DOI: 10.1111/mila.12191

WILEY

ORIGINAL ARTICLE

From punishment to universalism

David Rose[1] | Shaun Nichols[2]

[1]Philosophy, Neuroscience and Psychology Program, Washington University, St. Louis, Missouri

[2]Department of Philosophy, University of Arizona, Tucson, Arizona

Correspondence
David Rose, Department of Philosophy, Washington University, One Brookings Drive, St. Louis, MO 63130.
Email: david.rose@wustl.edu

Funding information
Office of Naval Research, Grant/Award Number: #11492159

Many philosophers have claimed that the folk endorse moral universalism. But while some empirical evidence supports the claim that the folk endorse moral universalism, this work has uncovered intra-domain differences in folk judgments of moral universalism. In light of all this, our question is: *why* do the folk endorse moral universalism? Our hypothesis is that folk judgments of moral universalism are generated in part by a desire to punish. We present evidence supporting this across three studies. On the basis of this, we argue for a debunking explanation of folk judgments of moral universalism.

KEYWORDS
debunking, folk meta-ethics, punishment, universalism

1 | INTRODUCTION

Are the folk moral relativists? A number of contemporary philosophers (e.g., Joyce, 2002, p. 97; Mackie, 1977, p. 33; Shafer-Landau, 2003, p. 18), maintain that ordinary people presuppose that morality is not relativistic. Although Mackie and Joyce argue that commonsense is mistaken, others take commonsense to provide support for the denial of moral relativism. In effect, they make an inference from our ordinary view to the way the world actually is (e.g., Dancy, 1986, p. 172). In a similar vein, we find Ross claiming that those who would depart from ordinary belief owe an account of why it is that people could have been so badly misled (Ross, 1930, p. 81). And we find Mackie telling us that the error theorist "must give some account of how other people have fallen into what he regards as an error, and this account will have to include some positive suggestions … about what has been mistaken for, or has led to false beliefs" (pp. 17–18; see also Olson, 2014).

Though many philosophers have claimed that the folk do not endorse moral relativism, others have held that the folk meta-ethics reflects a deep confusion. So we see a dispute among philosophers over folk meta-ethics. Is the folk view borne out of confusion or is it perfectly sensible? We see little hope of settling this issue by carefully reflecting on what it might be that the folk are up to when considering the nature of morality. Instead, we take it that psychological work on what the folk think about morality and why they do so can help move the discussion forward.

Recent empirical work suggests that the folk do not endorse moral relativism. As unsurprising as this may be at first glance, the empirical work on folk meta-ethics has uncovered surprising intra-

domain differences in folk meta-ethical judgments. For instance, this work has suggested that while the folk classify euthanasia and racial discrimination as moral issues, they are more inclined—with respect to the former—to give a relativistic response, saying that if two individuals disagree, it is possible that neither one is wrong. But what might be underpinning these differences? Our view is that these judgments are facilitated by a motivation to punish and that differences in meta-ethical intuitions are generated, at least in part, by differences in the motivation to punish. If this is right, then this will not only further our understanding of the psychological processes underpinning folk meta-ethics, it will also bear on philosophical discussions.

The Plan: We will begin in sections 2 and 3, by briefly considering some of the background work in psychology on folk meta-ethics before turning to our own studies in sections 4–6. We will present several studies that show, in different ways, that the motivation to punish causally influences meta-ethical judgments. We will then go on, in section 7, to discuss how this kind of psychological research might bear on philosophical discussions.

2 | EMPIRICAL WORK ON FOLK META-ETHICS

Several recent studies suggest that ordinary people do, at least in some cases, deny that moral claims are relative. These researchers have suggested that the folk take moral claims to be objective. Precisely characterizing objectivity is, of course, itself a contested philosophical issue. But roughly speaking, the notion of objectivism is that the truth conditions for objective claims are independent of the attitudes and feelings people have toward the claim (e.g., Shafer-Landau, 2003). To determine whether people embrace objectivism for moral claims, the notion of objectivism is operationalized in different ways in different studies. But typically the studies draw on the philosophical strategy of deploying intuitions about disagreement to get at issues about objectivity: if a claim is objectively true, then anyone who denies the claim is mistaken. As a result, if two people disagree about some objective statement, then at least one of them has to be wrong. This is illustrated by uncontroversial cases of objectively true claims like "A hydrogen atom has one electron" or "7*5 = 35." The truth of these claims holds independently of anyone's attitudes about the claims. And if an alien denies that 7*5 = 35, then at least one of us has to be wrong. If we disagree about an objective claim, we both cannot be right.

If a moral claim is objective, then if two people disagree about the claim, at least one of them has to be wrong. Goodwin and Darley (2008) rely on this fact to explore lay attitudes about objectivism. They presented participants with a series of statements from different classes. Some were factual (e.g., "The earth is not at the center of the known universe"); some were social-conventional (e.g., "Calling teachers by their first name, without being given permission to do so, in a school that calls them 'Mr.' or 'Mrs.', is wrong behavior"); some were ethical (e.g., "Consciously discriminating against someone on the basis of race is morally wrong"), and some were matters of taste (e.g., "Frank Sinatra was a better singer than is Michael Bolton"). For each statement, participants were asked whether they agreed with the statement, and they were then told that another respondent said the opposite. After this, the participant was asked whether they think "the other person is surely mistaken" or that "it is possible that neither you nor the other person is mistaken." To count as an objectivist response, the participant had to reject the option that "it's possible that neither you nor the other person is mistaken" (p. 1352).

