Biology & Philosophy (2019) 34:39
https://doi.org/10.1007/s10539-019-9693-z

Two of a kind: Are norms of honor a species of morality?

Toby Handfield[1] · **John Thrasher**[2]

Received: 6 November 2018 / Accepted: 12 June 2019 / Published online: 14 June 2019

Abstract
Should the norms of honor cultures be classified as a variety of morality? In this paper, we address this question by considering various empirical bases on which norms can be taxonomically organised. This question is of interest both as an exercise in philosophy of social science, and for its potential implications in meta-ethical debates. Using recent data from anthropology and evolutionary game theory, we argue that the most productive classification emphasizes the strategic role that moral norms play in generating assurance and stabilizing cooperation. Because honor norms have a similar functional role, this account entails honor norms are indeed a variety of moral norm. We also propose an explanation of why honor norms occur in a relatively unified, phenotypically distinctive cluster, thereby explaining why it is tempting to regard them as taxonomically distinct.

Keywords Honor · Social norms · Morality · Signaling · Scientific categories · Cooperation

Sometimes honor demands virtue. An "honor code" requires individuals to abstain from opportunities to cheat, in line with widely shared moral ideas. In many cultures, to be honorable requires generous hospitality, abstinence from theft, and a firm commitment to honesty: all typical moral virtues. But often, norms of honor and norms of morality appear to be in conflict. The early modern European duel encouraged men to engage in lethal violence, sometimes over trivial or contrived slights (Allen and Reed 2006). In cultures that practice "honor killings", victims of rape are sometimes murdered in order to repair family reputation (Chesler 2010). In cultures that practice blood feuds, the retaliatory cycle can lead to dozens of deaths over a period of decades (Boehm 1986). Not only are these seemingly immoral acts

✉ Toby Handfield
toby.handfield@monash.edu

John Thrasher
Johnthrasher23@gmail.com

[1] Philosophy Department, Monash University, Clayton, Australia

[2] The Smith Institute, Chapman University, Orange, USA

occasionally tolerated, they are positively *required* by the social norms of the honor society. Those who do not follow the relevant norms of honor suffer consequent social disadvantage, ranging from outright punishment to shunning and ostracism.

We look at characteristic features of honor norms across different societies and different times to determine whether there is a principled way of distinguishing these norms from morality. This question is of interest because of the general methodological issue it raises for social science: What constitutes evidence that a set of norms should be considered a distinct cluster or kind?

We argue attempts to demarcate honor and morality based on their psychological features, such as their first-order content or their motivational profile, are unlikely to succeed. Paradigmatic examples of honor psychology and moral psychology are each too heterogeneous to permit very strong generalisations, and they have a good deal of overlap. We further argue that honor and morality have some important functional similarities, which suggests they are better regarded as alternative solutions to a common set of core social problems. Specifically, both honor and moral norms contribute to sustaining cooperation by overcoming *assurance problems* and stabilizing cooperation. Assurance problems arise in strategic interactions where the more beneficial behaviour (individually, collectively, or both) makes the actor vulnerable to the choice of the other party. Different environments—both physical and social—will emphasise different types of assurance problem, generating demand for different mechanisms to solve those problems. We argue this explains why honor cultures tend to be associated with particular sorts of ecological and political environments.

We interpret these findings in light of recent attempts by evolutionary psychologists to map the domain of morality and argue that these give some support for the idea that honor is indeed a species of morality, though the same evidence suggests morality should be construed very broadly. Specifically, the descriptive extension of "morality" covers a set of norms that are considerably more diverse than most accounts of moral normativity are willing to allow. In the conclusion, we discuss some implications of this view. In particular, our conclusion suggests that if we also accept moral realism, the moral landscape will be characterized by a plurality of conflicting norms. Tragic conflict is therefore likely to be an inevitable feature of the moral realm. A second implication relates to how we should go about studying moral and social norms from a naturalistic point of view. Categories derived from traditional moral theory and metaethics are, we argue, ill-suited to the scientific study of human norms.

What are norms of honor?

Members of honor cultures are, by and large, highly concerned to defend a certain sort of social status: that of being "honorable".[1] The exact meaning of honorable status varies across cultures, but themes related to honesty, bravery, and female

[1] Note that we use the terms "norms of honor", "honor codes", and "honor norms" interchangeably. Similarly, we make no distinction between "morality" and "moral norms" for present purposes.

chastity recur frequently. This social status, and the concern to preserve it, undergird a set of social practices and norms which help to constitute the meaning of honor, while also prescribing the appropriate actions to defend, maintain, or threaten someone's honor. So understood, probably all cultures have some concern for honor. For present purposes, we wish to isolate particularly extreme examples of devotion to honor. Accordingly, we define a *culture of honor* to be one where widely accepted social norms sometimes demand *violent behaviour* to defend honorable status.[2] Feuding Balkan herders (Boehm 1986), Californian prison gangs (Skarbek 2014), duelling European aristocrats (Allen and Reed 2006), South Asian practitioners of honor killing (Chesler and Bloom 2012), and the Scots-Irish settlers of the American south (Nisbett and Cohen 1996; Grosjean 2014) are all members of honor cultures in this sense.

Within honor cultures, we see two main types of norm that require violence: norms of revenge and norms of purification. *Revenge norms* require individuals to avenge slights, insults and other wrongdoing against one's self or extended family by harming outgroup members. *Norms of purification* prescribe expulsion or violence towards in-group members who bring dishonor to one's group. In both cases, failure to comply with these norms will lead to feelings of shame for the transgressor, attitudes of contempt from others, and consequent social disadvantage for the transgressor.

Having characterized some of the central features common to honor cultures, it is also important to stress their heterogeneity. Some cultures of honor revolve primarily around revenge norms, others around purity norms, and yet others have both.[3] Cultures of honor are sometimes informal, such as the Pashtunwali (Ginsburg 2011) or the culture of Montenegrin herders (Boehm 1986), but sometimes they are formalized, as in codes of duelling (Allen and Reed 2006; LaVaque-Manty 2006) and *bushido*, the samurai warrior code (French 2004, Chapter 8). We will use the term "honor code"—for convenient contrast with the notion of a moral code—to refer to the norms characteristic of an honor culture, in an attempt to generalize over this significant heterogeneity.

Honorable status is partly constituted by compliance with the local norms of purity and revenge. It is also characterised in part by the distinctive emotions and reactions to the loss of that status. Losing honor is stereotypically associated with shame, rather than with guilt, fear, anger, sadness, or other negative emotions. That

[2] Leung and Cohen (2011) distinguish three types of culture: honor, dignity, and *face*. The culture of face is, like an honor culture, highly concerned about socially onferred status, and many socially important behaviors are motivated primarily by emotions of pride and shame, rather than by conscience and guilt. Here we effectively treat the concerns of face cultures as a species of honor norms, and do not attempt the analytic task of distinguishing the primary norms of the two. That said, Leung and Cohen suggest a key difference between honor and face cultures is that the social hierarchy is steeper, more pronounced, and relatively stable in a face culture, whereas honor cultures are relatively egalitarian. When we restrict our focus to cultures that defend honor with *violence*, we appear to have converged on a very similar understanding as Leung and Cohen, because the use of violence is likely a symptom of an unstable status hierarchy.

