

Are ethical explanations explanatory? Meta-ethical beliefs shape judgments about explanations for social change

Casey Lewry^{a,*}, George Tsai^b, Tania Lombrozo^a

^a Department of Psychology, Princeton University, United States

^b Department of Philosophy, University of Hawaii at Manoa, United States

ARTICLE INFO

Keywords:

Explanation
Ethics
Moral progress
Moral principles
Meta-ethics

ABSTRACT

Why were women given the right to vote? “Because it is morally wrong to deny women the right to vote.” This explanation does not seem to fit the typical pattern for explaining an event: rather than citing a cause, it appeals to an ethical claim. Do people judge ethical claims to be genuinely explanatory? And if so, why? In Studies 1 ($N = 220$) and 2 ($N = 293$), we find that many participants accept ethical explanations for social change and that this is predicted by their meta-ethical beliefs in moral progress and moral principles, suggesting that these participants treat morality as a directional feature of the world, somewhat akin to a causal force. In Studies 3 ($N = 513$) and 4 ($N = 328$), we find that participants recognize this relationship between ethical explanations and meta-ethical commitments, using the former to make inferences about individuals’ beliefs in moral progress and moral principles. Together these studies demonstrate that our beliefs about the nature of morality shape our judgments of explanations and that explanations shape our inferences about others’ moral commitments.

Why were women in the United States given the right to vote? Imagine your friend explains that “women were given the right to vote because denying this right is morally wrong.” Compare this with the explanation that “women were given the right to vote because the United States entered World War I, which allowed more women to take over men’s roles in the workforce and have a public voice.” What differs between these two types of explanations? Does the former seem genuinely explanatory, even though it explains the occurrence of an event by appeal to an ethical claim, rather than (non-ethical) causal factors? What would you assume about your friend’s ethical beliefs if they offered the first explanation, versus the second?

Ethical explanations, such as “women were given the right to vote because denying that right is morally wrong,” appeal to the morality of a practice as an explanation for social change.¹ But it is not clear whether people deem these explanations genuinely *explanatory* (akin to citing a causal force that accounts for the change) or merely *evaluative* (akin to “women were given the right to vote *and* denying that right is morally wrong”). Additionally, the implications of using ethical explanations to explain social change are unknown. When we hear someone give an ethical explanation, what do we infer about the explanation-provider’s moral commitments? For instance, do we think they are more likely to

believe in moral progress and endorse the existence of objective moral truths? Across four studies, we investigate whether and why people judge ethical explanations explanatory (Studies 1 and 2) and whether people use ethical explanations to draw inferences about others’ meta-ethical beliefs (Studies 3 and 4). These studies shed light on how our moral commitments shape explanatory judgments, and how our explanations communicate our moral commitments to others.

1. Are ethical explanations explanatory?

Explanations for events typically appeal to causes (e.g., Salmon, 1984; Woodward, 1989, 2003). For example, a good explanation for why Cheryl is tired, has gained weight, and has an upset stomach might cite her pregnancy as the cause (Read & Marcus-Newhall, 1993). Explanations for events can also appeal to generalizations or other claims with more indirect causal content. For instance, an explanation might appeal to a disposition (e.g., explaining that Jim helped his neighbor move their piano because he is a kind person; Budesheim & Bonnelle, 1998), mental reasons (e.g., explaining that the boy didn’t run in the mud because he didn’t want to ruin his shoes; Kalish, 1998; Malle, 1999, 2011), structural or situational factors (e.g., explaining giving someone

* Corresponding author at: Peretsman Scully Hall, Princeton, NJ, USA.

E-mail address: lewry@princeton.edu (C. Lewry).

¹ Throughout this paper we use “ethical” and “moral” to refer to the same concept, however we recognize that philosophers do not always use these terms interchangeably.

roses for Valentine's day because one lives in a country where marketers promote roses for this day; e.g., Vasil & Lombrozo, 2020; Tworek & Cimpian, 2016), or natural laws (e.g., explaining why a bowling ball slows down and pins go flying when the two collide by appealing to the law of the conservation of energy; Clayton, 2004).² The socio-historical explanation offered above, that women were given the right to vote because World War I allowed more women to enter the public sphere, is causal in this sense. By contrast, ethical explanations appeal to an ethical claim or the (un)ethical status of a practice, and, at least on the surface, do not offer causes or causal generalizations.

While there is some consensus among psychologists about how people typically explain everyday events (Lombrozo, 2012), there is disagreement among philosophers over whether ethical explanations are genuinely explanatory. For example, Harman (1986, 2007) argues that moral facts (e.g., that it is morally wrong for women not to have the right to vote) have no place in explanations of social change (e.g., why women were given the right to vote) because there is no plausible mechanism to identify how a moral fact could cause this change. But some have countered that there is an identifiable mechanism. Luco (2019) argues that there are moral facts that cause change through moral-cognitive mechanisms, akin to natural facts (like the fact that the earth revolves around the sun), which cause change through natural mechanisms. According to Luco's view, it could be a moral fact that it is unjust for women not to have the right to vote. Unjust conditions sometimes create unhappy societies, which can lead to unrest and ultimately social activism aimed at eliminating the injustice. Luco differentiates this mechanism from the alternative hypothesis that social changes are merely caused by widespread changes in individuals' beliefs (e.g., that women were given the right to vote because many people came to believe that denying this right is morally wrong, see Leiter, 2001). As discussed below, we also differentiate these hypotheses in our studies.

Sturgeon (1988) also assumes that moral facts exist and argues that they are necessary in our 'best explanatory picture' of the world. Consider the explanation that "slavery ended because slavery is morally wrong." Sturgeon asks us to consider the explanation that opposition to slavery in the US grew between the Revolutionary and Civil wars because slavery became a more oppressive (i.e., unjust) institution during that timeframe. He hypothesizes that most people would agree this is true (and even those who disagree would concede that the moral fact underlying the claim has some truth and forms part of a larger explanation). He takes this as evidence that people appeal to moral facts in standard explanations for social change; thus, moral facts are a necessary component of some explanations. Whether moral facts exist, and whether they have a role in our best explanatory picture of the world, is a common theme in these debates (Brink, 1989; Leiter, 2001).

The fact that experts disagree about the status of ethical explanations – while assuming that they are important in everyday reasoning – suggests that ethical explanations are a particularly informative test case for theories of explanation more broadly. In particular, are ethical explanations regarded as explanatory only to the extent that moral facts are taken to exist and to play a causal role in driving change? Or might ethical explanations for social change present an exception to the generalization that events are typically explained in causal terms? Moreover, disagreement among philosophers offers at least anecdotal evidence for individual variation in judgments about the status of ethical

explanations. Is this variation mirrored in laypeople's evaluations of ethical explanations, and if so, what drives it?

Given these open questions about the psychological status of ethical explanations, we consider a few hypotheses. One possibility is that people view ethical explanations as shorthand for ethical belief explanations, such as "women were given the right to vote because people came to believe that denying this right is morally wrong." Such an explanation would in fact be causal: it appeals to mental states that impacted behavior to bring about change. Another possibility is that ethical explanations are judged explanatory provided that an individual holds meta-ethical commitments (i.e., beliefs about the nature of morality) that make citing ethical content somewhat analogous to citing a causal force. That is, just as explaining a ball's fall by appeal to gravity is explanatory (provided that one thinks of gravity as a real and directional force that applies to the ball), explaining women's suffrage by appeal to its moral status could be regarded as explanatory as well (provided one accepts the immorality of denying women's suffrage as a moral fact that shapes the direction of social change).

2. Meta-ethical commitments

In the studies that follow we consider two meta-ethical commitments: belief in moral progress and belief in moral principles. For someone who holds such beliefs, an ethical truth might be seen as playing an explanatory role akin to a directed, causal force.

The first meta-ethical commitment that we consider is moral progress. Moral progress is the notion that the world is morally better today than it was in the past, and that the world will be morally better in the future than it is today. In other words, there is positive moral change over time. For instance, one might take women gaining the right to vote as clear-cut evidence that moral progress has occurred. There has been relatively little prior work on moral progress, but Lewry, Asifriyaz, and Lombrozo (2023) found that about two-thirds of adults in the United States (recruited on Prolific) believe that moral progress occurs. Rutjens and colleagues have also shown that belief in moral progress is predicted by feelings of personal control (Rutjens, van Harreveld, & van der Pligt, 2010) and the salience of death (Rutjens, van der Pligt, & van Harreveld, 2009; Rutjens, van Harreveld, van Der Pligt, van Elk, & Pyszczynski, 2016). We predict that variation in the extent to which an individual believes that moral progress occurs will partly account for the extent to which that individual judges ethical explanations explanatory, since moral progress could be thought of as a directional force that is capable of guiding morality forward. In this way, moral progress might contribute to an explanation of social change. We assess participants' belief in moral progress using measures adapted from Rutjens et al. (2016) in Study 1 and Lewry et al. (2023) in Study 2.