Goodwin and Darley found that people tended to give objectivist responses for both the ethical and factual statements but not for the statements about taste or social convention (pp. 1352–3). They summarize as follows: "individuals seem to treat core ethical beliefs as being almost as objective as

scientific or plainly factual beliefs, and reliably more objective than beliefs about social convention or taste." One of the striking findings from Goodwin and Darley—since replicated by Wright, Cullum, and Grandjean (2014)—is that there is diversity in the degree of objectivism *within* the domain of ethics (p. 1346). For instance, people are strongly objectivist about racial discrimination but not very objectivist at all about abortion (p. 1347) and euthanasia (p. 1351). Indeed, Wright finds that people classify issues like the death penalty and euthanasia as *both* moral and nonobjective (Wright et al., 2014). Perhaps this result should not really be so surprising to philosophy teachers. It is a familiar feature of teaching undergraduate ethics that students respond as relativists for many ethical issues, but few of them sustain their relativism when it comes to Hitler.

As interesting as these results are, we think the terminology is suboptimal. In the literature on the folk psychology of meta-ethics, researchers have tended to use the term "moral objectivism" as the contrast to moral relativism. This is in keeping with some philosophical discussions (e.g., Smith, 1993). However, the term "moral objectivism" often implies something stronger than the rejection of relativism; on one such usage, "objective" moral claims purport to describe facts or properties that are independent of anyone's feelings or attitudes about the claims (Finlay, 2007, pp. 821–822). One can, however, reject relativism without committing to mind-independent moral facts. The core claim that relativism rejects is that there is a *single true morality* (e.g., Harman, 2000). We will use the term "universalism" to refer to this anti-relativist view (e.g., Wong, 2006, p. xii).

3 | WHY DO PEOPLE BELIEVE IN UNIVERSALISM?

Given that the empirical work on folk meta-ethics has found intra-domain differences, we are interested in uncovering why there are these intra-domain differences. More generally, we are interested in discerning why people make the universalist judgments they do.

The extant work that attempts to determine why we believe in universalism largely pursues the idea that the belief in universalism is driven by emotional processes (e.g., Cameron, Payne, & Doris, 2013; Nichols, 2004; Prinz, 2007). The basic idea here seems plausible—emotions affect a wide range of attitudes about morality. There is now a bit of evidence in favor of the view that emotions impact universalist judgments. In what is perhaps the most rigorous study to date, Cameron et al. (2013) induced disgust in participants and probed about moral universalism by asking whether certain cultural practices were only wrong relative to the culture. Participants were shown a disgusting picture (or a control picture) and on the image would appear some activity practiced in a foreign culture (e.g., "Thieves have their hands cut off"). Participants were asked, "To what degree is the behavior morally wrong regardless of the culture in which it is practiced?" Cameron et al. found that disgust primes led to stronger judgments that the act was universally wrong. This shows that inducing disgust increases moral universalist responses. The study is elegant, but the actual effect is very small indeed. In one representative study, the people in the disgust prime condition gave responses that were on average higher than in the control condition—by .09 on a 5-point scale.[1] So, while the results suggest that emotions may have some impact on people's tendency to treat morality as universal, the results certainly do not indicate a large role for emotions in the processing that generates universalism judgments.

[1] One possibility is that disgust is the wrong emotion to prime for these violations. Seidel & Prinz (2013) have shown that anger has a much stronger effect than disgust on harm-based moral judgments. The acts used in Cameron et al. tended to be harm-based. So perhaps using an anger prime, rather than a disgust prime, would lead to larger effects of emotion on universalism judgments. We leave this as a question for future research.

In light of the disappointing findings on emotion and universalism, we will pursue a somewhat different line of explanation, one rooted in motivation. The basic idea of a motivational explanation was suggested already by Mackie:

> There are motives that would support objectification. We need morality to regulate interpersonal relations, to control some of the ways in which people behave towards one another, often in opposition to contrary inclinations. We therefore want our moral judgments to be authoritative for other agents as well as for ourselves: objective validity would give them the authority required (Mackie, 1977, p. 43).

As a psychological hypothesis, Mackie's proposal is rather vague.

We want to present a more specific version of the motivational hypothesis. In particular, we suggest that motivation to punish drives judgments of universalism. That is, we propose that the motivation to punish causally contributes to the belief in universalism. There are different ways this could hold. The motivation to punish might affect universalism judgments via basic emotions (e.g., anger) or by a process similar to dissonance reduction (e.g., Cooper, 2007), in which the subject wants to bring his universalist beliefs in line with his goal of punishing. Along the temporal dimension, it might be that in the course of development, the motivation to punish helps to establish intuitions about the universality of certain moral claims. Another (compatible) possibility is that the motivation to punish has an on-line effect on one's occurrent judgments about universalism, such that an occurrent motivation regarding punishment affects the extent to which an act is regarded as wrong. The developmental hypothesis is difficult to test, so we will focus largely on the on-line hypothesis. We now turn to our own empirical studies investigating this hypothesis.

4 | STUDY 1: INTRA-DOMAIN DIFFERENCES

As we have noted, not all moral claims are treated as equally universal. Given these intra-domain differences, our question is why the folk view some moral claims as being more (or less) universal than others. Our hypothesis is that folk judgments of moral universalism are infused with the motivation to punish. Accordingly, we expect that intra-domain differences in judgments of moral universalism are generated by differences in the motivation to punish.