[3] For a wide-ranging discussion of the variety of different honor cultures and codes, see Tamler Sommers *Why Honor Matters* (2018).

said, shame is not uniquely associated with transgressions of honor norms: cross-cultural evidence suggests shame is a robust reaction to local disapproval (Sznycer et al. 2018). But it is striking that honor is regained by a defiant act—an act of revenge or of purification—rather than by apology, reparation, or otherwise conciliatory behaviour. Indeed, honor cultures are sometimes characterised by uniting positive and negative reciprocity: avenging wrongs done is seen as evidence one will return favours also, and vice versa (Leung and Cohen 2011). Avenging an insult or slight is not just a way of demonstrating that one should not be predated against, but also a way of maintaining a reputation as a person who can be trusted to honor debts and promises. This shame–defiance pattern distinguishes honor's reactive footprint from alternative patterns of response such as guilt-then-apology or embarrassment/downplaying/concealment which might ensue from other types of normative transgression.

Defining morality

A definition of morality can be either normative or descriptive. A normative definition may make use of normative language in its formulation. So, for instance, if we define morality as that set of codes which rational individuals *should* regard as overriding, we have adopted a normative definition. If we instead define it as a code of behavior which (some subset of) individuals *do* regard as overriding, then we have a descriptive definition. For our purposes, we need a definition which will allow it to be, in principle, empirically tractable whether or not a given phenotype is "moral", and also allow it to be a non-trivial question whether or not honor norms are examples of morality: it should not be settled a priori by the definition. We therefore employ a descriptive definition.[4]

In this section, we consider three approaches to defining morality with these desiderata in mind. First, we consider the possibility that moral norms have a distinctive motivational force. Second, we consider the possibility that moral norms have a distinctive type of content. Finally, we consider the possibility that moral norms can be understood in terms of a social evolutionary framework, as a set of social technologies adapted to address a certain class of strategic problems. Only the third of these approaches is fruitful, we argue.

Terminologically, the following discussion is complicated by our intending the reader to have in mind certain candidate "moral" norms and "honor" norms so as to evaluate the implications of the candidate proposals, while those very proposals might—if adopted—significantly erode or even collapse the distinction between the two classes. We have tried to avoid an abundance of scare quotes to make the text more readable, but this does come at some cost of ambiguity.

[4] The relationship between "morality", descriptively defined, and "morality", normatively understood, takes us into fraught meta-ethical debates which we dare not enter at present. Suffice to say we think there is *some* relation between the two, and this is part of why we think the present project has interest for debates about moral realism. But for all we say here, it is possible the descriptive definition is entirely irrelevant to the normative understanding.

Can morality be defined by the motivational force of moral norms?

One allegedly distinctive feature of moral norms that has attracted considerable attention from philosophers is their motivational force. This broad idea is variously expressed by saying moral norms are overriding, categorical, or authoritative. Individuals who are moved by a moral commitment appear willing to forego substantial personal benefits, and even to incur painful and burdensome costs, in order (1) to fulfil their moral obligations, (2) to punish those who transgress moral obligations, and/or (3) to advocate their moral beliefs.

Relatedly, in social psychology, moral norms are distinguished from non-moral, conventional norms by their *unconditional* form (Nucci and Turiel 1978; Turiel 1983). In this vein, Cristina Bicchieri argues moral norms are "unconditional" (Bicchieri 2006).[5] According to Bicchieri, what makes norms conditional or not is how sensitive one's compliance and endorsement are to expectations that *others* will comply with and endorse the norm. If we behave in Rome as the Romans do, we are directed by conditional norms. If, however, when confronted with those who disagree with our norms, we respond with "here I stand, I can do no other," we are following an unconditional norm.[6]

This suggests the hypothesis that honor norms are conditional while moral norms are unconditional. If this hypothesis were true, it would enable us to make sense of the apparent conflict between honor and morality, and why this conflict can persist in a society. The norms can have different first order content: they require incompatible behaviours, but they have different psychological and sociological foundations: the psychology of conditional versus unconditional norms.

This hypothesis has several shortcomings. First, there may be no straightforward way to classify an entire community's norm as conditional or unconditional, as this hypothesis presupposes. For instance, in Indian society, Hindu children are trained from a very young age not to use one's left hand for eating. For many, this norm will be adopted conditionally. A Hindu migrant may lose any compunction about using his or her left hand in eating once they move to Hong Kong, for instance. However, we can also imagine this individual still refusing to use his or her left hand while eating, even though the context has shifted and there is no danger of anyone socially sanctioning this person for non-compliance. In this case the norm

[5] See also Brennan et al. (2013) and O'Neill (2018, Sect. 3). Southwood (2011) uses the terms "practice-dependent" and "practice-independent" in a somewhat more subtle way—distinguishing them from conditionality/unconditionality by referring to what *grounds* the normative judgment that such-and-such should be done. According to Southwood, a normative judgment is practice-dependent just in case it *appears* to the judging agent that a social practice plays a non-derivative role in justifying acting in accordance with a corresponding principle (p. 778). While this strategy has some evident attractions for distinguishing the moral from the conventional in comparison with alternative proposals, for our purposes we can treat it along with other proposals relating to the motivational force of norms. We also register some concern that this criterion is both highly individualistic and introspective: it turns on how the justification of a norm *appears* to a given *individual*. We suspect for purposes of explaining social phenomena, a concept like this unlikely to be particularly fruitful on this point, see Gaus (2014).

[6] Bicchieri (2006) discusses two categories of non-conditional norms: personal norms and moral norms. Personal norms, for instance how one takes their tea, do not concern us here.

appears unconditional. Given this individual level heterogeneity, it is hard to identify a non-arbitrary way of classifying an entire community's norm as conditional or unconditional.

Second, even if we grant that norms can be suitably categorised, there is limited evidence for the unconditional character of stereotypically moral norms.[7] The classic distinction between the moral and conventional has become increasingly unstable in the face of new experimental evidence (Kelly et al. 2007; Fraser 2012; Quintelier et al. 2012; Quintelier and Fessler 2015). The situationist attack on stable character traits undermines the notion that moral norms are viewed as unconditional (Harman 1999; Doris 2005). Social psychology (Milgram 1974) and experimental economics (McCabe et al. 2003; Smith 2008; Abbink et al. 2017) have also shown in numerous ways that moral norms are followed and dropped for surprising reasons in different contexts.

Third, there is evidence that honor norms are sometimes treated as unconditional: individuals are motivated to act on perceived obligations of honor, at odds with all other considerations. Honor killings in western democratic nations provide a stark example: a father who kills his "too westernized" daughter in suburban Arizona, for instance, does not earn any admiration from neighbours, does not improve his family's social standing, and will almost certainly be incarcerated for a lengthy period (Gaynor 2011). Appiah (2011) raises the case of the Duke of Wellington who engaged in a duel, despite clearly understanding it was a substantial risk to his life (given his poor marksmanship), it jeopardized the political causes that he valued, and it was contrary to the teaching of his church—of which he was a devout member. Wellington seems to have acted as if the norms of honor to which he subscribed were unconditional. In literature, characters like John Proctor in *The Crucible* and the titular character of *Lord Jim* are willing to die rather than violate norms of honor. Proctor in particular illustrates the apparent power of honor to outrank a competing demand that might be thought "moral": he is willing to tell a blatant lie, guaranteeing his damnation (given circumstances in which it will save his life and that of others), but cannot bring himself to *publish* the lie, and thus dishonor his name.

Beyond these anecdotal examples, further evidence that honor norms promote unconditional compliance comes from an experiment in which members of honor cultures were primed so as to make salient normative ideals of retaliation and revenge (Leung and Cohen 2011). After priming, those who *approved* of retaliatory violence were more likely to behave honestly in a task where there was an opportunity to cheat for material gain. In contrast, members of honor cultures who disapproved of retaliatory violence (and hence were reacting against their culture's primary normative logic) were more likely to cheat after exposure to the same prime. Given there was no evident opportunity for honest behaviour in the first group to be socially recognised, it is difficult to explain the marginal effect except in terms of an unconditional norm.