The second meta-ethical commitment that we consider is belief in moral principles. We use the term "moral principles" to capture a cluster of related commitments that are plausibly correlated and largely undifferentiated for laypeople. Some of these commitments concern what some have called moral "objectivism." For example, Goodwin and Darley (2008) characterize moral objectivism as the position that moral claims derive their truth independently of human minds. They assessed moral objectivism using two tasks. In one task, a moral claim had to be classified as true, false, or an opinion; responding "true" or "false" was interpreted as the objectivist response. In another task, participants were asked whether two individuals who disagree about a moral claim can both be right, or whether at least one of them must be mistaken. The latter response was taken to reflect objectivism. A similar task has been used in educational and developmental psychology to assess what has instead been called "absolutism," described as the belief that "assertions are facts that are correct or incorrect in their representation of reality" and that "knowledge comes from an external source and is certain" (Kuhn, 2000, p. 311; see also Kuhn, 1991; Rosman, Peter, Mayer, & Krampen, 2018; Metz, Liquin, & Lombrozo, 2023; Weinstock, Israel, Tabak, & Harari, 2020; Huiders, Voet, & De Wever, 2021; Mason,

² A more subtle case to consider concerns explanations for events that appeal to probabilistic generalizations. For example, explaining that a fair coin landed heads by saying that it had a 50% chance of doing so is not obviously causal, even though the flip causally led to the outcome. There is a lot more to be said about how non-causal factors might contribute to event explanations (e.g., Prasada & Dilligham, 2006; Strevens, 2018; Lombrozo & Vasil, 2017), but these nuances are not central to the contrast we draw between causal explanations and ethical explanations.

Boldrin, & Zurlo, 2006; Ionescu, Furdul, Gavreliuc, Greenfield, & Weinstock, 2023; Klopp & Stark, 2022; Felton & Kuhn, 2007).

Arguments for the explanatory role of ethical claims typically assume the existence of moral facts; both objectivism and absolutism plausibly track the extent to which laypeople endorse such an assumption. And matching the heterogeneity of views within the philosophical literature itself, work by Goodwin and Darley (2008) and others suggests considerable variation in the extent to which laypeople endorse moral objectivism and moral absolutism (Heiphetz & Young, 2017; Moss, Montealegre, Bush, Caviola, & Pizarro, 2023; Stanley, Marsh, & Kay, 2020). This variation allows us to test the hypothesis that belief in moral truths is a prerequisite to accepting ethical explanations. Specifically, we predict that those individuals who more strongly endorse independent moral truths should be more likely to judge ethical explanations explanatory.

Although the “opinion” and “disagreement” measures mentioned above have been used in subsequent work on moral objectivism (e.g., Goodwin & Darley, 2012; Sarkissian, Park, Tien, Wright, & Knobe, 2011), researchers have also raised questions about their relationship and whether they measure objectivism, absolutism, or something else (e.g., Pölzler & Wright, 2019, 2020; Pölzler & Wright, 2019; Uttich, Tsai, & Lombrozo, 2014; Wright, 2023; Zijlstra, 2019). For example, Pölzler and Wright (2019) argue that these measures do not successfully differentiate objectivism from some forms of subjectivism, such as forms of cultural relativism according to which a moral claim is (say) true within one given cultural context, but false within another. Zijlstra (2019) uses exploratory factor analysis to identify different aspects of folk moral objectivism, and on this basis develops a scale with three dimensions: universal truth (the idea that there are moral facts that are true no matter what), independent truth (the idea that there are moral facts independent of cultural variation), and divine truth (the idea that there are moral facts that depend on the existence of a divine entity). Despite these developments, the field currently lacks a generally accepted measure of moral objectivism or a universal taxonomy of meta-ethical commitments.

In light of the ongoing controversies concerning the measurement and characterization of moral objectivism, we use multiple measures across our studies: those from Goodwin and Darley (2008) in Study 1, and the scale developed in Zijlstra (2019) in Study 2. We also acknowledge that these are imperfect measures of moral objectivism as understood within philosophy (Pölzler & Wright, 2019, 2020; Pölzler & Wright, 2020). This motivates our use of “moral principles” (rather than moral objectivism or absolutism) as our preferred nomenclature. For our purposes, it is not critical that our measures capture objectivism per se, so long as they capture meaningful variability across individuals in the extent to which they endorse some or all of the suite of meta-ethical commitments that constitute or support belief in moral facts.

In sum, in Studies 1 and 2, we consider whether ethical explanations are indeed judged explanatory, whether they are differentiated from ethical belief explanations, and whether variation in their endorsement is predicted by variation in beliefs concerning moral progress and moral principles. For someone who holds such beliefs, an ethical truth might be seen as playing an explanatory role akin to a directed, causal force.

3. Communicating through (ethical) explanations

If ethical explanations are associated with meta-ethical commitments to moral progress and moral principles, do people recognize this association and use it to make inferences about others? In other words, if your friend says that “women were given the right to vote because denying them the right to vote is morally wrong,” rather than citing economic factors, are you more likely to infer that your friend believes in moral progress and that the immorality of denying suffrage is a moral fact?

Explanations clearly offer evidence about what the explanation-provider believes – in this case, that your friend believes women were

given the right to vote and that denying women the right to vote is morally wrong. These first-order inferences can be derived directly from the content of the explanation. But what kinds of second-order inferences are licensed by explanations? That is, what can one infer about an explanation-provider’s beliefs, desires, or intentions based on the fact that they chose to provide *that* explanation as opposed to another (see Fig. 1)? In this case, does the fact that someone provides an ethical explanation, rather than a non-ethical explanation, offer evidence about their meta-ethical commitments?

There is some evidence that explanation choice leads to these second-order inferences. Kirfel, Icard, and Gerstenberg (2022) illustrate this through the following thought experiment: imagine Suzy needs to pass two exams, physiology and anatomy, to be accepted into medical school. Suzy has also expressed that she feels much more confident in one subject than the other. One day, Suzy enthusiastically tells you that she got into medical school because she passed anatomy. Beyond communicating the literal content – that Suzy got into medical school and that she passed anatomy – we are also likely to infer that anatomy was the subject Suzy was less confident in. Other research suggests that the language used in an explanation, whether it is easy or hard to understand, and whether it is assigned a label can all impact the degree to which the explanation-provider is judged to be trustworthy (Elsbach & Eloffson, 2000). Thus, explanations can lead to judgments about the explanation-providers (in this case, their confidence and trustworthiness) even when this information is not explicitly stated.

To our knowledge, however, it is unknown whether the form of an explanation can be used to infer the underlying commitments that would render that explanation explanatory (i.e., belief in moral progress and moral principles in the case of ethical explanations). We expect that while first-order inferences about ethical explanations are derived from

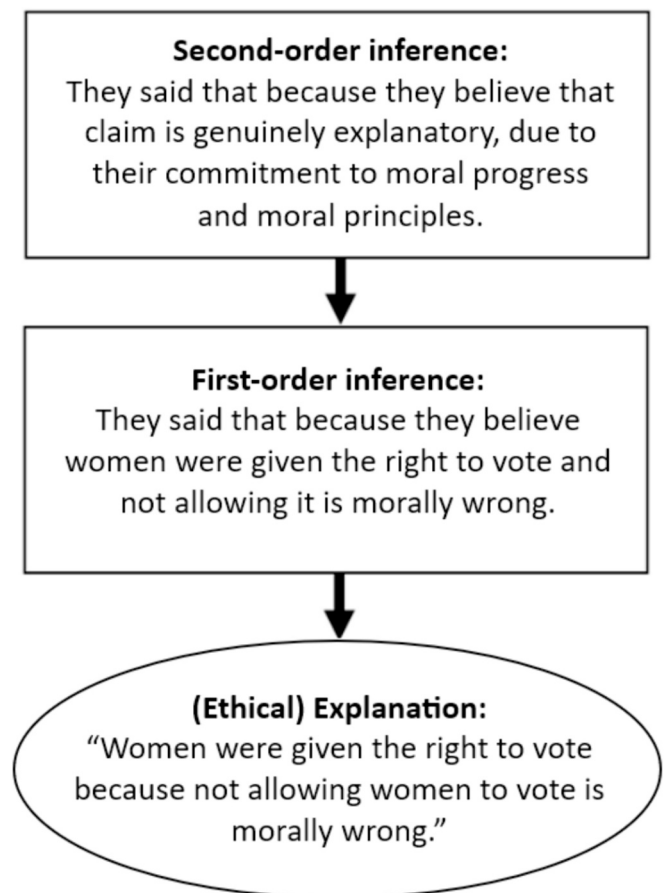


Fig. 1. Illustrative Example of the Relationship between Explanations and Inferences.