4.1 | Method

4.1.1 | Participants

Seventy-one people participated (aged 18–62, M_{age} = 38 years, 43 female, 100% reporting English as their native language). Participants were U.S. residents, recruited through Amazon Mechanical Turk, tested online using Qualtrics, and compensated $0.40 cents for approximately 2–3 minutes of their time. The same basic recruitment and testing procedures were used in all subsequent studies. Repeat participation was prevented.

4.1.2 | Materials and procedure

Our strategy was to select two cases from Goodwin and Darley which display intra-domain differences in judgments of universalism. We selected one moral case that is known to attract a high proportion of universalism ratings (discrimination) and one case that is known to attract a low proportion of universalism ratings (euthanasia). Participants were given one of the two following scenarios:

Discrimination: Don consciously discriminated against someone on the basis of race.

Suppose that one day your classmate said "Don's behavior of consciously discriminating against someone on the basis of race is morally wrong." But another classmate, Chris, said "Don's behavior of consciously discriminating against someone on the basis of race is not morally wrong."

Euthanasia: Keith ethically assisted in the death of a terminally ill friend who wanted to die.

Suppose that one day your classmate said "Keith's ethical assisting in the death of a terminally ill friend who wanted to die is morally wrong." But another classmate, Chris, said "Keith's ethical assisting in the death of a terminally ill friend who wanted to die is not morally wrong."

After reading the case, participants were presented with two probes:

Universalism Probe: Given that these individuals have different judgments about this case, we would like to know whether you think at least one of them must be wrong, or whether you think both of them could actually be correct. In other words, to what extent would you agree or disagree with the following statement concerning such a case:

"Since your classmate and Chris have different judgments about this case, at least one of them must be wrong."

Punishment Probe: How much should Don/Keith be punished?

This formulation of the Universalism Probe was adapted from Sarkissian, Park, Tien, Wright, and Knobe (2011). Ratings were made on a 6-point scale anchored with 1 = completely disagree, 6 = completely agree. The Punishment Probe utilized a 7-point scale, anchored with 1 = not at all, 7 = very much.[2] Both probes were presented on separate screens and in a fixed order (Universalism first, Punishment second).

4.2 | Results

First, we replicated the basic finding from Goodwin and Darley, finding a significant difference in Universalism between Discrimination ($M = 3.79$, $SD = 1.52$) and Euthanasia ($M = 3.02$, $SD = 1.49$), $t(70) = 2.14$, $p < .05$, $d = .512$. Participants were more inclined to view discrimination as universal (see Figure 1).

Second, we found a significant difference in Punishment between Discrimination ($M = 5.12$, $SD = 1.53$) and Euthanasia ($M = 1.94$, $SD = 1.41$), $t(70) = 9.12$, $p < .001$, $d = 2.16$, with participants being more inclined to view the subject who engaged in discrimination as deserving of punishment than the subject who engaged in euthanasia (see Figure 1). Most importantly, to better understand the relationships among the variables, we ran a causal search on the data.[3] The search returned the model in Figure 2.

[2] The difference in scale length between the Universalism and Punishment Probes was accidental. In the remaining studies, we corrected this and 6-point scales were used for both questions. We find the same basic pattern of results.

[3] We ran a Greedy Equivalence Search (GES) using Tetrad (http://www.phil.cmu.edu/tetrad/). Roughly, GES operates by considering the possible models available given the different variables. GES assigns an information score to the null model (i.e., a disconnected graph) and then considers various possible arrows ("edges") between the different variables. To do so, it begins by adding the edge that yields the greatest improvement in the information score (if there is such an edge) and repeats the process until additional edges would not further improve the information score. GES then considers deletions which would yield the greatest improvement in the information score (if there is such an edge), repeating this procedure until no further deletions will improve the score. In all cases, the orientation of the edges is given by edge-orientation rules in Meek, 1997. It has been shown by Chickering (2002) that, given enough data, GES will return the true causal model of the data. Moreover, GES is often interpreted as returning the best fitting causal model, given the data. (For further details and some applications, see Chickering, 2002; Rose, Livengood, Sytsma, & Machery, 2011; Rose & Nichols, 2013.) Finally, we would note that we are fitting structural equation models, rather than running a series of regressions to test for mediation because structural equation models are more discriminating, offering the advantage of providing a measure of overall fit for a model and in many cases structural equation models outperform mediation analyses (Iacobucci, Saldanha, & Deng, 2007; see also Rose & Nichols, 2013).

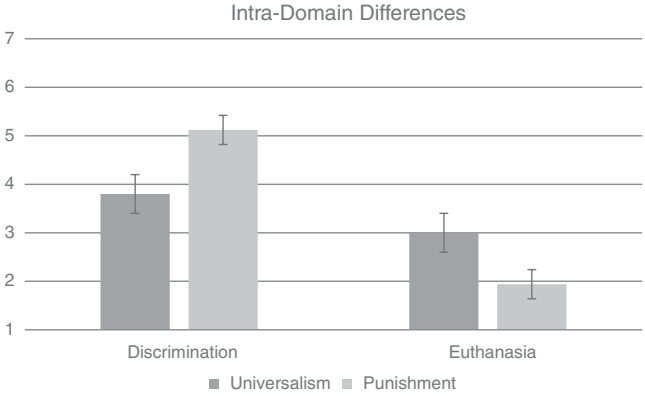

FIGURE 1 Intra-domain differences in judgments of moral universalism and punishment with 95% confidence intervals

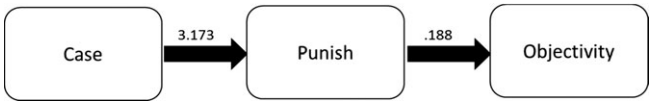

FIGURE 2 Model with punishment mediating effects of case on moral universalism judgments

This model fits the data well: df = 1, $\chi^2 = .4965$, $p = .4810$, BIC = −3.780.[4] It shows that assignment to one of the cases has a direct influence on people's inclinations to punish. More importantly for present purposes, the model also shows that the inclination to punish has a direct influence on people's judgments of universalism; the more one wants to punish, the more likely one is to offer a universalist judgment.[5] By contrast, a model that posits a causal arrow from universalism judgments to punishment judgments is rejected, df = 1, $\chi^2 = 51.5884$, $p = .0000$, BIC = 47.3117.