It is possible, however, that even though honor norms are viewed as unconditional, they only generate unconditional demands for members of the honor society, i.e., the norms are not used to evaluate the conduct of outsiders, and so are not

[7] See O'Neill (2018) for a recent review.

universal. This suggests another possible distinction: moral norms may be universal, while honor norms are not.

Many moral theories, however, eschew universalism. Explicitly relativist theories of morality are non-universal (Harman 1975, 2015), but so are plausible versions of constructivism (Street 2010) and contractarianism (Gauthier 1986; Moehler 2018), as well as conventionalist theories (Vanderschraaf 2018). These are all normative theories of morality of course, but we suggest the case for abandoning universalism becomes all the stronger if one adopts a descriptive approach, as we do. (Though see "The functions of honor norms" section where we discuss a reason why this distinction may prove to be symptomatic of a more fundamental underlying difference between honor and morality.)

None of the foregoing arguments are decisive, but given how thin the evidence is for each of the proposals surveyed, it appears unmotivated to place a great deal of weight on allegedly distinctive motivational features of different types of norms. It remains an important and interesting open question: what is the psychology of an unconditional norm for? Why did humans evolve the ability to internalise a norm and treat it as an objective demand?[8] If this question can be answered, it might guide us to better understand why some types of norm are more or less likely to be held unconditionally. But without such guidance, and the very mixed empirical evidence, there appears little basis for distinguishing our paradigms of honor norms and moral norms in these terms.

Do moral norms have distinctive first-order content?

Rejecting the hypothesis that the norms of morality and honor have fundamentally different modes of motivating behavior, we might instead think the *contents* of honor and moral norms systematically conflict. Typical moral norms might be: "Avoid harming innocents", "Apologize and be remorseful for your own transgressions", "Don't take unfair advantage of others", and so on.[9] These are obviously rather different from norms of revenge and purification. Second, honor codes tend to require violent or conflictual behaviour, often as a *response* to conflict or transgression (Sommers 2009) whereas it might (perhaps naively) be thought that moral codes typically involve a number of requirements to *prevent* or diminish conflict.[10]

[8] On this point see: Joyce (2006), Gavrilets and Richerson (2017) and Stanford (2018).

[9] Of course, there are norms for which there seems a good case to categorise them either as a norm of honor or as a moral norm. Keeping promises for instance, may sound like a moral norm, but what about swearing an oath, which is usually construed as a matter of honor? Any attempt to distinguish honor and morality on the basis of content would need to adjudicate such cases, but because we think this approach ultimately unsatisfactory, we will leave the matter unresolved. See Scanlon (1998, 323–326) for an attempt to make good the distinction between promises, which are entered into freely, and oaths made on the basis of honor. Although both are related to assurance in some way, Scanlon argues that the first create specifically moral obligations, while the second are related to aretaic values, but not distinctly moral obligations.

[10] There is good reason to think violence is frequently "moralistic" in the sense that it is committed to achieve what is seen as a moral goal (Black 1983). Fiske and Rai (2014) argue moral motivation is actually essential to the vast majority of violent behaviour. Nonetheless, we suppress these concerns at present for the sake of testing the hypothesis that morality and honor can be distinguished along these lines.

This touches on a third possible point of difference: moral motivation is frequently theorized as a variety of motivation that is *not solely* directed towards personal benefit. Someone whose motive is solely self-interested, either through pursuit of future reciprocal benefit, or through enhanced reputation, is regarded as a poor moral exemplar.[11] But a person motivated by honor can—at least in many cases—be transparently engaged in the cultivation of an enhanced social image without the loss of honor. A duellist, for instance, need have no compunction about engaging in the duel for the purposes of preserving his social status, i.e. for purely personal benefit. In contrast, if someone confessed that they acted morally only out of an expectation of praise or esteem, this would undermine our moral assessment of them.

To distinguish honor and morality on the basis that moral motivation can *never* be solely self-interested, while honourable motivation *can* be, makes for an empirically feeble dichotomy. One type is defined by a universal negative property, while the other is a mere compatibility with the corresponding positive property. Hence in any given case where self-interest is absent, it remains unresolved which kind is being dealt with. More fruitful proposals might suggest that honour-compliance is always motivated by self-interest, or that moral behaviour is always motivated by a positive concern for the collective good.

None of these proposals fit comfortably with examples of honor codes motivating honest behaviour, the forswearing of personal gain, and willingness to accept significant costs, as discussed above. More fundamentally, however, in all these cases, we cannot identify a methodological principle that explains why any differences of this sort are significant enough to warrant regarding honor and morality as constituting distinct kinds. For any set of norms, it will be possible to gerrymander two sets that differ in content. Insisting on these sorts of differences as definitive of the kinds conflicts with our requirement not to beg the question against those who think it at least an open possibility that honor is a species of morality. What is required is a substantive reason for why a given distinction in content warrants treating the classes as distinct.

In recent ethnographic work, Purzycki et al. (2018) have attempted to give a detailed, empirical account of morality, by comparing the moral "models" of eight diverse cultures using a very simple open-ended question: "What makes a good/bad person?" This methodology is motivated by a number of concerns, such as the absence of "morality" and "moral" in some languages, whereas "good/bad" appear to be cultural universals. They find generosity and honesty are the most salient "good"-making features, while deceit, theft, and violence were the most salient "bad"-making features. This leads to their conclusion that "[c]ross-culturally, the most salient components of individuals' mental models of morality revolve around

[11] There are any number of examples of this in the history of moral philosophy. Kant (1785) argues in the first lines of the *Groundwork of the Metaphysics of Morals* that nothing can be called "good" without qualification except the "good will," which is aimed at the moral law, not personal benefit. More generally, Baier (1954, 104–105) argues that to act from the "moral point of view" is to act on the basis of principle or distinctively moral reasons. Whatever "moral" principles and reasons are, they are not merely prudential. Even if prudence is the ultimate foundation and justification of morality (e.g., Gauthier 1986; Moehler 2018), the moral norms are not identical with prudential norms.

the provisioning of material resources in the form of generosity, helpfulness, and theft" (Purzycki et al. 2018, 8, 4.2.1).

Although these findings are of significant intrinsic interest, they are not sufficient to deliver a categorisation of norms that adjudicates our question. First, these data are not enough to draw any sharp boundary around the moral domain as a whole, since the sample of just eight cultures is still only a very small portion of global variation. Second, since the method used here does not distinguish between the specifically "moral" and the "good" or "appropriate", the results are compatible with thinking either that honor and moral norms are part of a larger moral domain or that they are distinct. Nevertheless, it is noteworthy that some of the most salient traits appear to track concerns that are associated with honor in many cultures: e.g. honesty, deceitfulness, hospitality, respectfulness, and arrogance.

A second cross-cultural study (Buchtel et al. 2015) asks what behaviours are *immoral*, and finds there is significant variation between Chinese and Western subjects. In particular, "mainland Chinese were more likely to describe uncivilized behaviours as 'immoral' compared with harmful behaviours, whereas Westerners did the opposite" (Buchtel et al. 2015, 1389). For the Chinese subjects, incivility rather than harm seems to be the basis of judging many behaviours to be morally wrong. Aside from further challenging the strict moral/convention distinction, these results are striking in that they find the "Chinese lay concept of 'immorality'… is more applicable to spitting on the street than killing people" (Buchtel et al. 2015, 1392). This suggests significant diversity must be tolerated in any account of the moral domain. And again, several of the behaviours identified as "immoral" by Chinese subjects were potentially examples of honor-norm violations: being unfilial, selling out/betraying others, adultery, back-stabbing, and lying.