judgments about what the explanation-provider believes such that they are *willing to make that claim* (i.e., that women were given the right to vote and that denying it is immoral), second-order inferences are derived from judgments about what the explanation-provider believes such that they *find that claim explanatory* (i.e., that they believe in moral progress and moral principles, and as such, view an ethical explanation as genuinely explanatory). In the second half of this paper, Studies 3 and 4 assess whether offering an ethical explanation is taken as evidence of the meta-ethical beliefs of the explanation-provider, as our analysis would predict. This is valuable not only as additional evidence for a link between ethical explanations and beliefs about moral progress and moral principles, but as a step towards understanding the communicative role of explanations: beyond their surface content, what is conveyed by the fact that an individual deems a claim explanatory?

3.1. Study 1: Meta-ethical commitments and explanations for social change

Study 1 had two aims. The first aim was to test the hypothesis that participants accept ethical explanations as partially or fully explanatory. To do so, we asked participants to consider why some social change occurred or might occur, such as women gaining the right to vote or the death penalty being abolished. We provided four possible explanations for this change, including one *ethical* explanation (which cited the morality of the practice), one *ethical belief* explanation (which cited changes in the population’s moral beliefs as responsible for the change), one *non-ethical* explanation (citing economic or social factors), and one *poor* explanation (which essentially described rather than explained the change; see Table 1). If participants accept ethical explanations as at least somewhat explanatory, they should rate the quality of ethical explanations reliably higher than poor explanations.

The second aim of Study 1 was to test the hypothesis that participants who endorse moral progress and moral principles are more likely to find ethical explanations explanatory. To test this, we also asked participants about their beliefs in moral progress and moral principles. We predicted that stronger endorsement of moral progress and moral principles would be associated with higher ratings for ethical explanations. Moreover, if these moral belief ratings do not predict *ethical belief* explanation ratings, this may provide some evidence that ethical explanations are not treated as equivalent to ethical belief explanations, but instead judged explanatory for other reasons.

4. Method

All studies were approved by the IRB at Princeton University (#10662). All studies were pre-registered, and all pre-registrations, data, research materials are available at <https://osf.io/f9n3q/>. Data were analyzed using R, version 4.1.3 (R Core Team, 2022).

4.1. Participants

Participants in Study 1 were 220 adults recruited via Prolific (93

Table 1
Explanation Types.

Explanation type	Explanation text
Ethical	Because it is morally wrong to not allow women to vote.
Ethical belief	Because people came to believe it was morally wrong to not allow women to vote.
Non-ethical	Because the United States entered World War I, which allowed more women to take over men’s roles in the workforce and have a public voice.
Poor	Because it became legal for women to have the right to vote.

Note. The four explanations shown to participants who were asked, “Why were women given the right to vote in the United States?”

women, 119 men, 7 non-binary people, 1 genderfluid person; mean age 35 years, age range 18–77). Five additional respondents were excluded for failing an attention check (described below). Participants were paid at a rate of \$7.50 per hour, pro-rated to our 8-min task. Participation in all studies was restricted to workers in the U.S. who had completed at least 100 prior tasks with a 95% approval rating.

4.2. Materials and Procedure

All participants completed two tasks, an Explanation Rating Task and a Moral Beliefs Task, the order of which was counterbalanced.

4.3. Explanation Rating Task

In this task, participants were randomly assigned to consider one of four social changes: the abolition of slavery, women gaining the right to vote, the legalization of same-sex marriage, or the abolition of the death penalty. They were asked to consider why this social change occurred or might occur in the United States (e.g., “Why were women given the right to vote in the United States?”).

Participants rated the quality of four possible explanations for the social change, presented in a random order. One explanation was an *ethical* explanation, citing just the ethics of the practice (“Because it is morally wrong to not allow women to vote.”). Another was an *ethical belief* explanation, citing a change in people’s beliefs about the ethics of the practice (“Because people came to believe it was morally wrong to not allow women to vote.”). Another was a *non-ethical* explanation, citing non-ethical facts (e.g., “Because the United States entered World War I, which allowed more women to take over men’s roles in the workforce and have a public voice.”). And finally, we included a *poor*, or circular, explanation (“Because it became legal for women to have the right to vote.”). Participants were asked, “How good is this explanation?”, and rated each answer on a scale from “1-Poor explanation” to “7-Good explanation” with a midpoint at “4-Average Explanation.”

4.4. Moral Beliefs Task

In this task, participants answered three questions about their personal moral beliefs, presented in a random order.

For the moral progress question, participants rated the extent to which they agree that moral progress occurs (“Do you think that people will necessarily advance morally, ethically, and socially, or decline?”; adapted from Rutjens et al., 2016) on a sliding scale from –10 (“Decline”) to 10 (“Advance”).

For the moral principles disagreement question (adapted from Sarkissian et al., 2011), participants read a vignette in which a person similar to themselves disagrees with an imagined friend of the participant about whether slavery, denying women the ability to vote, denying same-sex couples the right to marry, or the death penalty (depending on their social change condition) is morally wrong. For example, when the social change was women’s right to vote, participants read:

Imagine a person named Allison, a fairly ordinary student from your town who enjoys watching sports and hanging out with friends. Consider Allison’s views concerning the moral status of the following: denying women the right to vote.

Allison thinks that denying women the right to vote is not morally wrong.

Imagine that one of your friends thinks that denying women the right to vote is morally wrong. Given that these individuals (Allison and your friend) 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 friend and Allison have different judgments about this

case, at least one of them must be wrong.”

Participants then provided a rating on a scale from 1 (“definitely disagree”) to 7 (“definitely agree”) with a midpoint at 4 (“neither agree nor disagree”).

In the moral principles truth-aptness question (adapted from Goodwin & Darley, 2008), participants rated whether they believe the statement “[Slavery/Denying women the ability to vote/Denying same-sex couples the ability to marry/The death penalty] is morally wrong” is “true,” “false,” or “an opinion.”

Finally, participants answered demographic questions before being debriefed and exiting the study. In all studies, we included an attention check at the beginning of the survey in which participants read distractor text about fruit preferences, then were asked to write the word “instructions” in a text box.

5. Results

First, to test the hypothesis that participants will accept ethical explanations as partially or fully explanatory (i.e., endorsed more strongly than poor explanations), we performed a mixed ANOVA with explanation ratings as the dependent variable, explanation type (ethical, ethical belief, non-ethical, poor) as a within-subjects factor, and social change (slavery, suffrage, marriage, death penalty) as a between-subjects factor (see Fig. 2). We found a main effect of explanation type, $F(3,863) = 67.77, p < .001$, such that ethical explanations ($M = 3.90, SD = 1.88$) were indeed judged as significantly better than poor explanations ($M = 2.69, SD = 1.90; t(437) = 6.70, p < .001$). Additionally, non-ethical explanations were judged as significantly better than all other explanation types ($ps < 0.001$) and poor explanations were judged as worse than all other types ($ps < 0.001$). The ANOVA also revealed a significant interaction, $F(9, 863) = 6.04, p < .001$, reflecting variation across social change vignettes.

Next, we tested whether participants who more strongly endorse moral progress and moral principles are more likely to endorse ethical explanations. To test this, we ran three independent regressions, each treating ethical explanation rating as the dependent variable and one of the following as a predictor: moral progress rating, moral principles disagreement rating, or moral principles truth-aptness rating (with “true” and “false” collapsed to compare against “opinion”; see Fig. 3 and Table 2). We used a centered score for each of these predictors. Each regression also included social change as a predictor. We then compared each model to a reduced model that excludes the relevant predictor, using likelihood ratio tests. We report the variance explained by each model in OSF. Our hypothesis predicts that moral progress and at least one moral principles measure will be positive and significant predictors of ethical explanation score. In three of the final models, participants judged ethical explanations as significantly better when they endorsed

the corresponding meta-ethical beliefs (moral principles–truth-aptness, was marginally significant; see Table 2).