4.3 | Discussion

We have some initial evidence that the motivation to punish plays a causal role in judgments of universalism. Given that intra-domain differences in universalism seem to be explained in part by differences in the motivation to punish, we now want to consider whether the motivation to punish affects judgments of universalism for the *same* moral transgression.

5 | STUDY 2: PUNISHING THE YOUTH

In this study, we wanted to investigate whether punishment affects judgments of universalism for the same moral transgression.

5.1 | Method

5.1.1 | Participants

Ninety-three people participated (aged 18–67, $M_{age} = 36$ years, 47 female, 100% reporting English as their native language).

5.1.2 | Materials and procedures

To vary the motivation to punish, in one case we used an elderly character, and in the other case we used a young character. Our intuition was that even though the transgression was the same,

[4] Roughly, the null hypothesis for the chi-square goodness-of-fit is that the model fits the data. So $p > .05$ indicates that the model is a good fit; $p < .05$ indicates that the model is a poor fit. The Bayesian Information Criterion (BIC) provides an additional measure of model fit. For a discussion of BIC, see Kass and Raftery (1995).

[5] Note too that this model reverses the order in which the variables were measured.

participants would be more sympathetic to the older person, thinking he is less deserving of extensive punishment, while for the younger person, participants would be less sympathetic, thinking that the person is more deserving of extensive punishment. Thus, participants received the following case (variations in brackets):

[Old/Young]. In May of 2011, Don, a [70/20]-year-old employee of LLC Inc, who was in [poor/good] health, showed up for his last day of work.

Don had been struggling financially for some time now. His bank account was almost completely drained and he was becoming increasingly concerned about how he would make ends meet. Given that this was his last day at LLC Inc, he decided that this would be his one and only chance to get some extra cash.

He had info on one of LLC's wealthy investors. The investor was so wealthy that Don thought that if he transferred some money from the investor's account into his, that it would likely go unnoticed. So, Don decided to transfer $5,000 from the investor's account to his. This was the only time that he had ever stolen.

One month after the incident, Don was arrested and charged with grand theft. There was no evidence that Don had stolen in any other cases, but the evidence on this case was extremely clear.

After reading each case, participants were presented with the following information:

Suppose that one day your classmate said "Stealing $5000 from a company client is morally wrong." But another classmate, Chris, said "Stealing $5000 from a company client is not morally wrong."

Participants were then given two probes:

Universalism Probe: Given that these individuals have different judgments about this case, we would like to know whether you think at least one of them must be wrong, or whether you think both of them could actually be correct. In other words, to what extent would you agree or disagree with the following statement concerning such a case:

"Since your classmate and Chris have different judgments about this case, at least one of them must be wrong."

Punishment Probe: How much should Don be punished?

As in Study 1, both probes were presented on separate screens and in a fixed order (Universalism first, Punishment second). Rating for both probes was made on 6-point scales, utilizing the same anchors reported in Study 1.

5.2 | Results

We found a significant difference in Universalism between the Young ($M = 4.80$, $SD = 1.31$) and Old ($M = 4.15$, $SD = 1.67$) Don cases, $t(92) = 2.07$, $p < .05$, $d = .433$. Moreover, we found a significant difference in Punishment between the Young ($M = 4.80$, $SD = .865$) and Old ($M = 4.11$, $SD = 1.36$) Don cases, $t(92) = 2.92$, $p < .01$, $d = .605$. In short, participants were more inclined to give universalist judgments and assign more punishment when Don was described as being young. This can be seen in Figure 3.

To find out whether the motivation to punish plays a causal role in judgments of universalism, we ran a causal search on the data. The search returned the model in Figure 4.

This model fits the data well: df = 1, $\chi^2 = .7514$, $p = .3860$, BIC = −3.791.[6] Finally, as a point of comparison, we also constructed a structural equation model to see if a model with Universalism mediating the effect of the case on Punishment fit the data. This model is rejected, df = 1, $\chi^2 = 4.7669$, $p = .0290$, BIC = .2236. As with Study 1, the model in Figure 4 shows that

[6] Again, note that this model reverses the order in which the variables were measured.

