

Martina Valković

May 2019

Word count: 5453

The role of norm externalisation in the evolution of cooperation

In a recent article, Kyle Stanford gives an evolutionary account of what he terms “externalisation”, understood as “our distinctive tendency to objectify or externalise moral demands and obligations” (Stanford, 2018, p.1). According to Stanford, externalisation is a distinctive feature of human moral psychology, which consists in our experience of the demands of morality as externally imposed on us. That is, we see ourselves as obliged to conform to these demands no matter what our subjective preferences are and, furthermore, we regard them as imposing the same unconditional obligations on other agents as well. Externalisation is adaptive, he claims, because it enables cooperation and protects it against free riders by “establishing and maintaining a connection between the extent to which an agent is herself motivated by a given moral norm and the extent to which she uses conformity to that same norm as a criterion in evaluating candidate partners in social interaction generally” (p.1), and this way of choosing partners makes cooperation possible.

Stanford’s proposal, although certainly innovative and intriguing, is not without its problems. In this paper, I will argue that the most important of these are a lack of conceptual clarity, confusion over the norm domain to which externalisation applies, and doubtful usefulness of externalisation in partner selection in early human societies, which is the evolutionary role assigned to it by Stanford. In the following, I will first present Stanford’s account of externalisation in more detail, after which I will discuss the main conceptual difficulties it encounters and offer an interpretation of his theory that I deem the most charitable. Then I will proceed to assess Stanford’s view on the adaptive role of externalisation and argue that it is doubtful that it had a role in partner selection. Finally, I will sketch out a potential

mechanism which could result in norm externalisation, and propose some possible alternative functions of it in the evolution of cooperation.

Stanford sets out to explore what he believes may be the most important distinctive feature of human moral psychology: externalisation, meaning our experience of moral demands as externally imposed on us, as an obligation to act regardless of our subjective preferences, and at the same time, as imposing the same unconditional obligations on everyone else (p.1). To further delineate the phenomenon in question and provide empirical support for his claims, Stanford cites studies (e.g. Turiel, 1983; Turiel et al. 1987) which allegedly show that children between 2.5 and 3 years old already begin to “reliably and systematically” differentiate between moral norms (e.g. pulling hair or stealing) and purely conventional norms (e.g. talking out of turn or drinking soup from a bowl). Violations of moral norms are considered “more serious and more deserving of punishment” than conventional norms, and their wrongness is typically justified by appealing to harm, fairness, justice, rights, or the welfare of victims (p.2-3). In addition to seeming more “serious”, moral norms are systematically viewed both as more generalizable, that is, as applying to people in other places and periods, as well as authority-independent, meaning that they cannot be suspended by some authority figure or institution (p.3).

However, Stanford also refers to more recent studies (e.g. Haidt et al., 1993; Kelly et al., 2007) that have shown that norms which do not include considerations or justifications based on harm, fairness, justice, welfare or rights, can also be seen as moral. Similarly, norms that do include these considerations and justifications need not be seen as moral, but can also be thought of as purely conventional. For these reasons, Stanford concludes that, although moral norms are frequently or prototypically connected to harm, justice, welfare, or fairness, these are not defining features of the norms themselves (p.3). Instead, he turns to the scale developed by Goodwin and Darley (2008; 2012), whose experiments have shown that subjects “reliably locate moral judgments at a particular point along a scale of increasing objectivity” (Stanford, p.3), ranging from judgments of taste or preference, to judgments of social convention, to moral judgments, to judgments of empirical or scientific facts. Stanford’s claim is that these

“categorical differences” in subjects’ willingness to tolerate the possibility of disagreement without error provide “perhaps the clearest way to characterize the sense in which we treat moral norms and judgments as systematically more objective than judgments of taste or preference or judgments of social convention” (p.3).