Reflecting on studies like these reinforces that raw ethnographic data are inevitably doomed to underdetermine the sort of theoretical distinction we are seeking. While we may wish for a more empirically grounded conception of morality, theoretical commitments cannot be avoided outright. We suggest an evolutionary framework provides the necessary overarching constraints for any proposals to identify a class of norms as worthy of distinct study. Gene-culture co-evolution by natural selection and social transmission provides a well-confirmed theoretical basis for understanding the social world which unifies the social and physical sciences, a basis which has had increasing success at predicting and explaining a range of cultural phenomena such as culinary practices, religious doctrine, and ethnic markers (e.g. Boyd and Richerson 1988; Henrich and Henrich 2007; Henrich 2009; Bowles and Gintis 2011; Gintis and Fehr 2012).

An evolutionary framework for categorising norms

Drawing on the failure to find robust conceptual analyses of morality, on cross-cultural variance in "moral" intuitions, and on the failure of various attempts to identify the moral domain either with certain distinctive content or distinctive force, Stich (2017) argues there is no moral domain: the kind is too heterogeneous to be the

subject of well-constituted scientific debate. While sharing Stich's pessimism about approaches based on psychological content, we believe there is significant potential for a functional account of norms more generally, based on their role in the evolution of human sociality. Equipped with such a framework, we will be better able to assess the differences and similarities between different classes of norms, such as the stereotypically moral and stereotypically honor-based.

Given many norm-governed behaviours appear to serve a group-beneficial function at a cost of individual fitness, they *prima facie* require special explanation in an evolutionary framework. The following are the best-established elements of the current orthodoxy. First, many "moral" behaviours are altruistic in the sense that they sacrifice the *wellbeing* of the actor for the wellbeing of another, but need not be *fitness-sacrificing* in the sense of reducing the number of the actor's surviving offspring. That is: what is psychologically "sacrificial" or altruistic need not be biologically sacrificial (Kitcher 2011, 18–20). Second, the concept of *inclusive* fitness can be used to show how helping those in a group to which one is closely related can increase the reproductive success of one's genes, even if it involves a lower rate of reproduction by the individual (Hamilton 1964a, b). Third, many "moral" behaviours involve a mutually beneficial exchange of favours—these behaviours need involve no sacrifice of fitness or wellbeing, but merely require an environment in which there are mechanisms to maintain trust and to reduce associated risks (Trivers 1971).

These mechanisms that explain how various aspects of our norm-psychology can be compatible with Darwinian evolutionary processes require that behaviour be *motivated* in particular ways, that it be *targeted* to particular individuals, and that it be *regulated* by certain contextual features. Broadly speaking, these are all *strategic* considerations. From this perspective, morality can be fruitfully regarded as a sort of social technology to implement the necessary strategic checks and measures for altruism, reciprocity, and related behaviours to be fitness-enhancing (Gaus 2015). Norms relating to fairness are important in building trust and reducing the risk of defection in cases of reciprocal exchange (McCabe et al. 2003; Henrich 2004; Nowak and Sigmund 2005). Norms relating to disgust and purity may be important in building group identity and directing altruistic concern towards associates who are more closely related (Haidt et al. 1993; Nichols 2002; Kelly 2011). Norms relating to harming other in-group members are important in reducing conflict and injury.[12]

Norms are not the only important aspect of morality as a social technology. *Emotions* such as guilt and anger are important in proximate motivation of post-transgression behaviours which are necessary to rehabilitate the transgressor to a trusted status or to minimize the danger they will pose in future. Although we doubt morality always and uniquely involves *unconditional* norms, to the extent that this

[12] Obviously, norms can be stupid, inefficient, and fail to serve any useful purpose—our claims about the functions of these norms should not be taken to imply that we neglect that possibility. On "bad norms" see Brennan et al. (2013, Chapter 8), Abbink et al. (2017) and Thrasher (2018). But at the same time, many norms clearly do serve functions that may be of individual and social benefit.

phenomenon exists, it may serve to bolster the motivation to comply with cooperative norms even in cases of high temptation (Joyce 2001). The social *transmissibility* of normative attitudes is important in preserving a social order over generations, enabling multi-generational cooperation in long-term projects such as maintaining a fortress, building a religious monument, or cultivating land, and capable of effectively competing with other groups (Gintis 2003; Boyd and Richerson 1988; Bowles and Gintis 2011; Birch 2017; Rozin 1991).

Thinking in terms of morality as a social technology also invites the use of game theory to analyse the strategic circumstances that "morality" is best suited to address. It is no coincidence that the prisoner's dilemma is the most discussed game in academic literature: its central feature is the conflict between individual self-interest and collective benefit (Gauthier 1986). The rational pursuit of what is best for each individual leads to a worse outcome for all. Conversely, achieving a better outcome for all requires sacrifice on the part of one or more individuals. Norms appear to play a significant role in motivating cooperative behaviour in circumstances analogous to the prisoner's dilemma (Vanberg and Congleton 1992; Bendor and Swistak 1997; Bicchieri 2006). In addition to this, social structures that increase the likelihood of repeat interactions, the tracking of reputations, and the punishing of transgressors, transform the prisoner's dilemma into related games, where fitness promoting behaviour for individual and group are better aligned (Skyrms 2004; Sterelny and Fraser 2016, 45).

We propose to adopt the *Morality as Cooperation* (MAC) framework, developed by Curry et al. (2019a, b), as a basis for systematising the space of norms.[13] Morality as Cooperation posits seven domains of moral thinking:

1. Family values
2. Group loyalty
3. Reciprocity
4. Heroism
5. Deference
6. Fairness
7. Property rights.

[13] The principal competitor to MAC at time of writing is Moral Foundations Theory (MFT), developed by Graham et al. (2011, 2013). Compared to MFT, MAC has a clearer rationale for what it would take to identify a new domain of moral thinking. Although the general theory motivating MFT is evolutionary, in practice its proponents have relied heavily on factor analysis of survey responses to questions like "Is the following factor relevant to whether or not a given behaviour is moral" to identify moral domains, but such a process is highly sensitive to the inventory of questionnaire items used. MAC uses a similar methodology, but is more explicit and consistent in its evolutionary criterion guiding the development of the questionnaire items. More work needs to be done showing the links between the empirical data in MFT and the underlying theory, and there are questions about the replicability and robustness of MFT cross culturally (Curry et al. 2019a; Purzycki et al. 2018). For these reasons, MAC is a better basis for our inquiry, though it too is almost certainly not the final word on these matters. In recent work using US populations, Davis (unpublished manuscript) has found further evidence that folk use of the term "moral" is inconsistent with understanding morality in terms of distinctive content or force of norms, and is consistent with the more pluralistic domains of morality associated with either MAC or MFT.

The MAC domains correspond to lessons from evolutionary game theory. Each domain of moral thinking relates to a distinct configuration of strategic interests that poses challenges to cooperation. The sorts of solution called for in each case are sufficiently different that it is plausible to think different social and cognitive mechanisms have developed to address each one. For instance, the domain of "family values" are the values explained by mechanisms of kin selection; heroism and deference are associated with the two possible strategies in pure strategy equilibria of a hawk–dove interaction, whereas property rights correspond to a correlated equilibrium of the same encounter; and so on for each domain. In this way, there is a theoretical criterion for when we should be open to the possibility of a new domain: whether or not the strategic situation is different in the problems to which it gives rise.