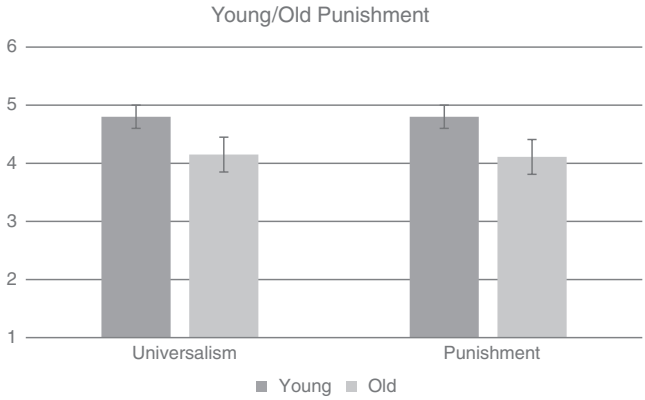

FIGURE 3 Effect of young/old Don on judgments of moral universalism and punishment with 95% confidence intervals

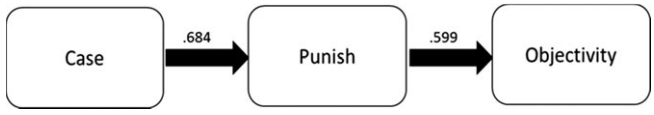

FIGURE 4 Model with punishment mediating effects of case on moral universalism judgments

assignment to one of the cases has an influence on people's judgments of punishment. And it shows that judgments of punishment play a causal role in people's judgments of universalism; the more one wants to punish, the more likely one is to give a universalist judgment.

5.3 | Discussion

Both Studies 1 and 2 provided support via causal modeling for the hypothesis that the motivation to punish plays a causal role in judgments of moral universalism. Given that the motivation to punish has an effect on judgments of universalism, this raises the intriguing possibility that if we intervene directly on the motivation to punish, then we should be able to see differences in universalist judgments. We will take this up in the next study.

We also want to address two main concerns about our studies thus far.[7] The first is that though we have been probing judgments about universalism via disagreement, disagreement is not a perfect measure of universalism.[8] To investigate whether disagreement is indeed tapping into intuitions about universalism, we will introduce a new universalism probe in the next study. Second, we wanted to explore whether the relationship between the motivation to punish and universalism was explained by affective responses.

6 | STUDY 3: OVERPUNISHING

6.1 | Method

6.1.1 | Participants

One hundred and twenty-six people participated (aged 18–67, M_{age} = 39 years, 58 female, 100% reporting English as their native language).

[7] We would like to thank two anonymous referees for raising these issues.

[8] For instance, it's not a universal fact that it's wrong to drive on the left side of the road. But if two people in the United States disagree about whether it's wrong to drive on the left, participants might well say that one of them must be mistaken.

6.1.2 | Materials and procedure

To directly intervene on the motivation to punish, we were guided by the idea that if an individual is severely punished for a transgression, then this should reduce our motivation to punish and thus our tendency to treat the behavior as universally wrong. But if an individual is not punished for a transgression at all, we will be left with the motivation to punish, and express this in universalist judgments. Thus, our strategy was to present participants with either a case where an individual is severely punished or a case where an individual is not punished at all. Thus, participants received one of the following two cases:

Severe Punishment. In May of 2011, Don, who was the manager of LLC Inc, consciously discriminated against Alvin on the basis of race and refused to hire him. The incident was reported and Don was arrested. There was no evidence that Don had discriminated in any other cases, but the evidence on this case was extremely clear. The state law allowed for punishments from probation up to lengthy prison term. The judge sentenced Don to 20 years to life in prison.

No Punishment. In May of 2011, Don, who was the manager of LLC Inc, consciously discriminated against Alvin on the basis of race and refused to hire him. There was no evidence that Don had discriminated in any other cases, but the evidence on this case was extremely clear. However, the incident was never reported and so Don never got caught.

After reading each case, participants were presented with a new universalism probe:

Universalism Probe. Please indicate the extent to which you think the statement "Don's behavior of consciously discriminating against someone on the basis of race is morally wrong" is an absolute truth.

Ratings were made on a 6-point scale anchored with 1 = There is no absolute truth about whether Don's behavior of consciously discriminating against someone on the basis of race is morally wrong and 6 = There is an absolute truth about whether Don's behavior of consciously discriminating against someone on the basis of race is morally wrong.

Finally, on a separate screen, participants filled out the Positive and Negative Affect Scale or PANAS (Watson, Clark, & Tellegen, 1988). The PANAS provides a measure of current affective states through 20 self-report measures, half of which target negative affect and half which target positive affect. Participants were presented with negative (e.g., upset, hostile) and positive affect words (e.g., enthusiastic, excited) and instructed to indicate the extent to which they felt that way right now, at the present moment. Ratings were made on the following 5-point scale: 1 = very slightly or not at all, 2 = a little, 3 = moderately, 4 = quite a bit, 5 = extremely.

6.2 | Results

First, we found that whether Don was severely punished ($M = 4.60$, $SD = 1.47$) or not punished at all ($M = 5.11$, $SD = 1.17$) produced a significant effect on judgments of moral universalism, $t(124) = 2.13$, $p < .05$, $d = .384$. This can be seen in Figure 5.

This result suggests two things. First, manipulating the motivation to punish affects universalism judgments. Second, given that we continue to find differences in universalism judgments, even when using a new measure, this suggests that our measure of universalism via disagreement does indeed tap into intuitions about universalism and not merely intuitions about disagreement.

Third, we ran a series of correlations between Universalism and Negative and Positive Affect as measured by the PANAS.

We found that neither Negative nor Positive Affect was correlated with Universalism (Table 1).