Stanford believes that this kind of externalisation presents us with an evolutionary puzzle, since merely subjective preferences for interaction with those who act prosocially would serve equally well to steer us towards interacting with those willing to cooperate, and away from those that are not. The question is, then, why did we develop this particular kind of motivation instead of a purely subjective one? Stanford’s answer is that externalisation was beneficial because it ensured the correlated interaction between agents willing to act cooperatively, thus protecting them from exploitation by free riders (p.8). Correlation in this context means that, for whatever reason, there is some tendency to interact with similar individuals (Skyrms, 2014). In the case of externalisation, this tendency is achieved by establishing and maintaining a connection between an agent’s motivation to adhere to a certain moral norm, and her use of conformity to that norm as a criterion when choosing potential interaction partners. In other words, if an agent is motivated by a norm and experiences it as somehow externally imposed on herself and others, she will have a preference for interacting with those who respect the norm, rather than those that do not.

In this way, those who conform to the same norms will prefer to interact with each other, instead of interacting with those who do not respect the norm. This is a beneficial outcome for them, as interacting with those who do not respect the norm could make them lose out by cooperating when the other does not. For example, consider an individual who adheres to the norm of fair division which makes her divide a jointly obtained resource (say, from collective hunting or gathering) in two equal parts. When such an individual interacts with another adherent of the same norm, each of them gets half of the resource. But, if she were to interact with someone who takes a great majority of the resource for himself, she would obviously lose out and, in future, it would be beneficial for her to avoid this individual and interact with those who respect her norm. This would explain why, although it is in principle possible for almost any norm to become

externalised, norms concerning harm, fairness, justice, rights and welfare are typically externalised: in order to maintain the evolutionary benefits of externalisation, norms protecting prosociality and cooperation from exploitation need to be preserved, and the above norms are vital for this (p.9).

Having provided this brief overview, let us now turn to the problems Stanford's theory encounters. First of all, it seems that Stanford's account is vulnerable to the objection of begging the question and suffering from a lack of conceptual clarity. His distinction between "moral" and "conventional" relies on externalisation. As we have seen above, although he enumerates criteria such as "seriousness", considerations of harm, fairness, justice, rights and welfare, which were all previously used by researchers to establish the moral/conventional distinction, more recent research compels him to conclude that norms need not exhibit these features to be considered moral. Instead, Stanford focuses on features like generalisability and authority-independence, opting for a criterion based on Goodwin and Darley's research (2008), and claiming that moral norms are those that are "reliably and systematically" located on the scale of increasing objectivity between social conventions and empirical facts. However, "reliably and systematically" is more than an overstatement, since Goodwin and Darley (2012) find considerable variation between subjects' judgments on the objectivity of moral beliefs. Setting this issue aside, another problem is Stanford's equation of "externalised", or "externalised to a specific degree", with "moral". As long as the norms that are externalised are also those that are typically considered, in everyday life as well as in philosophical discussions, as having a moral character, this equation looks unproblematic, and throughout the text Stanford uses examples that fit this picture. However, if it turns out that people externalise some norms which would not be considered moral either pre-theoretically or in philosophical theorising, as we will later see is the case, and also that they do not externalise some norms which would be considered moral on these pre-theoretic and theoretic grounds, it is doubtful that "externalised" should mean "moralised".

A further concern is that, notwithstanding his equation of "moral" with "externalised", Stanford uses the term "moral" in at least four senses: (1)

what people would *consider* to be or count as moral (e.g. p.1, 3, 7, 8, 13); (2) “phenomenological”, or what people *experience* as moral (e.g. p.2, 8); (3) exhibiting features of authority-independence and generalizability, which Stanford equates with being located on a certain point on a scale of objectivity (e.g. p.3, 4); (4) prosocial (e.g. p.2, 4, 13). These senses are often conflated, and Stanford seems to believe that they all pick out the same category. This makes it difficult to interpret Stanford’s use of phrases such as “distinctively moral”, or “specifically moral” as opposed to “merely conventional”, which he uses to assert that moral norms are externalised, whereas conventional ones are not. If “moral” here is understood as “externalised”, the argument becomes trivial, if not, it begs the question, as Stanford gives no delineation of moral and conventional norms apart from the claim that the former are externalised while the latter are not.