We also repeated the three independent regressions and likelihood ratio tests just reported using ethical belief explanation ratings rather than ethical explanation ratings as the dependent variable. In the final models, participants did *not* judge ethical belief explanations as significantly better when they endorsed meta-ethical beliefs. This suggests that participants distinguish ethical explanations from ethical belief explanations, although we are cautious to avoid strong claims from null findings. To further explore this difference between ethical and ethical belief explanations, we ran additional regressions predicting explanation ratings from each measure for both kinds of explanations; these analyses revealed significant interactions between explanation type and each predictor (moral progress: $p = .03$; moral principles–disagreement, $p = .04$; moral principles–truth-aptness, $p = .02$), indicating that moral beliefs were significantly stronger predictors of ethical explanation ratings than of ethical belief explanation ratings.

6. Discussion

The results from Study 1 support our first hypothesis that participants are willing to accept ethical explanations – explanations that simply appeal to the morality of a practice – as explanations for social change in the sense that such explanations are rated more explanatory than poor (circular) explanations. Study 1 also supports our second hypothesis: endorsing moral progress and moral principles was associated with higher ratings for ethical explanations.

Notably, our findings challenge the plausible idea that insofar as participants find ethical explanations explanatory, this is because they treat them as shorthand for ethical belief explanations. Ethical belief explanations were indeed rated similarly to ethical explanations, but they were not predicted by either moral progress or moral principles beliefs. Thus, it appears that at least some people consider ethical explanations to be genuinely explanatory, and that they do so for reasons related to their beliefs in a world that is driven to conform to objective moral truths.

6.1. Study 2: Conceptual replication of Study 1

Our aim in Study 2 was to conceptually replicate Study 1 using different stimuli, to assess whether our results generalize to other ways of measuring explanation ratings, moral progress, and moral principles. Specifically, rather than measuring explanatory goodness, we used measures adapted from Sulik, van Paridon, and Lupyán (2023) to measure explanatory satisfaction, which is related to curiosity and perceived learning (Liquin & Lombrozo, 2020, 2022; Lombrozo & Liquin, 2023). For moral progress, we asked about participants’ confidence

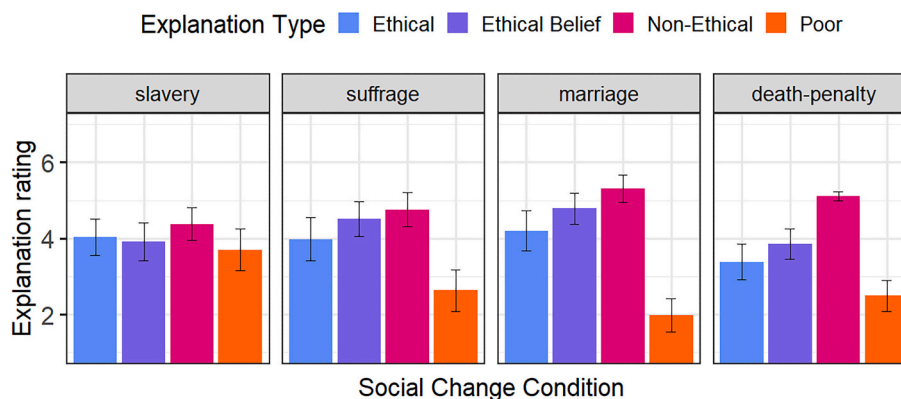


Fig. 2. Explanation Ratings across Social Change Conditions and Explanation Types.

Note. This graph depicts mean ratings of the quality of ethical, ethical belief, non-ethical, and poor explanations across social change conditions. Error bars indicate 95% CI.

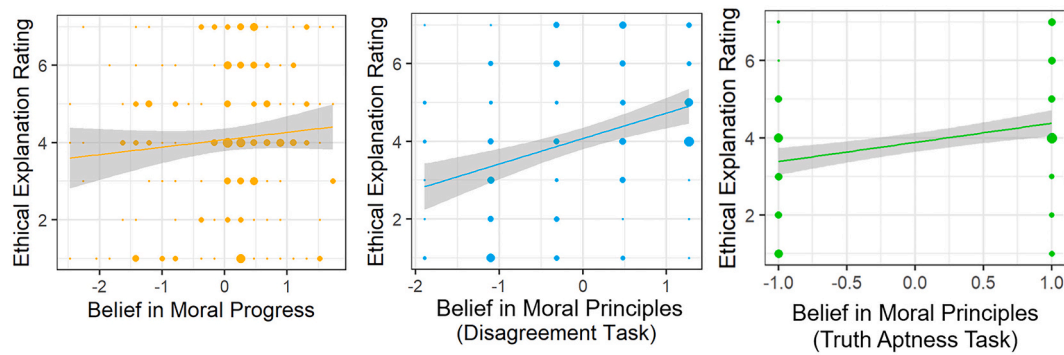


Fig. 3. Relationships between Explanation Rating and Meta-ethical Beliefs.

Note. This graph depicts the relationship between individual-level ethical explanation ratings and belief ratings with the best fit line from Pearson regression analysis (for the truth aptness task, -1 corresponds to ‘an opinion’ and 1 corresponds to ‘true’ or ‘false’). Error bars indicate standard errors and point size indicates the relative number of participants for each response value.

Table 2
Effects of Metaethical Beliefs on Ethical and Ethical Belief Explanations.

Dependent variable	Metaethical belief predictor	β	p-value	Adjusted R ²
Ethical explanations	Moral progress	0.06	0.02	0.02
	Moral principles–disagreement	0.29	0.004	0.02
	Moral principles–truth-aptness ^b	0.53	0.059	0.07
Ethical belief explanations	Moral progress ^a	-0.02	0.43	0.04
	Moral principles–disagreement ^a	0.10	0.25	0.05
	Moral principles–truth-aptness ^a	0.18	0.11	0.04

Note. This table reports the β and p-value for the effect of each metaethical belief predictor on ethical explanations and ethical belief explanations. The values are taken from the final model retained after model comparisons.

^a Final model retained metaethical belief and social change as predictors.

^b Final model retained metaethical belief, social change, and their interaction as predictors.

that moral progress is occurring, using questions modeled on Study 2 from Lewry et al. (2023). Finally, as noted in the introduction, there has been debate among philosophers and psychologists about the best way to measure folk moral objectivism (Hopster, 2019; Pözlzer & Wright, 2020; Sarkissian et al., 2011; Zijlstra, 2023). Zijlstra (2019) proposes that folk moral objectivism may be multidimensional, and that typical disagreement and truth-aptness judgments are inadequate to capture these dimensions. For this reason, in Study 2 we used Zijlstra’s (2019) measure which examines three varieties of belief in moral truths: universal truth, independent truth, and divine truth, described in detail below. Given that Study 2 was designed as a conceptual replication of Study 1, our predictions mirrored those for Study 1.

7. Method

7.1. Participants

Participants in Study 2 were 293 adults recruited via Prolific (142 women, 146 men, 5 non-binary people; mean age 36 years, age range 18–78). Seven additional respondents were excluded for failing an attention check (described below). Participants were paid at a rate of \$12.50 per hour, pro-rated to our 6-min task. Participation in all studies was restricted to workers in the U.S. who had completed at least 100 prior tasks with a 95% approval rating.

7.2. Materials and Procedure

The methods and procedure followed those of Study 1 closely, with the following modifications.

First, explanation ratings were elicited differently in the Explanation Rating Task. Following Sulik et al. (2023), participants were instructed to rate how satisfying they find each explanation to be. They were told the following: “Please rate this explanation on how satisfying you think it is. The answer could be true or accurate, but still be unsatisfying. For instance, if someone explains why deer have antlers by simply saying ‘Evolution’, then this answer is correct, but it wouldn’t satisfy someone who wonders why they evolved that way. So try to think about how appealing you think the answer is as a whole, not just whether it is true.” Finally, they rated each answer on a scale from “1-Very unsatisfying” to “7-Very satisfying” with a midpoint at “4-Neither satisfying nor unsatisfying.”

Second, in the Moral Beliefs Task, participants answered two sets of questions about their personal moral beliefs – one about their belief in moral progress and one about their belief in moral principles, the order of which was randomized across participants.