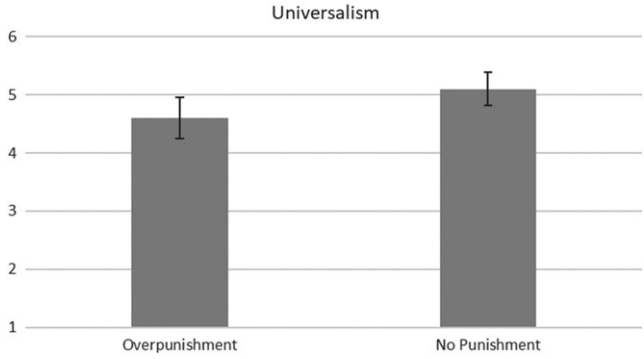

FIGURE 5 Effect of overpunishing on judgments of moral universalism with 95% confidence intervals

6.3 | Discussion

By directly intervening on the motivation to punish, we continue to find differences in universalist judgments, with people being less inclined to view behavior as wrong when the motivation to punish has been reduced, in this case through overpunishment. Moreover, we continued to find that the motivation to punish plays a role in universalist judgments even when utilizing a different measure of universalism. We also found that reported emotion on the PANAS was not related to universalism judgments. We would emphasize though that we are not denying that emotion plays any role in universalist judgments. Indeed, some of the evidence discussed above suggests that it does. Moreover, motivation is presumably connected to emotion. So, while we did not uncover a direct connection between reported emotion and universalism, it is likely that emotion is playing a role in the motivation to punish. We view our motivational hypothesis and the role of emotion in universalist judgments as entirely complementary. Indeed, insofar as our motivational hypothesis is correct, it may well be that emotion is related to the motivation to punish and runs through this to affect universalist judgments.

7 | GENERAL DISCUSSION

In a range of studies, we have provided evidence for a robust effect of the motivation to punish on judgments of universalism. We began by investigating, in Study 1, the intra-domain differences found in Goodwin and Darley, utilizing cases that are known to elicit different universalist judgments, and finding that punishment judgments about those kinds of cases play a causal role in universalist judgments. Study 2 looked at whether the motivation to punish would produce differences in universalism for the same moral transgression. Here, we found that by changing the sympathy for the criminal, we affect punishment judgments and that this in turn drives universalist judgments. Study 3 introduced something (overpunishment) that we would expect to reduce the motivation to punish, and this affected universalist judgments. Taken together, one important feature is that all of our

TABLE 1 Correlations between affect and universalism ($N = 126$)

Variables	1	2	3
1. Negative affect	—		
2. Positive affect	−.046	—	
3. Universalism	−.104	.009	—

studies are structurally quite different. Nonetheless, we find that a consistent, robust pattern emerges: The motivation to punish plays a causal role in generating judgments of universalism.

Our results add to the literature on motivated cognition and suggest that the motivation to punish can have surprising effects on ordinary judgments of moral universalism (e.g., Alicke, 1992; Alicke, 2000; Alicke, Rose, & Bloom, 2011; Clark et al., 2014; Ditto & Lopez, 1992; Ditto, Pizarro, & Tannenbaum, 2009; Kunda, 1990). In particular, they extend recent results by Clark and colleagues. They found, across a range of studies, that punitive motivations led to increased beliefs in free will. These findings suggest that the motivation to punish plays a role in free will beliefs. Our results suggest that the motivation to punish also plays a role in universalist judgments. And they might also explain, in part, why some previous research has uncovered intra-domain differences in judgments of moral universalism. These intra-domain differences appear to arise, in part, because of the motivation to punish. Indeed, our findings cohere well with the finding from Goodwin and Darley (2008) that morally wrong actions are seen as more universal than morally right actions. That said, we now want to consider an issue set out at the beginning of the paper in order to illustrate how work in psychology can contribute to disputes in philosophy and in particular to philosophical disputes over folk meta-ethics.

Some philosophers invoke the commonsense intuition that morality is not relative in order to provide support for the claim that morality is not relative. But philosophers disagree over whether the folk view morality as relative. For instance, Pojman claims that:

> [The] rejection of ethnocentrism in the West has contributed to a general shift in public opinion about morality, so that for a growing number of Westerners, consciousness raising about the validity of other ways of life has led to a gradual erosion of belief in moral objectivism. (Pojman, 2004, p. 239)

He continues: "in polls taken in my ethics and introduction to philosophy classes over the past several years…students by a two-to-one ratio affirmed a version of moral relativism over moral absolutism with hardly 3 percent affirming something in between these two polar opposites." By contrast, Michael Smith claims that the folk:

> …seem to think moral questions have correct answers; that the correct answers are made correct by objective moral facts; that moral facts are wholly determined by circumstances and that, by engaging in moral conversation and argument, we can discover what these objective moral facts determined by the circumstances are. (Smith, 1994, p. 6)

Moreover, philosophers who wish to depart from the folk view are thought to be required to provide some explanation of how the folk are mistaken. Psychological work can help move these discussions forward by uncovering not only what the folk think but why they do so. Indeed, in light of our evidence, we want to suggest that our results might serve to debunk philosophical views which are based on folk meta-ethical intuitions.

We noted in section 1 that some philosophers invoke commonsense intuitions in order to provide support for meta-ethical views. But we never discussed why folk meta-ethical intuitions might be thought to provide support for meta-ethical views. One standard approach holds that intuitions provide evidence for the truth of some philosophical claim or theory.