Another, related difficulty arises from Stanford’s claim that externalisation is unique to our experience of moral norms. While it is true that “humans experience the demands of morality as somehow imposed on us externally” (p.1), this experience is hardly limited to moral norms. As several authors have pointed out, externalisation is a feature of norms in general, not only moral ones (Davis & Kelly, 2018; Isern-Mas & Gomila, 2018; Patel & Machery, 2018). Thus, epistemic, aesthetic, prosocial, antisocial and all other norms can be and are experienced as being externally imposed on us. That is, in all of these cases, there is a psychological, as well as a conceptual, distinction between one’s subjective preferences and normative standards. Take, for example, aesthetic norms. One can believe that Beethoven’s music is absolutely superior to Britney Spears’ based on some objective, externally imposed standards valid for everyone. Yet, one may at the same time have a much stronger preference for indulging in a guilty pleasure of listening to “Hit Me Baby One More Time” for hours on repeat, while never doing so with the 9th Symphony. Or, to borrow an example of an epistemic norm from Davis & Kelly (2018), consider the norm of inductive inference that prescribes basing your extrapolations on a large rather than a small sample. If one adheres to this norm, one does not experience this as a matter of taste, habit, or prejudice, a subjective preference which might unproblematically change, but as an externally imposed norm with which everyone should comply, because it generally leads to more accurate results.

These examples show that norms which do not fall within the moral domain can also be experienced as externally imposed, objective and exerting their force on everyone. Once we realise that the fact that a norm is externalised does not mean it is moralised, we can say that, for instance, norms of respecting personal space or handshake norms may in some cases be externalised, which does not mean they are moral on anyone's account. In other words, externalisation does not separate moral norms from other types, such as social, epistemic, aesthetic ones, and it is present in all of these kinds. While Stanford may object that in moral norms externalisation is present to a higher degree, the basis for this claim is unclear. As Davis and Kelly (2018) have pointed out, in the studies cited by Stanford the respondents were not asked to judge degrees of objectivity, but rather to categorically judge the claim as either objective or not. Furthermore, it is hard to see what would "more objective" or "less objective" even mean, and it might be more plausible to interpret these judgments of degree as in fact referring to importance or significance of the norm or, alternatively, the scope of norm application. Thus, when someone claims that a moral norm is more objective than an aesthetic one, he might in fact mean that the moral norm is more significant, or that it is more important to obey such norms, and not that it is actually more objective than the aesthetic norm. Similarly, he might hold that those norms that apply to a larger number of people, for example all members of the society, are "more objective" than those that only apply to a smaller group, such as a particular subgroup within the larger society. In short, unless we are willing to consider all norms moral, it would be false to equate externalisation with moralisation.

While Stanford presents externalisation as being about moral norms, it actually does not have to be connected to it. I believe that one of the merits of Stanford's account is that he does not need to keep the conventional/moral distinction in order to describe the phenomenon of externalisation and to provide an evolutionary account of it as enabling and maintaining cooperation. In this text, I will preserve the distinction between the term "externalised" in the sense used by Stanford, that is, as referring to our specific experience of norms imposing themselves on us, and the term "moral", understood in our everyday sense, or in what I previously designated as the first sense in which Stanford uses it. I will then interpret Stanford's theory in the way I consider to be the most charitable, that is, by

focusing on the phenomenon of externalisation while abstaining from equating it, or even necessarily connecting it, with morality. The questions we then need to answer is why we are prone to externalising some norms but not others, and why this externalisation is adaptive.

Let us start with the second part of the question. According to Stanford, the role of externalisation in the specific aspect of the evolution of cooperation that Stanford assigns to it, namely partner selection, is questionable. I argue that externalisation will likely not do the job Stanford would like, or at least not in the way he proposes. It is unlikely that, in the circumstances in which human cooperation first evolved, avoiding those who do not adhere to the same norms as we do would be possible, let alone beneficial.