For the moral progress question, participants rated the extent to which they agree that moral progress occurs. Participants were asked the following: “We would like you to think about the progress that has been made over the course of history in terms of humans advancing morally and ethically. Many people would agree that we live in a world that is, overall, morally better today than it was hundreds of years ago, and that this moral progress will continue into the future. In general, do you think that moral progress has occurred and will continue to occur in the future?” (adapted from Lewry et al., 2023). Participants could choose “yes” or “no.” If they chose “yes,” they were asked to give a confidence rating on a scale from “1-Not at all confident” to “6-Very confident.”

For the moral principles question set (adapted from Zijlstra, 2019), participants rated their agreement with twelve statements presented in a random order (see OSF for all items). Four statements reflected Universal Truth (the idea that there are moral facts that are true no matter what; e.g., “Certain actions are morally wrong and they remain morally wrong even in the rare case that no one believes so”), four reflected Independent Truth (the idea that there are moral facts independent of cultural variation; e.g., “All ideas about what is morally right and morally wrong are products of individuals, cultures, and communities and nothing more”, reverse scored), and four reflected Divine Truth (the idea that there are moral facts that depend on the existence of a divine entity; e.g., “The only actions that are ultimately morally right or wrong are those actions that God prescribes”). Each question was answered on a 6-point scale from “Strongly disagree” to “Strongly agree.”

8. Results

First, to test the hypothesis that participants will accept ethical explanations as partially or fully explanatory (i.e., endorsed more strongly than poor explanations), we performed a mixed ANOVA with explanation ratings as the dependent variable, explanation type (ethical, ethical belief, non-ethical, poor) as a within-subjects factor, and social change (slavery, suffrage, marriage, death penalty) as a between-subjects factor (see Fig. 4). We found a main effect of explanation type, $F(3,987) = 112.66, p < .001$, such that ethical explanations ($M = 4.31, SD = 1.74$) were indeed judged as significantly more satisfying than poor explanations ($M = 2.38, SD = 1.73; t(578) = 10.63, p < .001$). Additionally, non-ethical explanations were judged as significantly more satisfying than all other explanation types ($ps < 0.001$) and poor explanations were judged as worse than all other types ($ps < 0.001$). The ANOVA also revealed a significant interaction, $F(9, 313) = 11.89, p < .001$, reflecting variation across social change vignettes.

Next, we tested whether participants who more strongly endorse moral principles and moral progress are more likely to endorse ethical explanations. To test this, we ran four independent regressions, each treating ethical explanation rating as the dependent variable and one of the following as a predictor: moral progress rating (with “no” converted to a score of 1, and “yes” confidence ratings converted to a score of 2–7), moral principles—universal truth rating, moral principles—-independent truth rating, or moral principles—divine truth rating (see Fig. 5 and Table 3). We used a centered score for each of these predictors. Each regression also included social change as a predictor, dummy coded and with the death penalty vignette as a reference group. We then compared each model to a reduced model that excludes the relevant predictor, using likelihood ratio tests. We report the variance explained by each model in OSF. Our hypothesis predicts that moral progress and at least one moral principles measure will be positive and significant predictors of ethical explanation score. In each of the three final models, participants judged ethical explanations as significantly better when they endorsed meta-ethical beliefs (with the exception of moral principles— independent truth, which was not significant; see Table 2).

We also repeated the four independent regressions and likelihood ratio tests just reported using ethical belief explanation ratings rather than ethical explanation ratings as the dependent variable. In the final models, participants did *not* judge ethical belief explanations as significantly better when they endorsed meta-ethical beliefs. As in Study 1, this suggests that participants distinguish ethical explanations from ethical belief explanations. To further explore this difference between ethical and ethical belief explanations, we ran additional regressions predicting explanation ratings from each measure for both kinds of explanations; these analyses revealed significant interactions between explanation type and each predictor except moral

principles— independent truth (moral progress: $p = .04$; moral principles—universal truth, $p = .01$; moral principles— independent truth, $p = .31$; moral principles—divine truth, $p = .02$). This indicates that for moral progress and two measures of moral principles, the association between moral beliefs and explanation ratings was significantly stronger for ethical explanations than for ethical belief explanations.

9. Discussion

The results from Study 2 lend further support to our hypotheses from Study 1, suggesting that some participants find ethical explanations satisfying and that endorsing moral progress and moral principles are associated with higher ratings for ethical explanations. These results expand on Study 1 by showing that these results replicate using different measures of explanation ratings, moral progress, and moral principles. Future work could explore why the universal truth and divine truth, but not independent truth, dimensions of principles are associated with ethical explanation endorsement.

9.1. Study 3: Using ethical explanations to infer moral commitments

Studies 1 and 2 found a reliable relationship between an individual’s beliefs about moral progress and principles and their willingness to endorse an ethical explanation. In Study 3, we test whether participants recognize this relationship; that is, whether participants use the kinds of explanations that others provide to infer their moral beliefs. Specifically, we hypothesized that hearing an ethical explanation for why social change occurs (versus a non-ethical explanation that appeals to non-ethical factors) makes participants more likely to infer that the individual who offered the explanation believes in moral progress and moral principles.

To test this, we introduced participants to a character who expresses an opinion on some potential social change, such as the banning of handgun ownership. Across participants, the character either provides an ethical explanation (“handgun ownership will eventually be made illegal because it is morally wrong to own handguns”) or a non-ethical explanation (“[...] because social and economic pressures will lead to legislative changes that make it illegal”). Next, we explain the concepts of moral progress and moral principles and ask participants the extent to which they think the character holds each of these moral beliefs.

If people use the explanations that others provide to infer their moral beliefs, specifically reflecting the relationship between ethical explanation endorsement and moral progress and moral principles beliefs found in Studies 1 and 2, then participants should rate the character higher on moral principles and moral progress when the character provides an ethical explanation (vs. a non-ethical explanation).

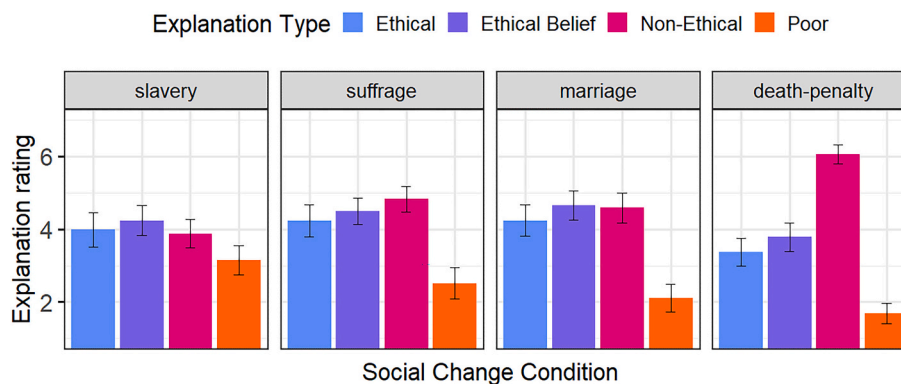


Fig. 4. Explanation Ratings across Social Change Conditions and Explanation Types.

Note. This graph depicts mean satisfaction ratings of ethical, ethical belief, non-ethical, and poor explanations across social change conditions. Error bars indicate 95% CI.