We propose to use MAC as a framework for the question whether honor is part of morality. The key question, therefore, is not merely whether there is some sort of connotative association between the domains, as labelled in MAC, and notions of honor. Rather, we need to understand what social problems, if any, norms of honor are equipped to solve. If honor norms play a functional role, what is the strategic situation in which they arise? Can they be characterised as broadly "cooperative", and are they substantially different from the other types of cooperation that are relevant to morality?

The functions of honor norms

There is no consensus on the role of honor in strategic settings, but one common thought is that honor has a place in competitive settings, where agents vie for a resource that cannot be shared and the ensuing competition leads to costly losses for both parties. The classic game to model such interactions is known variously as Chicken, Hawk–Dove, or Snowdrift (Maynard Smith and Price 1973; Maynard Smith 1982).

It has been suggested that honor cultures dominated by norms of revenge arise in settings where important economic assets are easy to steal, but exogenous and formal enforcement is absent or weak, and thus the best mechanism to defend one's personal property is "self-help" (Black 1983; Nisbett and Cohen 1996; Brown and Osterman 2012). This sort of setting can be modelled by a sequential version of Hawk–Dove (Fig. 1). The standard Nash solution concept predicts that Player 1 will aggress, and it will then be rational for Player 2 to defer. But if Player 2 can convincingly commit to meet any future aggression with retaliation, then it becomes rational for Player 1 to defer.

It has been suggested that the emotional responses we have to transgressions—anger, outrage and the like—are commitment devices to enable us, in the position of Player 2, to credibly threaten that we will undertake irrationally costly revenge against an initial transgression (Frank 1988). That is, by being captive to his or her emotions, Player 2 cultivates a reputation as being the sort of agent who will always respond to aggression with similar aggression, even though this will deliver a lesser payoff in the short term. Convincing a potential aggressor of this disposition,

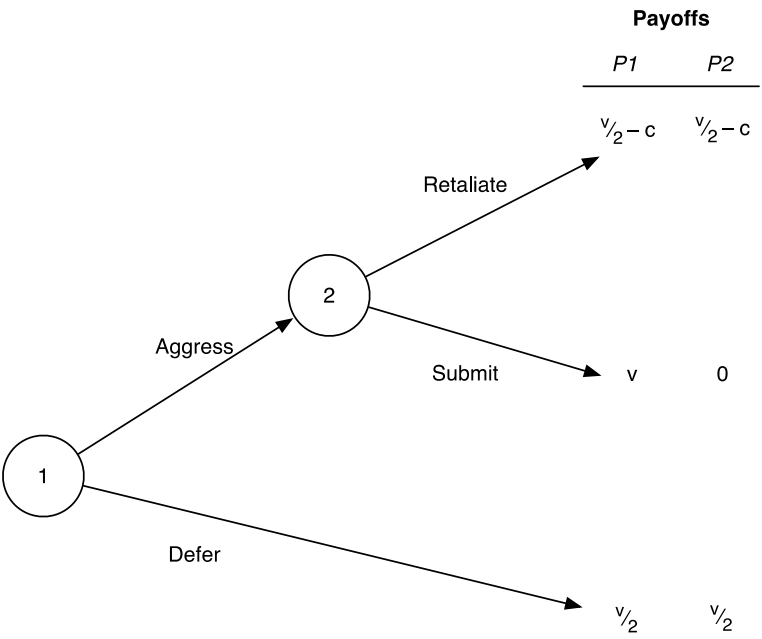

Fig. 1 Sequential hawk-dove game. Assuming $2c > v$, the only equilibrium is (Aggress, Submit)

however, should effectively deter aggressors, and will justify what may otherwise appear an irrational tendency to violence. Honor norms requiring revenge may play a similar role: they demand from individuals revenge against slights that appear to outsiders to be disproportionate and dangerous—but they may serve a useful function for the agents of revenge if they enable them to cultivate a sufficiently fearsome reputation that they thereby deter future attacks (Nisbett and Cohen 1996).

We conjecture that revenge norms are particularly important in settings where there are *collectives* that wish to protect their property. In the same way that an individual will deter aggression by developing a reputation for fearsome commitment to vengeance, a collective has reason to carry out individual acts of retaliation that are costly and seemingly irrational. But internally, such collectives may suffer from conflicts of interest and mixed motives that give rise to standard collective action problems. For instance, a family of shepherds may all benefit from carrying out an act of violent retaliation for having had one of their sheep stolen, but no particular shepherd wishes to be the one to carry out the dangerous act of violence. Each individual is tempted to free-ride on the behaviour of others, and the net result is likely to be a complete absence of retaliation. In such a setting, social norms within the collective, requiring defence of collective honor, may provide a crucial additional motivation that overcomes the temptation to free-riding.

This conjecture entails an intriguing feature of norms of revenge—these norms have different strategic functions at different levels of analysis. Between honor-bearing collectives (i.e. between potential targets and perpetrators of aggression), the norms play the role of ensuring families carry out irrational retaliation, serving a

deterrent role against other potential aggressions. Within collectives (among potential victims of future aggression whose fates are collectively tied to shared property), the norms play the role of ensuring the collective takes optimal action, overcoming individual temptations to defect. The second of these functions is similar to solving the prisoner's dilemma: hence there turns out to be a powerful analogy to the central paradigm for studying moral norms. But at the same time, between families, these norms serve a function that is not stereotypically moral, and can be related to a game of conflict (e.g., Hawk–Dove).

This may explain why honor norms appear to have similar motivational force as "moral" norms: they are, at one level of analysis, solving similar problems of free-riding as moral norms are believed to solve. So it is unsurprising that they are similar in their psychological profile. It also suggests a reason why honor norms might often appear less universal than stereotypical moral norms: their function is to promote the competitive performance of one collective *against* another, so the norms have no proper function outside the relevant collective.

What about the second paradigmatic honor norm: norms of purification? We believe this is best related to another sort of competitive paradigm: a competitive matching market for trading or exchange opportunities. In some cultures, marriage provides families with opportunities to improve economic prospects for the entire family. In many such cultures, for reasons that are not entirely clear, there is a very high premium placed on female sexual chastity as a prerequisite for marriage (Mackie 1996; Edlund 2018). This leads parents to restrict movement and education of female children from an early age, to arrange very early marriages to reduce opportunities for pre-marital sex, and other measures to control the sexual behaviour of adolescent females (Rai and Sengupta 2013). In some cultures of this sort, rumour of transgression by a female daughter can lead to forced marriage, to beating, and even to murder. Filicide of this sort is a particularly extreme incarnation of a purification norm. It is frequently referred to as "honor-killing" and it is typically undertaken in a way to ensure that the relevant community knows the murder has occurred. The family will often go out of its way to publicize rather than hide the killing. It is also usually done with implicit or explicit approval of the broader community, who regard the family's honor as being to some degree restored as a result.

Purification norms are even less well understood than revenge norms, but some work has been done to show that they may be intelligible in a strategic setting involving problems of trust and cooperation, broadly similar to a repeated prisoner's dilemma (Thrasher and Handfield 2018). A family (typically an extended family) whose child is suspected of transgression needs to convince other families that its remaining children are of good standing. If this can be demonstrated, then it stands to make favourable marriages for those remaining children, and this will be of mutual benefit. But in a setting where other families are competing for the same marriage opportunities, it is not easy to demonstrate the chastity of one's children, and even harder to convince that they will remain so in future. Honor killing may function as a costly signal of a family's confidence in the chastity of its remaining children: families that expect their other children will behave dishonorably have less to gain by sending a costly signal which will later be undermined, relative to a family confident that its future children will behave impeccably.