Firstly, in small groups, the cost of disagreement would select against shunning due to norm externalisation. As Brusse and Sterelny (2018, p.17) note, externalisation is potentially dangerous in small groups, since shunning everyone with whom you disagree over norms, without having the majority of the group on your side, might result in you effectively excluding yourself from the group, as you voluntarily deprive yourself of cooperation partners. Likewise, it is likely that, as a consequence of the appearance of new coordination problems, there would be disagreement among group members when it comes to the optimal response, and even if the group agrees on their common norms, they can still disagree on the way in which these norms apply in unfamiliar situations. For example, in a group that abides both by the norm of taking care for one's kin and punishing premarital sex, there will likely be a lack of consensus on which of these norms takes precedence. Such disagreements would threaten to result in group fracture.

Furthermore, it is doubtful that individuals could normally just switch to another tribe whose norms they support or avoid contact with or reliance on untrustworthy partners. While in large communities it might be true that individuals are free to choose their cooperation partners, at least to a degree, it is unlikely that this was the case in the early hunter-gatherer societies in which cooperation and prosociality evolved. Usually, members could not simply join a different tribe if they disagreed with the norms prevalent in their own group (van Prooijen, 2018, p.31). Neither it is likely that they had the luxury to stop interacting with or relying on individual members of

their tribe with whom they disagreed (Voorhees et al., p.32). If this line of reasoning is correct, human groups had an incentive to punish defectors rather severely, as they evolved to be obligatory co-operators – environmental and social challenges could only be met if the levels of defection were minimal.

Early hominins had to cope with threats such as extreme changes in climate, droughts, famines (van Prooijen, 2018, p.31), as well as rely on each other for their subsistence, considering that they depended both on collective hunting and on collective defense (Sterelny, 2012). No single individual would have much chance of surviving alone, outside of the group. In fact, humans came to rely more on each other to the extent that their social environment came to be more important to their fitness than the natural environment (Sterelny, 2012). These factors made going along with the group to avoid sanctions the adaptive strategy, which could also have included *internalisation* of the group norms, in the sense of acting in accordance with the norms of the majority in the group without much deliberation or thinking about the rightness or appropriateness of those norms. Thus, instead of Stanford's proposal that cooperation was feasible because of externalised norms, a simpler and more plausible explanation is that prosocial tendencies coupled with sanctions ensured that cooperation and prosocial norms were sustained.

Finally, if the norms to be externalised are indeed transmitted via enculturation, which seems the most plausible explanation, all or almost all members of a group would likely share the same norms, thus eliminating the option of differentiating between partners on the basis of the norms they hold dear. This is at least the case with relatively small and homogenous groups, and in the absence of a large-scale intergroup interaction. In small and tight-knit communities, where everyone is raised in similar ways and by similar people, and where people lead similar lifestyles that highly depend on other people in the group, it is likely that everybody will adhere to the same norms, which would make it impossible to choose between individuals on this basis. Norms and conventions evolve out of interactions between members of a group, and are sustained by their beliefs, expectations, and often sanctions. In the absence of these factors, norms simply cease to exist. At least in the case of individuals born within the

group, norms are mainly transmitted by observation and participation in the daily social life of the group rather than by formal instruction. Again, it might be the case that modern societies are comprised of different, more or less integrated social groups whose norms differ to a greater or lesser extent. But, as with partner selection, in this respect modern societies differ from ancient small-scale societies. If cooperation indeed evolved in small hunter-gatherer tribes, externalisation again could not have played an important role. If the pressure was on group members to remain close and coordinate smoothly to achieve their desired outcome, which would have to be the case in tribes that depend on pack hunting and collective defence, any significant creativity in norm creation or interpretation would be unwelcome. Thus, in small hunter-gatherer societies in which human cooperation first evolved, externalisation could not have played a meaningful role in partner selection, at least not through the process of shunning or avoiding group members whose norms differ from ours.