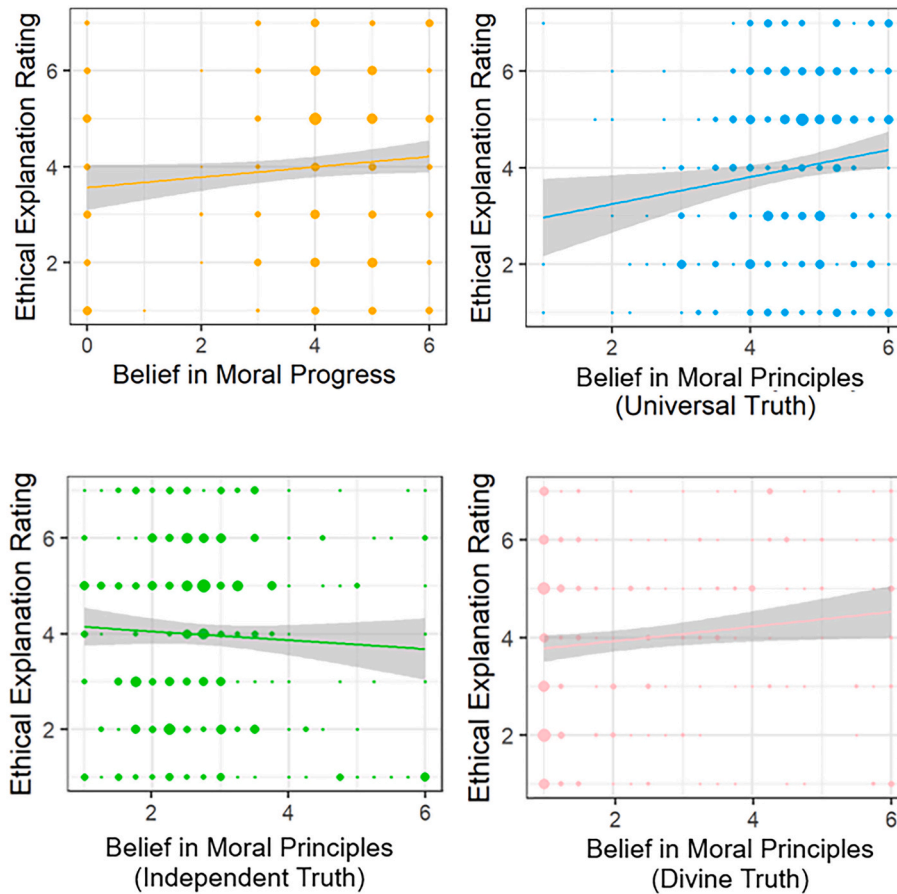


Fig. 5. Relationships between Explanation Rating and Meta-ethical Beliefs.

Note. This graph depicts the relationships between ethical explanation ratings and belief ratings with the best fit line from Pearson regression analyses. Error bars indicate standard errors and point size indicates the relative number of participants for each response value.

Table 3
Effects of Metaethical Beliefs on Ethical and Ethical Belief Explanations.

Dependent variable	Metaethical belief predictor	β	p -value	Adjusted R^2
Ethical explanations	Moral progress ^a	0.11	0.04	0.03
	Moral principles–universal truth ^a	0.28	0.01	0.04
	Moral principles–independent truth ^a	-0.08	0.39	0.03
	Moral principles–divine truth ^b	0.32	0.01	0.10
Ethical belief explanations	Moral progress ^b	0.06	0.55	0.05
	Moral principles–universal truth ^b	0.09	0.72	0.04
	Moral principles–independent truth ^b	-0.20	0.24	0.05
	Moral principles–divine truth ^b	0.20	0.09	0.09

Note. This table reports the β and p -value for the effect of each metaethical belief predictor on ethical explanations and ethical belief explanations. The values are taken from the final model retained after model comparisons.

^a Final model retained metaethical belief and social change as predictors.

^b Final model retained metaethical belief, social change, and their interaction as predictors.

10. Method

10.1. Participants

Participants in Study 3 were 513 adults recruited via Prolific (252 women, 254 men, 7 non-binary people; mean age 37 years, age range 18–85). Twelve additional respondents were excluded for failing an attention check (described below). Participants were paid at a rate of \$7.50 per hour, pro-rated to our 5-min task.

10.2. Materials and Procedure

Participants were randomly assigned to one of two explanation type conditions: ethical or non-ethical. Participants were also randomly assigned to read about one of five potential social changes: the legalization of abortion, the abolishment of the death penalty, the banning of handgun ownership, the legalization of marijuana, or the legalization of physician-assisted suicide. These social changes were chosen because approximately 40–60% of Americans support each change according to Pew Research Center as of 2021 (Pew Research Center, 2015, 2021).

First, participants were introduced to a character who provides an explanation for why a social change might occur. In the ethical explanation condition, the character explains that the social change will occur because the practice is morally wrong. For example, for the abolishment of the death penalty, participants read the following text, with the manipulated ethical/non-ethical explanation text in brackets:

One day, you meet someone from your town named James. You and James begin discussing the topic of the death penalty, and he tells

you that he thinks the death penalty will eventually be made illegal because [it is morally wrong to enact the death penalty/ social and economic pressures will lead to legislative changes that make it illegal].

Character names/pronouns were counterbalanced. Participants were then asked to summarize the character's view to ensure they read it carefully.

After completing this task, participants completed a moral progress inference question and a moral principles inference question, in random order. For the moral progress inference questions, participants read:

Moral progress is the idea that over time, humans advance morally. The world we live in today is morally better than it was in the past, and the world will be morally better in the future.

To what extent do you think [character] believes that moral progress occurs?

Participants responded on a scale from 1 (“[character] definitely does not believe that moral progress occurs”) to 5 (“[character] definitely believes that moral progress occurs”). For the moral principles inference question, participants read:

Moral objectivism is the idea that there are some moral facts that are true, no matter where you are from in the world. Regardless of the culture you were raised in, some things are just morally right or wrong.

To what extent do you think [character] believes that moral objectivism is true?

Participants responded on a scale from 1 (“[character] definitely does not believe that moral objectivism exists”) to 5 (“[character] definitely believes that moral objectivism exists”).

We also asked participants to judge the character's moral position on the social change on a scale from 1 (e.g., “[character] definitely believes that the death penalty is morally wrong”) to 5 (“[character] definitely believes that the death penalty is morally right”). Finally, we asked participants to rate the character on a variety of traits (not reported here; see OSF Materials), and to provide their personal views on this topic following a similar scale (1-“I strongly believe the death penalty is morally wrong” to 5-“I strongly believe the death penalty is morally right”).

11. Results

To determine whether inferences about a character's moral beliefs differ depending on whether the character provides an ethical or non-

ethical explanation for why social change might occur, we performed two between-subjects ANOVAs (see Fig. 6).

The first ANOVA had moral progress inference ratings as the dependent variable and explanation type (ethical, non-ethical) and social change (abortion, death penalty, gun ownership, marijuana, physician-assisted suicide) as between-subjects factors. We found a main effect of explanation type, $F(1,502) = 9.54, p = .002$, such that participants rated the character's belief in moral progress higher when the character gave an ethical explanation ($M = 3.86, SD = 0.96$) than a non-ethical explanation ($M = 3.61, SD = 0.99; t = 2.94, p = .003$). We also found a main effect of social change, $F(4,502) = 14.15, p < .001$, such that moral progress inference ratings were lowest in the abortion vignette ($ps < 0.001$) and highest in the death penalty vignette ($ps < 0.05$). We found no significant interaction.

The second ANOVA had moral principles inference ratings as the dependent variable and explanation type and social change as between-subjects factors. We found a main effect of explanation type, $F(1,503) = 12.03, p < .001$, such that participants rated the character's belief in moral principles higher when the character gave an ethical explanation ($M = 3.85, SD = 0.91$) than a non-ethical explanation ($M = 3.57, SD = 0.95; t = 3.45, p < .001$). We also found a main effect of social change, $F(4,503) = 2.78, p = .03$, such that moral principles inference ratings were higher in the abortion and death penalty vignettes than in the marijuana vignette ($ps < 0.05$), and higher in the abortion vignette than in the physician-assisted suicide vignette ($p = .04$). We found no significant interaction.

We also considered whether these results were moderated by participants' own views on the morality of the social practice under consideration. None of them were: moral progress, $F(1,499) = 0.07, p = .79$, moral principles, $F(1,500) = 3.00, p = .08$.

Finally, we performed a between-subjects ANOVA to determine whether providing an ethical or non-ethical explanation affects the extent to which participants infer that the character believes the practice is morally good or bad. The ANOVA had character's moral position inference as a dependent variable and explanation type and social change as between-subjects factors. Unsurprisingly, we found a main effect of explanation type, $F(1, 502) = 206.18, p < .001$, such that participants were more likely to rate the character as morally supportive of the change (e.g., morally opposed to the death penalty) when the character gave an ethical explanation ($M = 1.30, SD = 0.71$) than a non-ethical explanation ($M = 2.34, SD = 0.92; t = -14.30, p < .001$).