On this account, we can again distinguish two levels of function for such norms. Between groups, the norms serve as costly signals in a competitive marriage market. Because the signals are credible, they provide useful information, and depending on broader conditions, may even be welfare enhancing for families. This in no way justifies the practice but at least makes it potentially intelligible, from a biological perspective, how such a wasteful practice with respect to human life can persist.

Within groups, just like norms of revenge, norms of purification serve to overcome potential free-rider problems. While the extended family may be economically or biologically better off if an honor-killing is performed, this does not mean the task is easy to perform for any individual. In particular, honor-killers will often run the risk of judicial punishment. Thus, the norms of honor may be necessary to overcome the collective action problem and ensure that an individual performs the prescribed violence.

Does honor form a unified cluster?

Having argued that honor is part of the moral domain, because norms of honor serve a cooperative purpose, we now consider some more specific features of honor norms which might explain why it has been so tempting to regard them as a distinctive category. For reasons touched on already, it is unlikely anything like a satisfactory set of necessary and sufficient conditions could be offered to demarcate honor from the rest of morality, but it is certainly conceivable that honor norms could form a relatively stable cluster, maintained by processes that promote coinstantiation of certain phenotypic traits (Boyd 1991; Millikan 2017). Such clusters may not be especially natural (Barrett 2007; O'Connor 2018), but nonetheless apt for successful induction.[14] It is no doubt reasonable, having encountered a stable, unified cluster of phenotypic traits, over which one can make successful inductive inferences, to regard that cluster as itself a kind. We suspect this explains the continuing attraction of the view that honor and morality are distinct.

The unity of honor norms might be explained by their addressing a distinctive set of problems. Hence if honor norms were uniquely associated with some, but not all, of the MAC domains, this might be a basis for differentiation. But although honor does appear strongly associated with "heroism", deference, and loyalty to family, it seems (1) unlikely that honor norms occupy *only* these domains, especially when we note norms related to generosity and honesty; and conversely (2), it is unlikely that stereotypical norms of honor are the *only* candidates to occupy these particular domains. This suggests honor is just one possible set of solutions—likely incomplete – to the social challenges arising in these domains.

[14] Another alternative is that honor norms form a cluster whose unity is explained by *common ancestry* (Millikan 1999). But we find this implausible in the present case, given evidence that honor cultures arise spontaneously in the right environmental circumstances. One example is the US prison system, especially in states like California and Texas, where dramatic growth of the prison population has put traditional norms of prison behaviour under stress, and led to the emergence of gang organisations that are particularly florid examples of honor culture (Skarbek 2014).

Honor norms appear to be prominent in settings where, because exogenous and formal governance mechanisms are weak or non-existent, individuals rely on informal collectives such as the family, clan, gang, or tribe, in order to obtain protection (Weiner 2013; Skarbek 2014; Nisbett and Cohen 1996). Unsurprisingly, then, many norms of honor require profound sacrifices of individual wellbeing for the sake of the group—though we are aware of no way to systematically quantify the degree of sacrifice in honor cultures compared to other cultures. Indeed, in many honor cultures, individual identity appears to be much more closely integrated with group identity than in liberal, individualistic cultures (Sommers 2018). If these observations are correct, they might suffice to explain the relative unity of honor norms.

We also find important commonalities when we look in more detail at the strategic settings to which the norms appear to be adapted. Honor cultures are functional in more competitive, and less cooperative settings (Nowak et al. 2016). Although honor norms can make competitive situations relatively less conflictual, their home is in a setting of contest. This is obvious in our preferred model of norms of revenge: the Hawk–Dove game. In the marriage markets underlying violent purification norms, the conflict is less obvious, because markets are always in some sense cooperative. But in this case, we suggest, the limit on marriage opportunities and scarcity of elite male spouses entails significant amounts of competition for marriage to those elites (Edlund 2018). Thus, both purification and revenge norms appear to be characteristic of settings involving significant levels of competition.[15]

Finally, another mechanism which could give rise to clustering is if honor norms use a distinctive *means* of solving the social problems that they address. One promising hypothesis of this sort is that *signalling* is intrinsic to the functioning of honor norms. We have touched on signalling ideas in our discussion of purification norms above, but similar issues arise in the case of revenge norms. Revenge norms motivate behaviours which would reveal to potential aggressors that future aggressions would be met with dangerous retaliation. Given every family—even those who are not capable of retaliation—would like to make potential aggressors believe this message, any such attempt to communicate this idea is liable to be ignored.

Costly signalling theory can explain how it can be strategically stable to send signals that credibly indicate a disposition to retaliate (Thrasher and Handfield 2018). If the marginal cost of sending the signal is higher for non-vengeful families than vengeful ones, then it may be possible to find a signal that is sufficiently costly to "price out of the market" the non-vengeful families—they cannot afford to send a signal of equal strength as the vengeful families, so they will send no signal at all. Given that in many honor cultures, the relevant signal is itself a lesser act of revenge—such as provoking a fight over a minor insult—it is indeed plausible that families who are well-organized, committed to defending their honor, and disposed to vengeance will find it easier (less costly) to retaliate in lesser conflicts than less dangerous families. At its extreme, there may be acts of violence that less dangerous

[15] It is perhaps no surprise then that honor norms and honor language are still commonly used in explicitly martial contexts.

families simply cannot organize—performing acts such as this is a powerful demonstration (akin to a feat of strength or an ordeal) that is entirely credible.

Signalling appears to be central, then, to how norms of honor perform one of their key functions. Communities which embrace honor norms respond differentially to different types of agent (which may be groups or individuals), contingent on their honor status. The costly behaviours associated with defence of honor are the signals that make honorable status worth having, and make cooperative equilibria possible. Norms of morality, in contrast, do not appear to place such a central emphasis on signalling. While some accounts suggest particular altruistic or punitive behaviours have adaptive benefit because of their signalling value, (Henrich 2009; Jordan et al. 2016), the signalling role does not appear essential to the function of the norms but rather may be part of an ancillary apparatus which stabilizes the norms.

Conclusion

Norms of honour, like norms of morality, appear to play "pro-social" roles, promoting trust in institutions of property and marriage that are important for maintaining cooperation between families and clans.[16] To that extent, honor and morality appear similar: they are social technologies that facilitate cooperative solutions to social dilemmas. Indeed, we go further and say honour is best regarded as a subtype of morality understood generally as system of normative social technology.

If the above account is correct, it has potential implications for meta-ethical debate. In particular, this framework makes clearer what commitments will be acquired by anyone hoping to prosecute an argument for (Kumar and Campbell 2015) or against (Doris and Plakias 2008) moral realism on the basis of the apparent "conflict" between cultures of honour and cultures of morality. It certainly does not settle the realism/anti-realism debate, which turns on subtle issues about the semantics of moral language, but it does suggest that if a plausible naturalistic realism is acceptable [of the sort found in Sterelny and Fraser (2016)], it is likely to be a *pluralistic* or relativistic realism that embraces tragic conflict among moral requirements.[17]

In part, this follows from the theoretical presuppositions of the MAC framework. Even though MAC assumes moral technology is adaptive as a package, this is still consistent with there being significant *conflict* between norms adapted to different domains of cooperation. The domains identified by MAC are both sufficiently diverse and sufficiently ubiquitous that we may well find ourselves simultaneously subject to conflicting requirements. Morality may therefore be inherently tragic in this sense.

[16] Although we argue honour norms are functionally "pro-social", the norms themselves may require or allow what is typically considered "anti-social" and violent behaviour. But, as we have already argued, a considerable amount of violence is done is "moralistic" in one way or another (Black 1983; Fiske and Rai 2014).

[17] On this point, see also Harman (2015).