A possible reply is that this only means that norm externalisation is a relatively recent phenomenon, a product of the so-called broad-spectrum revolution (BSR) that occurred 100,000 to 400,000 years ago. According to Sterelny (2012, p.90-91), in this period there were several important changes in the way humans lived. One was the switch from hunting megafauna to hunting smaller prey, as well as hunting with more sophisticated weapons. The result was specialisation in diverse hunting and foraging niches and division of labour, as different techniques were needed for hunting different prey, and a shift to small-group or individual hunting. Another marked change was increased contact between different human groups, which led to trade but also competition (Zawidzki, 2013, p.104). The consequence of this BSR transition was a pervasive transformation in human group size, social organisation, technology and foraging practices (Sterelny, 2012, p.4). It is argued by Zawidzki (2013) and others that these new circumstances undermined the conditions which previously made cooperation cheap. Free riding, that is, exploiting cooperatively minded individuals without reciprocating, could not anymore be prevented by public food sharing and pooling of information. Information became "balkanized", transmitted only between smaller specialised groups within the society and rarely integrated. At the same time, increased interaction with unfamiliar individuals produced more opportunities for deception. It

could therefore be argued that externalisation developed in the post-BSR era as a mechanism to protect cooperation.

However, if information post-BSR was truly as balkanized as Zawidzki assumes, this would presumably also apply to information about one's conduct and, by extension, to norms to be externalised. We do not signal our adherence to norms by wearing colourful hats, but rather by acting in specific ways, expressing particular opinions in more or less public situations, joining in condemning others, and so on. Another source of this information are the accounts of others who have either interacted with *us* or have second-hand information regarding *our* conduct. Thus, in order to know to which norms a person adheres, we would have to have some knowledge of her previous conduct, which is allegedly not easily available in the post-BSR era of balkanized information.

However, it seems unlikely that information would become so compartmentalised that a decent estimate of how the other is going to behave would become unattainable. Notwithstanding specialisation and division of labour, humans still lived together and it is likely that they continued to interact across these and other divisions. But then, externalisation could have played a part at this stage of the evolution of human cooperation. Indeed, if it played a role at all, it must have been at this stage.

To summarise: externalisation could not have been helpful for partner selection in small, homogenous groups, due to the fact that norms to be externalised are acquired through enculturation, a process which would be more or less the same for every group member. There would be virtually no different partners with different norms to choose from, and shunning partners would likely not be possible to begin with. Once groups started interacting with each other on a more frequent basis, and became bigger and more heterogeneous, choosing and shunning partners based on their norms could become possible, but not if the information was as compartmentalised as Zawidzki and others suppose.

What about the mechanism underlying externalisation? Although Stanford does not speculate about a possible mechanism that could lead to externalisation, it may be instructive to do so. Notwithstanding the above insight that, in principle, all kinds of norms can be externalised, it may still

turn out to be true that, on average, people deem moral norms to be “more objective” than, for example, aesthetic norms, in the sense of being more important or as applying to a larger number of people. If this is so, it could be explained by the importance of “moral” norms for the group, that is for the group members’ interests, as in the case of prosocial norms such as those regarding welfare, justice, rights or fairness. In other instances, it might be explained by the value of signalling group membership, as may be the case when some norms that govern a group’s social life differ significantly from those of other groups. These factors make group enforcement of these norms beneficial in a way that enforcement of, say, aesthetic norms is not. Reaching common ground on the acceptability of theft is arguably more important than reaching it on the acceptability of Britney Spears’ music, as theft could be detrimental to the group members’ interests, while Britney’s music could not, at least not in most cases. Similarly, certain ways of greeting or dressing, for example a handshake or a headscarf, may be strongly externalised if they are taken to be a part of group identity, and thus as a signal as to with whom to cooperate and with whom not. In this sense, putatively moral and strongly externalised social norms may not only be experienced as externally imposed, they *are* externally imposed, in the sense that their violations are subject to social condemnation.

Van Prooijen (2018) develops this idea in terms of conformity pressures exerted on individuals by their immediate social environment in the form of imposing demands and obligations to follow them, which stimulates intrinsic agreement with these demands. He believes this is a more plausible and parsimonious explanation of the process of externalisation, especially considering that human morality evolved in small tribes that offered little choice as to whom to cooperate with, which makes adaptive maximising one’s own adherence to the demands of the group by objectifying them. Group members have an incentive to punish offenders, as violations of norms pertaining to harm, fairness, loyalty, obedience or contamination have a strong likelihood of reducing their evolutionary fitness. Furthermore, this punishment can be essentially cost-free for the group, as it may consist only in ridicule, gossip, or coalitional punishment, but at the same time it may be highly costly for the offender, since it can result in social exclusion, damaged reputation, or reduced access to resources of reproductive opportunities.