12. Discussion

Study 3 built on Studies 1 and 2 by showing that not only is there a

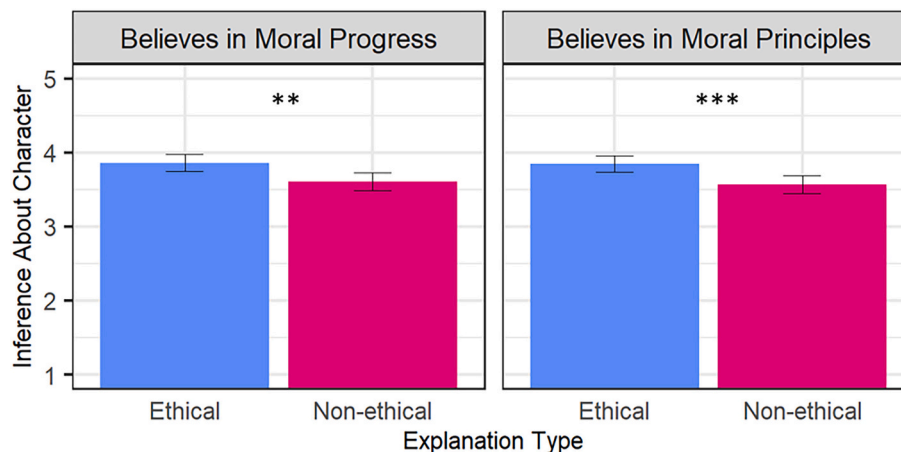


Fig. 6. Inferences about Character's Meta-ethical Commitments across Explanation Types.

Note. This graph depicts mean inference ratings about a character's belief in moral progress and moral principles when the character provides an ethical or non-ethical explanation for a social change. Error bars indicate 95%-CI. $**p < .01$, $***p < .001$.

relationship between endorsement of ethical explanations and beliefs in moral progress and moral principles, but that people recognize this relationship and use individuals' ethical explanations to infer their moral beliefs. For example, if someone explains that "handgun ownership will be made illegal because owning a handgun is morally wrong," others are more likely to infer that this person believes that the world will morally improve and that there are objective moral truths.

One limitation of this study is that conclusions rest on differences between the ethical explanation and non-ethical explanation conditions, so it is unclear whether ethical explanations elevate inferences about moral progress and moral principles from their default values, versus the alternative or additional possibility that non-ethical explanations depress such inferences. In Study 4, we disentangle these possibilities.

12.1. Study 4: The role of ethical and explanatory content in inferring moral commitments

Study 3 showed that participants attribute moral progress and moral principles beliefs to others differently depending on whether they provide an ethical versus non-ethical explanation for social change. In Study 4, we examine these belief inferences across different types of contrasts: when someone provides an ethical explanation versus ethical claim (i.e., no explanatory language), an ethical explanation versus neutral explanation (i.e., an explanation with no specified content), or an ethical explanation versus non-ethical explanation. We hypothesized that hearing an ethical explanation for why social change occurs makes people more likely to infer that the individual who offered the explanation believes in moral progress and moral principles, and that this holds against all contrast types. As such, both the ethical and explanatory content of an ethical explanation are necessary for people to infer others' beliefs in moral progress and moral principles.

To test this, we introduced participants to two characters who each expressed an opinion on one of the five social changes in Study 3. One character always provides an ethical explanation ("handgun ownership will eventually be made illegal because it is morally wrong to own handguns"). Across participants, the other character provides either an ethical claim ("[Amy/James] says that they think it is morally wrong to own a handgun"), a neutral explanation ("[Amy/James] explains why they think handgun ownership will eventually be made illegal"), or a non-ethical explanation ("[Amy/James] explains that they think handgun ownership will eventually be made illegal because social and economic pressures will lead to legislative changes that make it illegal"). As in Study 3, we then explain moral progress and moral principles and ask participants the extent to which they think the character holds each of these moral beliefs.

If both the ethical content of an ethical explanation, *and* the fact the ethical content plays an explanatory role, contribute to the inference that the explanation-provider believes in moral progress and moral principles, then participants should give higher progress and moral principles ratings to the characters who give ethical explanations than the characters who offer non-ethical explanations, ethical claims, or neutral explanations.

Study 4 thus goes beyond Study 3 in two important ways. First, by looking at the contrast between ethical explanations and neutral explanations, the study can address the question of whether ethical explanations elevate inferences concerning moral progress and moral principles, versus the alternative that non-ethical explanations merely depress them. Second, by looking at the contrast between ethical explanations and ethical claims, the study can address whether the *explanatory role* of the ethical content of the explanation – and not just the ethical content it presupposes – plays a role in driving inferences concerning moral progress or moral principles.

13. Method

13.1. Participants

Participants in Study 4 were 328 adults recruited via Prolific (158 women, 162 men, 8 non-binary people; mean age 38 years, age range 18–78). Two additional respondents were excluded for failing an attention check (described below). Participants were paid at a rate of \$7.50 per hour, pro-rated to our 5-min task.

13.2. Materials and Procedures

Participants were randomly assigned to one of three contrast conditions. Participants were also randomly assigned to read about one of the five potential social changes from Study 3.

First, all participants were introduced to two characters. For example, "One day, you meet two people from your town named Amy and James. The three of you begin discussing the topic of the death penalty." Amy and James then each made a statement (character names were counterbalanced). One of the statements was always an ethical explanation (e.g., "the death penalty will eventually be made illegal because it is morally wrong to enact the death penalty"). Depending on the contrast condition, the other statement was either an ethical claim (i.e., a statement with ethical content but no explanatory language; "[Amy/James] tells you that they think it is morally wrong to enact the death penalty"), a neutral explanation (i.e., an explanation with no specific content; "[Amy/James] explains why they think the death penalty will eventually be made illegal"), or a non-ethical explanation (i.e., an explanation citing non-ethical factors; "[Amy/James] tells you that they think the death penalty will eventually be made illegal because social and economic pressures will lead to legislative changes that make it illegal"). The order of the contrast statements was counterbalanced. Participants were then asked to summarize the character's view to ensure they read it carefully.

Second, participants completed the moral progress and moral principles measures from Study 3. They also completed the trait inference measures from Study 3; these are reported in OSF Materials. The measures were presented one at a time, below the original character statement text so participants could refer back to each of the characters' statements. Participants were asked to give moral progress, moral principles, and trait inference ratings for each of the characters. The order of the progress and moral principles measures was counterbalanced, followed by the trait inference measure. Finally, we asked participants to judge each of the characters' moral positions on the social practice and provide their own views on the topic, as in Study 3.

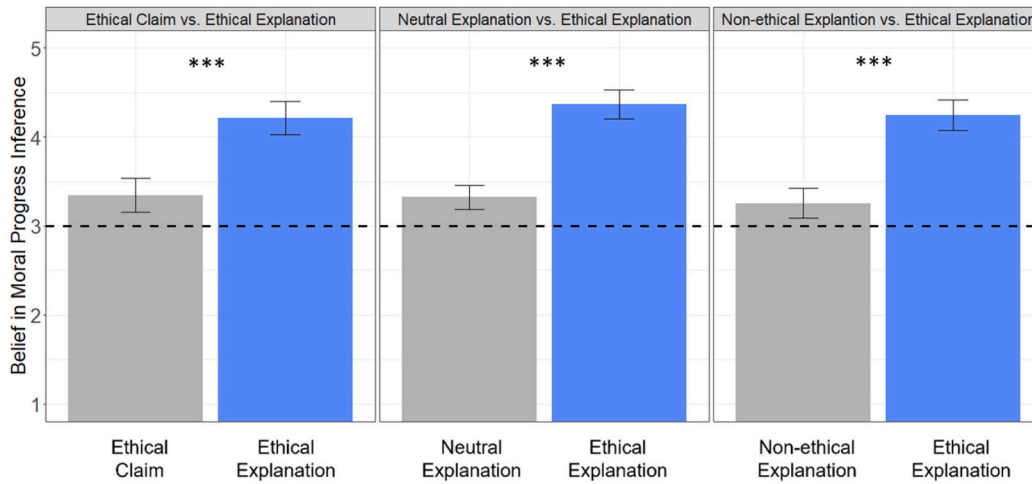
14. Results

To determine whether ethical explanations lead to stronger inferences about someone's moral progress and moral principles beliefs than non-ethical explanations, ethical claims, or neutral explanations, we performed six mixed ANOVAs.³ That is, for each contrast type (ethical claim, neutral explanation, non-ethical explanation), we performed one analysis to compare differences in moral progress ratings, and one analysis to compare moral principles ratings (see Figure 7).

First, we compared moral belief inferences in the ethical explanation versus ethical claim contrast. The first ANOVA had moral progress inference ratings as the dependent variable, statement type (ethical explanation, ethical claim) as a within-subjects variable, and social change as a between-subjects variable. We found a main effect of statement type, $F(1,103) = 70.59, p < .001$, such that participants rated the character's belief in moral progress higher when the character

³ To minimize the risk of a type II error, we preregistered the predicted direction of the effects.