Our central claims about the relationship between honour and morality are evidently empirical, and hence defeasible in light of future science. Whether the individual claims survive is less important than the broader project, which is to demonstrate that there are rich resources available within an evolutionary framework to improve understanding of the relationship between different types of norm. This approach is in tension with the normative project which tends to regard "moral" norms as importantly distinct from all other norms, and emphasises properties such as universality, categoricity, and objectivity. While we have done nothing to undermine that normative project, we advocate that the descriptive project engaged in here has a much stronger prospect of contributing to the explanation of actual human practices, with the goal being a truly scientific understanding of morality. This is a goal towards which, we suggest, philosophers should be willing to contribute.

Acknowledgements Thanks to Patrick Emerton, Taylor Davis, Erik Kimbrough, Matthew Lindauer, Elizabeth O'Neill, Jordan Adamson, Tom Parr, and audiences at Monash University and the California Workshop on Evolutionary Social Sciences for comments on previous versions of this paper. Funding was provided by Australian Research Council (Grant No. DP150100242).

References

Abbink K, Gangadharan L, Handfield T, Thrasher J (2017) Peer punishment promotes enforcement of bad social norms. Nat Commun 8(1):609. https://doi.org/10.1038/s41467-017-00731-0

Allen DW, Reed CG (2006) The duel of honor: screening for unobservable social capital. Am Law Econ Rev 8(1):81–115

Appiah KA (2011) The honor code: how moral revolutions happen. W. W. Norton & Company, New York

Baier K (1954) The point of view of morality. Australas J Philos 32(2):104–135

Barrett JA (2007) Dynamic partitioning and the conventionality of kinds. Philos Sci 74(4):527–546. https://doi.org/10.1086/524714

Bendor J, Swistak P (1997) The evolutionary stability of cooperation. Am Polit Sci Rev 91(2):290–307. https://doi.org/10.2307/2952357

Bicchieri C (2006) The grammar of society: the nature and dynamics of social norms. Cambridge University Press, Cambridge

Birch J (2017) The philosophy of social evolution. OUP Oxford, Oxford

Black D (1983) Crime as social control. Am Sociol Rev 48(1):34–45. https://doi.org/10.2307/2095143

Boehm C (1986) Blood revenge: the enactment and management of conflict in Montenegro and other tribal societies. University of Pennsylvania Press, Philadelphia

Bowles S, Gintis H (2011) A cooperative species: human reciprocity and its evolution. Princeton University Press, Princeton

Boyd R (1991) Realism, anti-foundationalism and the enthusiasm for natural kinds. Philos Stud 61(1–2):127–148. https://doi.org/10.1007/BF00385837

Boyd R, Richerson P (1988) Culture and the evolutionary process. University of Chicago Press, Chicago

Brennan G, Eriksson L, Goodin RE, Southwood N (2013) Explaining norms. Oxford University Press, Oxford

Brown RP, Osterman LL (2012) Culture of honor, violence, and homicide. In: Shackelford T, Weekes-Shackelford V (eds) The Oxford handbook of evolutionary perspectives on violence, homicide, and war. Oxford University Press, Oxford, pp 218–232

Buchtel EE, Guan Y, Peng Q, Su Y, Sang B, Chen SX, Bond MH (2015) Immorality east and west: are immoral behaviors especially harmful, or especially uncivilized? Personal Soc Psychol Bull 41(10):1382–1394. https://doi.org/10.1177/0146167215595606

Chesler P (2010) Worldwide trends in honor killings. Middle East Q 17:3–11

Chesler P, Bloom N (2012) Hindu vs. muslim honor killings. Middle East Q 19(3):43–52
Curry OS, Chesters MJ, van Lissa CJ (2019a) Mapping morality with a compass: testing the theory of 'morality as cooperation' with a new questionnaire. J Res Personal 78:106–124
Curry OS, Mullins DA, Whitehouse H (2019b) Is it good to cooperate? Curr Anthropol 60(1):47–69
Davis T The scope and structure of the moral domain: an empirical study (unpublished manuscript)
Doris JM (2005) Lack of character: personality and moral behavior. Cambridge University Press, Cambridge
Doris JM, Plakias A (2008) How to argue about disagreement: evaluative diversity and moral realism. In: Sinnott-Armstrong W (ed) Moral psychology, vol 2. The cognitive science of morality: intuition and diversity. MIT Press, Cambridge, pp. 303–331
Edlund L (2018) Cousin marriage is not choice: muslim marriage and underdevelopment. AEA Pap Proc 108:353–357. https://doi.org/10.1257/pandp.20181084
Fiske AP, Rai TS (2014) Virtuous violence: hurting and killing to create, sustain, end, and honor social relationships. Cambridge University Press, Cambridge
Frank RH (1988) Passions within reason: the strategic role of the emotions. W. W. Norton & Company, New York
Fraser B (2012) The nature of moral judgements and the extent of the moral domain. Philos Explor 15(1):1–16. https://doi.org/10.1080/13869795.2012.647356
French SE (2004) The code of the warrior: exploring warrior values past and present. Rowman & Littlefield Publishers, Lanham
Gaus G (2014) Review of review of explaining norms, by Geoffrey Brennan Southwood Lina Eriksson, Robert E. Goodin and Nicholas. https://ndpr.nd.edu/news/explaining-norms/. Accessed 7 Nov 2018
Gaus G (2015) The egalitarian species. Soc Philos Policy 31(2):1–27
Gauthier D (1986) Morals by agreement. Clarendon Press, Oxford
Gavrilets S, Richerson PJ (2017) Collective action and the evolution of social norm internalization. Proc Natl Acad Sci. https://doi.org/10.1073/pnas.1703857114
Gaynor T (2011) Iraqi guilty of murder in daughter's honor killing. *Reuters*, 22 Feb 2011. http://www.reuters.com/article/2011/02/22/us-arizona-iraqi-idUSTRE71L8IT20110222
Ginsburg T (2011) An economic interpretation of the pashtunwali. Univ Chicago Legal Forum 2011:89–114
Gintis H (2003) The hitchhiker's guide to altruism: gene-culture coevolution, and the internalization of norms. J Theor Biol 220(4):407–418. https://doi.org/10.1006/jtbi.2003.3104
Gintis H, Fehr E (2012) The social structure of cooperation and punishment. Behav Brain Sci 35(1):28–29. https://doi.org/10.1017/S0140525X11000914
Graham J, Nosek BA, Haidt J, Iyer R, Koleva S, Ditto PH (2011) Mapping the moral domain. J Personal Soc Psychol 101(2):366–385
Graham J, Haidt J, Koleva S, Motyl M, Iyer R, Wojcik SP, Ditto PH (2013) Chapter two—moral foundations theory: the pragmatic validity of moral pluralism. In: Devine P, Plant A (eds) Advances in experimental social psychology, vol 47. Academic Press, Cambridge, pp 55–130. https://doi.org/10.1016/B978-0-12-407236-7.00002-4
Grosjean P (2014) A history of violence: the culture of honor and homicide in the us south. J Eur Econ Assoc 12(5):1285–1316. https://doi.org/10.1111/jeea.12096
Haidt J, Koller SH, Dias MG (1993) Affect, culture, and morality, or is it wrong to eat your dog? J Personal Soc Psychol 65(4):613–628
Hamilton WD (1964a) The genetical evolution of social behaviour. I. J Theor Biol 7(1):1–16
Hamilton WD (1964b) The genetical evolution of social behaviour. II. J Theor Biol 7(1):17–52
Harman G (1975) Moral relativism defended. Philos Rev 84(1):3–22
Harman G (1999) Moral philosophy meets social psychology: virtue ethics and the fundamental attribution error. Proc Aristot Soc 99:315–331
Harman G (2015) Moral relativism is moral realism. Philos Stud 172(4):855–863
Henrich J (2004) Cultural group selection, coevolutionary processes and large-scale cooperation. J Econ Behav Organ 53(1):3–35. https://doi.org/10.1016/S0167-2681(03)00094-5
Henrich J (2009) The evolution of costly displays, cooperation and religion: credibility enhancing displays and their implications for cultural evolution. Evol Hum Behav 30(4):244–260
Henrich J, Henrich N (2007) Why humans cooperate: a cultural and evolutionary explanation. Oxford University Press, Oxford
Jordan JJ, Hoffman M, Bloom P, Rand DG (2016) Third-party punishment as a costly signal of trustworthiness. Nature 530(7591):473–476. https://doi.org/10.1038/nature16981