But now consider the results from our studies. We systematically found that variations in the motivation to punish led to variations in ordinary judgments about moral universalism. This makes it plausible that intuitions about moral universalism are—to some extent—caused by the motivation regarding punishment. And it is plausible that the kind of motivated reasoning in play here—from

the motivation regarding punishment to universalist beliefs—is irrelevant to the universality of moral claims. In other words, the motivation to punish is irrelevant to the truth of moral universalism and given that intuitions about universalism are affected by truth-irrelevant factors, in this case the motivation to punish, this undermines the evidential credentials of these intuitions.[9]

That is the basic idea behind the debunking argument we are setting out. But there are two objections we should address. The first is that we have cast our debunking explanation in terms of a motivation to punish. But one might point out that we can distinguish between two specific versions of the motivation-to-punish hypothesis. One version is that the motivation to punish is rooted in a desire to punish an agent; the other version is that the motivation to punish is rooted in a normative judgment that an agent should be punished. Perhaps our data suggest that it is a normative judgment that a subject should be punished which plays a role in universalist judgments. If that is right, then—so the objection might go—perhaps the motivation to punish does indicate the truth about whether moral claims are universal.

Even if the normative version of the motivational hypothesis is right, the debunking concern still applies. Consider the results from Study 2. We found that by inducing sympathy for an agent, judgments of universalism were affected. The motivation to punish (whether normative or desire based) is specific to a particular action (i.e., it was manipulated in the context of an individual engaging in a particular action). Yet the universalism question that participants were asked was not about the individual action but about the action type (i.e., stealing). The motivation to punish is extrinsic to the action type but nonetheless affects universalist judgments about an action type.[10] In light of this, the evidential relevance of the normative motivation to punish to the question of whether some moral claim is universal looks to be on par with the evidential relevance of whether someone should be blamed to the question of whether they caused some outcome.[11]

The second objection is that even if we are right that the motivation to punish—whether this be rooted in a desire or a normative concern—is evidentially irrelevant to the truth of universalism, our results only indicate that punishment plays a somewhat small role in universalist judgments. It is not clear that we should be at all concerned about this and so can still take intuitions of moral universalism to provide support for the truth of moral universalism. In response, we would point out that the crucial issue here is not the size. Instead, the important issue is how the motivation to punish is being used. And in this case, the issue is that people are moving from punishment to universalism: the motivation to punish is playing a causal role in generating judgments of moral universalism. So the fact that the motivation to punish plays a causal role in generating universalist judgments is enough to cast a shadow over the evidential standing of intuitions of moral universalism, given that it is unclear how the judgment that someone should be punished is evidentially relevant to the truth of moral universalism. Moreover, as noted in the introduction, we are only looking at the on-line effects of the motivation to punish. The on-line effect of the motivation to punish that we have demonstrated might

[9] See Nichols (2014) for an overview of debunking arguments in ethics.

[10] Perhaps here one might object that the use of "this case" by the disagreeing individuals leads participants to take this to refer to Don's action and not the action type of stealing from a company client. If that's right, then the token action is not affecting intuitions about the action type because participants are taking the question to be about the token act and not the action type. We doubt that this is the case. The information about Don was presented on a separate screen from the information about the disagreement. Had everything been presented together then perhaps it would be plausible to think that some people might take "this case" to refer to Don's specific action. Given that "this case" occurs on a different page where the two individuals are discussing the action type it is more plausible that participants take "this case" to refer to the action type that the two individuals are disagreeing over. Moreover, the disagreement between the classmates never involved Don at all. Nothing was said to indicate that his classmates knew anything about Don behaving in this way. Further, Don isn't mentioned again until after participants make their judgments about universalism. We would like to thank an anonymous reviewer for raising this concern.

[11] See Rose (2017) for a discussion of the irrelevance of blame to deciding whether an agent caused some outcome.

have a much more powerful effect over the course of development. That said, we do not doubt that judgments of universalism are influenced by a wide range of factors. The motivation to punish is surely not the complete psychological explanation of those judgments. Perhaps some of the factors that generate universalist beliefs are epistemically appropriate. However, it will be important to *show* that. By contrast, the influences on universalist judgments that we have shown are epistemically inappropriate.

We would also flag that we do not take our argument to debunk all folk intuitions about universalism. One reason is because our results clearly do not show that intuitions about universalism are always driven by a motivation to punish. Another is that we plausibly make universalist judgements in cases which involve morally good or morally right actions.[12] The debunking concern should thus be understood as follows: insofar as the motivation to punish plays a role in generating judgments of universalism, they are subject to debunking.

Finally, we would clarify that we are not taking our debunking argument to undermine the usefulness of *philosophers'* intuitions in disputes about meta-ethics. We only take our results to provide reason to worry that the psychological findings might provide the basis for an argument that debunks philosophical views that are based on commonsense intuitions about moral universalism. More importantly, we take our results to illustrate one key way in which work in psychology can contribute to philosophical discussions of meta-ethics.

8 | CONCLUSION

Many philosophers claim that the folk are committed to moral universalism. Some have taken this commitment to provide support for moral universalism. Others have taken the folk commitment to universalism to reflect a deep confusion. Our view is that work in psychology can help move this discussion forward.

We thus set out a range of empirical studies aimed at determining, at least in part, why the folk believe in moral universalism. Our first study revealed that the intra-domain difference in moral universalist judgments found by Goodwin and Darley are explained, in part, by the motivation to punish. In our second study, we looked at whether the motivation to punish would produce differences in judgments of universalism for the same moral transgression. We found that it did. Our third study approached the issue from a different direction. There, we intervened on the motivation to punish by presenting participants with a case where an individual is overpunished. Here, we found that overpunishing, in comparison to a case where an individual is not punished at all, reduces judgments of moral universalism.

These results provide support for the thesis that the motivation to punish affects judgments of moral universalism. The results also point toward a debunking argument for commonsense beliefs in universalism. Insofar as the commonsense belief in moral universalism is produced by the motivation to punish, it seems that that belief is not well justified. Our results thus suggest caution about relying on commonsense intuitions about moral universalism.