Isern-Mas and Gomila (2018) develop a similar idea with emphasis on connection between norms and values and motivations, and put it in terms of Darwall's second-person view of morality. According to this account, the objective character of norms aligns with subjective motivations because the norms themselves are grounded in claims and demands that emerge out of interaction between the community members, which the community eventually begins to sanction. The process through which this happens goes as follows. Individuals frequently interact out of evolved prosocial preferences, and when doing so they explicitly or implicitly make demands on each other. Over time, they form expectations about how others will or should act. At the same time, they are themselves motivated to comply with others' expectations, and know they can expect sanctions otherwise. Eventually, this results in group enforcement of these norms and in group members experiencing them as objective, while simultaneously being motivated by them. In this way, the experience of norms as externalised emerges as a side effect of the process of social interaction.

This kind of "externalisation" via group enforcement is not, of course, the psychological kind Stanford has in mind. However, it points to social and psychological mechanisms that might lead to our experience of norms as externalised. Gradually, through socialisation and enculturation and faced with widespread condemnation or approval of certain kinds of behaviour, we come to regard these norms as an (almost) objective feature of our world. Thus, it may be the case that common ground makes our prosocial preferences normative, as we come to internalise certain kinds of externally sanctioned behaviour. On this view, it would make perfect sense that the norms with consequences for the group benefit would be strongly externalised. Typically, whether you like ice-cream or not is of no importance to me, but whether you like Nazis or not matters substantially, as it says something about the ways in which you are likely to treat others. I, as well as the rest of the group, would then have interest in sanctioning your fondness to Nazis, while we would have no comparable interest in sanctioning your taste for ice-cream.

In short, the externalisation pattern that Stanford tries to account for could be explained by the process of enculturation, where group members punish some behaviours but not others, or punish some behaviours more strongly

than others, and it could even be coupled with some psychological dispositions to regard the group opinion as likely to be true. Over time, this results in our experience of these norms being as objective as empirical facts. Here then, is a potential mechanism which could lead to us externalising some norms, but not others. But is it adaptive, as van Prooijen thinks?

The evidence is inconclusive, but it seems at least possible. Externalisation could be one of the psychological traits that helps us get along with each other, as are, for example, prosocial tendencies, ability to empathise, inability to experience potentially destructive emotions, such as rage, for prolonged periods of time, and so on. The role of externalisation would then be to make conforming to the group opinion easier, as it comes to be seen as objective feature of our environment.

Furthermore, if we set aside the assumption that information in post-BSR societies was compartmentalised, externalisation could maybe contribute to established ways of protecting group members against exploitation, such as punishment and reputation-tracking. As already noted, punishment, which in highly social beings like us could consist in nothing more than a frown, ridicule, turning a cold shoulder, or coalitional punishment, is a highly efficient mechanism in that it can be exceptionally costly for the offender while being cost-free for the group. Similarly, reputation-tracking, which can be informed not only by direct contact with group members, but also by more indirect forms of gathering information about prospective cooperation partners, such as gossip, is another efficient mechanism for ensuring correlation between cooperative individuals.

It may be that externalisation could make these simple mechanisms even more effective. Coalitional punishment requires consensus among the punishers, which might be easier to achieve if the parties believe that they are acting on objective facts rather than on subjective beliefs that they just happen to share. Once people need to provide justification for their beliefs and actions, the simplest way to do so may be to claim that the normative beliefs they hold are a part of the objective world just like their empirical beliefs are. This could make them more confident in their beliefs, which can also result in greater willingness to punish and hold transgressors in contempt.