Joyce R (2001) The evolution of morality. Cambridge University Press, Cambridge
Joyce R (2006) Metaethics and the empirical sciences. Philos Explor 9(1):133–148. https://doi.org/10.1080/13869790500492748
Kant I (1785) Grounding for the metaphysics of morals. In: Ethical philosophy: the complete texts of grounding for the metaphysics of morals, and metaphysical principles of virtue, part II of the metaphysics of morals, with on a supposed right to lie because of philanthropic concerns (trans: Ellington JW). Hackett, pp 1–65
Kelly D (2011) Yuck!: The nature and moral significance of disgust. MIT Press, Cambridge
Kelly D, Stich S, Haley KJ, Eng SJ, Fessler DMT (2007) Harm, affect, and the moral/conventional distinction. Mind Lang 22(2):117–131. https://doi.org/10.1111/j.1468-0017.2007.00302.x
Kitcher P (2011) The ethical project. Harvard University Press, Cambridge
Kumar V, Campbell R (2015) Honor and moral revolution. Ethical Theory Moral Pract 19:147–159
LaVaque-Manty M (2006) Dueling for equality: masculine honor and the modern politics of dignity. Polit Theory 34(6):715–740. https://doi.org/10.1177/0090591706291727
Leung AK-Y, Cohen D (2011) Within- and between-culture variation: individual differences and the cultural logics of honor, face, and dignity cultures. J Personal Soc Psychol 100(3):507–526. https://doi.org/10.1037/a0022151
Mackie G (1996) Ending footbinding and infibulation: a convention account. Am Sociol Rev 61(6):999–1017
Maynard Smith J (1982) Evolution and the theory of games. Cambridge University Press, Cambridge
Maynard Smith J, Price GR (1973) The logic of animal conflict. Nature 246(5427):15–18
McCabe K, Rigdon M, Smith V (2003) Positive reciprocity and intentions in trust games. J Econ Behav Organ 52:267–275
Milgram S (1974) Obedience to authority: an experimental view. Harper & Row, New York
Millikan RG (1999) Historical kinds and the 'special sciences'. Philos Stud 95(1–2):45–65. https://doi.org/10.1023/A:1004532016219
Millikan RG (2017) Beyond concepts: unicepts, language, and natural information. Oxford University Press, Oxford
Moehler M (2018) Minimal morality: a multilevel social contract theory. Oxford University Press, Oxford
Nichols S (2002) Norms with feeling: towards a psychological account of moral judgment. Cognition 84(2):221–236. https://doi.org/10.1016/S0010-0277(02)00048-3
Nisbett RE, Cohen D (1996) Culture of honor: the psychology of violence in the south. Westview Press, Boulder
Nowak MA, Sigmund K (2005) Evolution of indirect reciprocity. Nature 437(7063):1291–1298. https://doi.org/10.1038/nature04131
Nowak A, Gelfand MJ, Borkowski W, Cohen D, Hernandez I (2016) The evolutionary basis of honor cultures. Psychol Sci 27(1):12–24. https://doi.org/10.1177/0956797615602860
Nucci LP, Turiel E (1978) Social interactions and the development of social concepts in preschool children. Child Dev 49(2):400–407. https://doi.org/10.2307/1128704
O'Connor C (2018) Games and kinds. Br J Philos Sci. https://doi.org/10.1093/bjps/axx027
O'Neill E (2018) Kinds of norms. Philos Compass 12(5):e12416. https://doi.org/10.1111/phc3.12416
Purzycki B, Pisor AC, Apicella C, Atkinson Q, Cohen E, Henrich J, McElreath R et al (2018) The cognitive and cultural foundations of moral behavior. Evol Hum Behav 39:101–132
Quintelier K, Fessler D (2015) Confounds in moral/conventional studies. Philos Explor 18(1):58–67. https://doi.org/10.1080/13869795.2013.874496
Quintelier K, Fessler D, De Smet D (2012) The case of the drunken sailor: on the generalisable wrongness of harmful transgressions. Think Reason 18(2):183–195. https://doi.org/10.1080/13546783.2012.669738
Rai B, Sengupta K (2013) Pre-marital confinement of women: a signaling and matching approach. J Dev Econ 105(November):48–63. https://doi.org/10.1016/j.jdeveco.2013.07.003
Rozin P (1991) Family resemblance in food and other domains: the family paradox and the role of parental congruence. Appetite 16:93–102
Scanlon TM (1998) What we owe to each other. Harvard University Press, Cambridge
Skarbek D (2014) The social order of the underworld: how prison gangs govern the American penal system. Oxford University Press, Oxford
Skyrms B (2004) The stag hunt and the evolution of social structure. Cambridge University Press, Cambridge

Smith V (2008) Rationality in economics: constructivist and ecological forms. Cambridge University Press, Cambridge
Sommers T (2009) The two faces of revenge: moral responsibility and the culture of honor. Biol Philos 24(1):35–50
Sommers T (2018) Why honor matters. Basic Books, New York
Southwood N (2011) The moral/conventional distinction. Mind 120(479):761–802. https://doi.org/10.1093/mind/fzr048
Stanford PK (2018) The difference between ice cream and Nazis: moral externalization and the evolution of human cooperation. Behav Brain Sci. https://doi.org/10.1017/S0140525X17001911
Sterelny K, Fraser B (2016) Evolution and moral realism. Br J Philos Sci 68(4):981–1006. https://doi.org/10.1093/bjps/axv060
Stich SP (2017) The moral domain. In: Gray K, Graham J (eds) The Atlas of moral psychology. Guildford Press, New York
Street S (2010) What is constructivism in ethics and metaethics? Philos Compass 5(5):363–384
Sznycer D, Xygalatas D, Agey Eizabeth, Alami S, An X-F, Ananyeva KI, Atkinson QD et al (2018) Cross-cultural invariances in the architecture of shame. Proc Natl Acad Sci. https://doi.org/10.1073/pnas.1805016115
Thrasher J (2018) Evaluating bad norms. Soc Philos Policy 35(1):196–216
Thrasher J, Handfield T (2018) Honor and violence: an account of feuds, dueling, and honor killing. Hum Nat 29(4):371–389
Trivers RL (1971) The evolution of reciprocal altruism. Q Rev Biol 46(1):35–57
Turiel E (1983) The development of social knowledge: morality and convention. Cambridge University Press, Cambridge
Vanberg VJ, Congleton RD (1992) Rationality, morality, and exit. Am Polit Sci Rev 86(02):418–431. https://doi.org/10.2307/1964230
Vanderschraaf P (2018) Strategic justice: convention and problems of balancing divergent interests. Oxford Moral Theory. Oxford University Press, Oxford
Weiner MS (2013) The rule of the clan: what an ancient form of social organization reveals about the future of individual freedom. Farrar, Straus and Giroux, New York