experimental Child Psychology*, *130*, 163–175.

Lamm, E. (2014). Forever united: The coevolution of language and normativity. In D. Dor, C. Knight, & J. Lewis (Eds.), *The social origins of language: Early society, communication and polymodality* (pp. 267–283). Oxford University Press.

Lemeire, O. (2023). "Philosophers care about the truth": Descriptive/normative generics. *Mind & Language*, *38*, 772–786.

Lindström, B., Jangard, S., Selbing, I., & Olsson, A. (2017). The role of a "common is moral" heuristic in the stability and change of moral norms. *Journal of Experimental Psychology: General*, *147*(2), 228–242.

Mathew, S., & Boyd, R. (2011). Punishment sustains large-scale cooperation in prestate warfare. *Proceedings of the National Academy of Sciences, USA*, *108*, 11375–11380.

Mathew, S., Boyd, R., & van Veelen, M. (2013). Human cooperation among kin and close associates may require enforcement of norms by third parties. In P. J. Richerson & M. Christiansen (Eds.), *Strüngmann Forum Report 12: Cultural evolution*. MIT Press. https://doi.org/10.7551/mitpress/9894.003.0006

Mercier, H., & Sperber, D. (2017). *The enigma of reason*. Harvard University Press.

Mikhail, J. (2011). *Elements of moral cognition: Rawls' linguistic analogy and the cognitive science of moral and legal judgment*. Cambridge University Press.

Millikan, R. W. (1989). In defense of proper functions. *Philosophy of Science*, *56*, 288–302.

Moya, C., & Henrich, J. (2016). Culture–gene coevolutionary psychology: Cultural learning, language, and ethnic psychology. *Current Opinion in Psychology*, *8*, 112–118.

Neander, K. (1991). Functions as selected effects: The conceptual analyst's defense. *Philosophy of Science*, *58*, 168–184.

Powell, R. (2023). Social norms and superorganisms. *Biology & Philosophy*, *38*, Article 21. https://doi.org/10.1007/s10539-023-09909-x

Roberts, S. O., Gelman, S. A., & Ho, A. K. (2017). So it is, so it shall be: Group regularities license children's prescriptive judgments. *Cognitive Science*, *41*(Suppl. 3), 576–600.

Roberts, S. O., Guo, C., Ho, A. K., & Gelman, S. A. (2018). Children's descriptive-to-prescriptive tendency replicates (and varies) cross-culturally: Evidence from China. *Journal of Experimental Child Psychology*, *165*, 148–160.

Schmidt, M., Butler, L., Heinz, J., & Tomasello, M. (2016). Young children see a single action and infer a social norm: Promiscuous normativity in 3-year-olds. *Psychological Science*, *27*(10), 1360–1370. https://doi.org/10.1177/0956797616661182

Schultz, P. W., Nolan, J., Cialdini, R., Goldstein, N., & Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological Science*, *18*, 429–434.

Sinnott-Armstrong, W., & Wheatley, T. (2012). The disunity of morality and why it matters to philosophy. *The Monist*, *95*(3), 355–377.

Sinnott-Armstrong, W., & Wheatley, T. (2014). Are moral judgments unified? *Philosophical Psychology*, *27*, 451–474. https://doi.org/10.1080/09515089.2012.736075

Smaldino, P. (2019). Social identity and cooperation in cultural evolution. *Behavioural Processes*, *161*, 108–116.

Sripada, C. & Stich, S. (2007). A framework for the psychology of norms. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The innate mind: Culture and cognition* (pp.280–301). Oxford University Press.

Stich, S. (2018). The quest for the boundaries of morality. In K. Jones, M. Timmons, & A. Zimmerman (Eds.), *The Routledge handbook of moral epistemology* (pp. 15–37). Routledge.

Tinbergen, N. (1963). On aims and methods of ethology. *Zeitschrift für Tierpsychologie*, *20*, 410–433.

Tworek, C. M., & Cimpian, A. (2016). Why do people tend to infer "ought" from "is"? *The Role of Biases in Explanation: Psychological Science*, *27*(8), 1109–1122.

Westra, E., & Andrews, K. (2022). A pluralistic framework for the psychology of norms. *Biology & Philosophy*, *37*, Article 40. https://doi.org/10.1007/s10539-022-09871-0

Young, L., & Dungan, J. (2012). Where in the brain is morality? Everywhere and maybe nowhere. *Social Neuroscience*, *7*(1), 1–10. https://doi.org/10.1080/17470919.2011.569146