ACKNOWLEDGEMENTS

We would like to thank two anonymous reviewers for helpful comments on an earlier version of this paper. This paper was supported by Office of Naval Research grant #11492159, awarded to S.N.

[12] We would like to thank two anonymous referees for raising this point.

REFERENCES

Alicke, M. D. (1992). Culpable causation. *Journal of Personality and Social Psychology*, *63*, 368–378.

Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological Bulletin*, *126*, 556–574.

Alicke, M. D., Rose, D. & Bloom, D. (2011). Causation, norm violation, and culpable control. *The Journal of Philosophy*, *108*, 670–696.

Cameron, C. D., Payne, B. K. & Doris, J. M. (2013). Morality in high definition: Emotion differentiation calibrates the influence of disgust on moral judgments. *Journal of Experimental Social Psychology*, *49*, 719–725.

Chickering, D. (2002). Optimal structure identification with greedy search. *Journal of Machine Learning Research*, *3*, 507–554.

Clark, C. J., Luguri, J. B., Ditto, P. H., Knobe, J., Shariff, A. F. & Baumeister, R. F. (2014). Free to punish: A motivated account of free will belief. *Journal of Personality and Social Psychology*, *106*, 501–513.

Cooper, J. (2007). *Cognitive dissonance*. London: Sage Publications.

Dancy, J. (1986). Two conceptions of moral realism. *Proceedings of the Aristotelian Society, Supplementary Volumes*, *60*, 167–188.

Ditto, P. H. & Lopez (1992). Motivated skepticism: Use of different decision criteria for preferred and nonpreferred conclusions. *Journal of Personality and Social Psychology*, *63*, 568–584.

Ditto, P. H., Pizarro, D. A. & Tannenbaum, D. (2009). Motivated moral reasoning. In D. M. Bartels, C. W. Bauman, L. J. Skitka & D. L. Medin (Eds.), *Moral judgment and decision making* (pp. 307–338). San Diego, CA: Academic Press.

Finlay, S. (2007). Four faces of moral realism. *Philosophy Compass*, *2*(6), 820–849.

Goodwin, G. & Darley, J. (2008). The psychology of meta-ethics. *Cognition*, *106*, 1339–1366.

Harman, G. (2000). Is there a single true morality? In G. Harman (Ed.), *Explaining value: And other essays in moral philosophy*. Oxford: Clarendon Press, 77–99.

Iacobucci, D., Saldanha, N. & Deng, X. (2007). A mediation on mediation: Evidence that structural equation models perform better than regressions. *Journal of Consumer Psychology*, *17*(2), 140–154.

Joyce, R. (2002). *The myth of morality*. Cambridge: Cambridge University Press.

Kass, R. & Raftery, A. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*, 773–795.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*, 480–498.

Mackie, J. L. (1977). *Ethics: Inventing right and wrong*. New York, NY: Penguin.

Meek, C. (1997). *Graphical models: Selecting causal and statistical models.* (PhD thesis). Carnegie Mellon University.

Nichols, S. (2004). *Sentimental rules: On the natural foundations of moral judgment*. New York, NY: Oxford University Press.

Nichols, S. (2014). Process debunking and ethics. *Ethics*, *124*, 727–749.

Olson, J. (2014). *Moral error theory*. Oxford: Oxford University Press.

Pojman, L. (2004). Who's to judge? In Sommers & Sommers (Eds.), *Vice and virtue in everyday life*. Belmont, CA: Wadsworth Publishing.

Prinz, J. (2007). *The emotional construction of morals*. Oxford: Oxford University Press.

Rose, D. (2017). Folk intuitions of actual causation: A two-pronged debunking explanation. *Philosophical Studies*, *174*(5), 1323–1361.

Rose, D., Livengood, J., Sytsma, J. & Machery, E. (2011). Deep trouble for the deep self. *Philosophical Psychology*, *25*, 629–646.

Rose, D. & Nichols, S. (2013). The lesson of bypassing. *Review of Philosophy and Psychology*, *4*, 599–619.

Ross, W. D. (1930). *The right and the good*. Oxford: Clarendon Press.

Sarkissian, H., Park, J., Tien, D., Wright, J. & Knobe, J. (2011). Folk moral relativism. *Mind & Language*, *26*, 482–504.

Seidel, A. & Prinz, J. (2013). Sound morality: Irritating and icky noises amplify judgments in divergent moral domains. *Cognition*, *127*(1), 1–5.

Shafer-Landau, R. (2003). *Moral realism*. Oxford: Oxford University Press.

Smith, M. (1993). Realism. In P. Singer (Ed.), *A companion to ethics* (pp. 399–410). Cambridge, MA: Blackwell.

Smith, M. (1994). *The moral problem*. Oxford: Blackwell.

Watson, D., Clark, L. A. & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070.

Wong, D. B. (2006). *Natural moralities: A defence of pluralistic relativism*. New York, NY: Oxford University Press.

Wright, J., Cullum, J. & Grandjean, P. (2014). The cognitive mechanisms of intolerance. In J. Knobe, T. Lombrozo & S. Nichols (Eds.), *Oxford studies in experimental philosophy* (Vol. 1). Oxford: Oxford University Press.

How to cite this article: Rose D, Nichols S. From punishment to universalism. *Mind Lang*. 2019;34:59–72. https://doi.org/10.1111/mila.12191