A further possibility is that the role of externalisation in the evolution of cooperation is akin to that of clothing styles, bodily adornment, and similar markers of group belonging and between-group difference. Considering that different groups have different norms, and that these are often used to discriminate between in- and out-group members, it is not implausible to think of norms in this way. However, it remains to be seen what would be the advantages of norms over other indicators of group difference, especially those hard or impossible to imitate, as is the case with, for example, some bodily modifications such as tattoos and scarification. A possible, though not necessarily particularly satisfactory answer is that, by styling their bodies or clothes in certain ways, people show they are different from those in other groups, whereas by externalising specific norms, and showing it, they indicate that they are not only different, but right, while members of the other group are in the wrong, which could motivate stronger attachment to the group. It is important to note that these are mere suggestions, and no matter how plausible or implausible one judges them to be, externalisation could still just as well be just a quirk of our psychology, with no adaptive use.

In summary, when we reconsider Stanford's account of externalisation in a way which does not include links to morality, we get a clearer idea of what externalisation could and could not accomplish. Pace Stanford, externalisation most probably did not play a role in partner selection in societies in which cooperation first evolved. In early human societies partner selection was, most likely, unattainable luxury. In the same environment, diversity of norms upon which this selection could operate is an equally implausible assumption. If externalisation did play a role, then that was in larger and more heterogeneous post-BSR societies, in that it would bolster ancient, tried-and-tested mechanisms such as punishment and reputation-tracking, and have some effect on partner selection. When we examine a possible way through which externalisation comes to exist, namely through social pressures combined with some of our psychological dispositions, we see that a possible adaptive role of externalisation is that it makes conforming to group opinion easier, which could confer significant evolutionary benefits in a species as obligatorily cooperative as ours. One of the other further possibilities is to think of norms as markers of group

belonging, and of externalisation as amplifying these markers from a level of difference to a level of disagreement.

Bibliography

- Brusse, C., & Sterelny, K. (2018). Moral externalisation fails to scale. *Behavioral and Brain Sciences, 41*, 17-18.
- Davis, T., & Kelly, D. (2018). Norms, not moral norms: The boundaries of morality do not matter. *Behavioral and Brain Sciences, 41*, 18-19.
- Goodwin, G. P. & Darley, J. M. (2008). The psychology of meta-ethics: Exploring objectivism. *Cognition, 106*, 1339–66.
- Goodwin, G. P. & Darley, J. M. (2012). Why are some moral beliefs perceived to be more objective than others? *Journal of Experimental Social Psychology, 48*, 250–56.
- Haidt, J., Koller, S. H. & Dias, M. G. (1993). Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology, 65*, 613–28.
- Isern-Mas, C., & Gomila, A. (2018). Externalization is common to all value judgments, and norms are motivating because of their intersubjective grounding. *Behavioral and Brain Sciences, 41*, 21.
- Kelly, D., Stich, S., Haley, K. J., Eng, S. J. & Fessler, D. M. T. (2007). Harm, affect, and the moral/conventional distinction. *Mind and Language, 22*, 117–31.
- Patel, S., & Machery, E. (2018). Do the folk need a meta-ethics? *Behavioral and Brain Sciences, 41*, 25-6.
- Skyrms, B. (2014). *Evolution of the Social Contract*. Cambridge University Press.

- Stanford, P. K. (2018). The difference between ice cream and Nazis: Moral externalization and the evolution of human cooperation. *Behavioral and Brain Sciences*, *41*, 1-13.
- Sterelny, K. (2012). *The evolved apprentice*. Cambridge: MIT Press.
- Tisak, M. S. & Turiel, E. (1988). Variation in seriousness of transgressions and children's moral and conventional concepts. *Developmental Psychology*, *24*, 352-57.
- Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge University Press.
- Turiel, E., Killen, M. & Helwig, C. (1987). Morality: Its structure, functions, and vagaries. In J. Kagan & S. Lamb (Eds.), *The emergence of morality in young children*, (pp. 155-243). University of Chicago Press.
- van Prooijen, J. (2018). Moral demands truly are externally imposed. *Behavioral and Brain Sciences*, *41*, 30-1.
- Voorhees, B., Read, D., & Gabora, L. (2018). The objectivity of moral norms is a top-down cultural construct. *Behavioral and Brain Sciences*, *41*, 31-40.
- Zawidzki, T. (2013). *Mindshaping*. Cambridge: MIT Press.