(A) Inferences about a Character's Belief in Moral Progress across Explanation Types



(B) Inferences about a Character's Belief in Moral Principles across Explanation Types

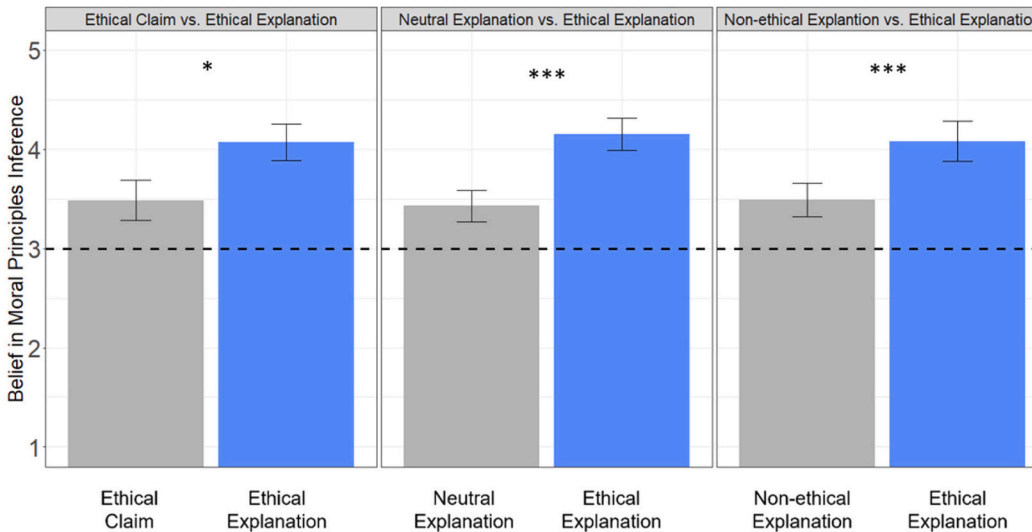


Fig. 7. (A) Inferences about a Character's Belief in Moral Progress across Explanation Types. (B) Inferences about a Character's Belief in Moral Principles across Explanation Types.

Note. These graphs depict mean inferences about a character's belief in moral progress (A) and moral principles (B) for a character who provides an ethical explanation versus a character who provides an ethical claim, a neutral explanation, or a non-ethical explanation. Error bars indicate 95%-CI. * $p < .05$, *** $p < .001$.

offered an ethical explanation ($M = 4.17, SD = 1.02$) than an ethical claim ($M = 3.33, SD = 1.07; t = 8.21, p < .001$). We also found a main effect of social change, $F(4,103) = 5.94, p < .001$, such that moral progress inference ratings were lowest in the abortion vignette ($ps < 0.05$). We found no significant interaction.

The second ANOVA had moral principles inference ratings as the dependent variable, statement type (ethical explanation, ethical claim) as a within-subjects variable, and social change as a between-subjects variable. We found a main effect of statement type, $F(1,102) = 25.20, p < .001$, such that participants rated the character's belief in moral principles higher when the character offered an ethical explanation ($M = 4.07, SD = 0.96$) than an ethical claim ($M = 3.49, SD = 1.05; t = 4.95, p < .001$). We found no significant effect of social change and no significant interaction.

Second, we compared moral belief inferences in the ethical

explanation versus neutral explanation contrast. The first ANOVA had moral progress inference ratings as the dependent variable, statement type (ethical explanation, neutral explanation) as a within-subjects variable, and social change as a between-subjects variable. We found a main effect of statement type, $F(1,106) = 127.47, p < .001$, such that participants rated the character's belief in moral progress higher when the character gave an ethical explanation ($M = 4.27, SD = 0.92$) than a neutral explanation ($M = 3.31, SD = 0.88; t = 9.75, p < .001$). We found no main effect of social change and no significant interaction.

The second ANOVA had moral principles inference ratings as the dependent variable, statement type (ethical explanation, neutral explanation) as a within-subjects variable, and social change as a between-subjects variable. We found a main effect of statement type, $F(1,106) = 41.57, p < .001$, such that participants rated the character's belief in moral principles higher when the character gave an ethical

explanation ($M = 4.14$, $SD = 0.91$) than a neutral explanation ($M = 3.49$, $SD = 0.74$; $t = 6.66$, $p < .001$). We found no main effect of social change and no significant interaction.

Finally, we compared moral belief inferences in the ethical explanation versus non-ethical explanation contrast. The first ANOVA had moral progress inference ratings as the dependent variable, statement type (ethical explanation, non-ethical explanation) as a within-subjects variable, and social change as a between-subjects variable. Consistent with the results of Study 2, we found a main effect of statement type, $F(1,103) = 64.42$, $p < .001$, such that participants rated the character's belief in moral progress higher when the character gave an ethical explanation ($M = 4.32$, $SD = 0.97$) than a non-ethical explanation ($M = 3.21$, $SD = 0.86$; $t = 8.39$, $p < .001$). We also found a main effect of social change, $F(4,103) = 5.37$, $p < .001$, such that moral progress inference ratings were lowest in the abortion vignette ($ps < 0.05$), and death penalty led to significantly higher ratings than handgun ownership ($p < .05$). We found no significant interaction.

The second ANOVA had moral principles inference ratings as the dependent variable, statement type (ethical explanation, non-ethical explanation) as a within-subjects variable, and social change as a between-subjects variable. Again consistent with Study 3, we found a main effect of statement type, $F(1,102) = 54.91$, $p < .001$, such that participants rated the character's belief in moral principles higher when the character gave an ethical explanation ($M = 4.15$, $SD = 0.98$) than a non-ethical explanation ($M = 3.15$, $SD = 0.83$; $t = 8.13$, $p < .001$). We found no main effect of social change and no significant interaction.

None of the reported results were moderated by participants' own views on the morality of the social practice under consideration.

15. Discussion

Study 4 supported the findings of Study 3 by showing that participants use others' ethical explanations to infer their moral commitments. As in Study 3, but using a within-subjects rather than between-subjects design, we found that participants attribute stronger beliefs in moral progress and moral principles to those who provide an ethical explanation, rather than a non-ethical explanation, for social change. Moreover, we expanded beyond Study 3 by showing that these attributions are stronger for ethical explanations than two other types of statements: ethical claims and neutral explanations.

As such, both the ethical and explanatory content of a statement such as, "the death penalty will eventually be made illegal because it is morally wrong to enact the death penalty," play important roles in driving inferences about others' moral commitments. A mere ethical claim without reference to explanation, a mere explanation without ethical content, or an explanation explicitly citing non-ethical factors all led to weaker belief inferences than did ethical explanations.

15.1. General Discussion

"Women were given the right to vote in the United States because it is morally wrong to deny women the right to vote." In this paper, we asked whether participants judge such ethical explanations genuinely explanatory, whether the explanations' explanatory power relates to belief in moral progress and moral principles, and whether participants recognize this relationship and use it to make inferences about others' moral commitments.

In Studies 1 and 2, we compared participants' endorsement of ethical explanations for social change to their endorsement of ethical belief, non-ethical, and poor explanations for the same event. Participants judged ethical explanations as better and more satisfying than poor explanations, providing evidence that they are deemed at least somewhat explanatory. Additionally, we found that participants who reported stronger beliefs in moral progress and moral principles were more likely to endorse ethical explanations, but not more likely to endorse ethical belief explanations. This provides evidence that

perceiving moral truths as guiding and objective (that is, like a causal force) fulfills the criteria for deeming ethical explanations explanatory. Moreover, it suggests that ethical explanations are not mere shorthand for ethical belief explanations: whether explaining that "women were given the right to vote because people came to believe that denying it was morally wrong" meets the criteria to be explanatory does not depend on these meta-ethical commitments. In sum, Studies 1 and 2 demonstrated that participants do sometimes judge ethical explanations genuinely explanatory and are more likely to do so if they believe in moral progress and moral principles.

In Studies 3 and 4, we use between- and within-subjects designs to present a character who either offered an ethical explanation for a social change or made an alternative statement, and we asked participants to rate the extent to which the character believed in moral progress and moral principles. Participants gave higher ratings on both of these dependent variables when the character offered an ethical explanation than when the character offered a non-ethical explanation, ethical claim, or neutral explanation. This provides evidence that people use ethical explanations to make inferences about others' meta-ethical commitments, and that these inferences depend on both the ethical and explanatory components of ethical explanations.

These findings broaden our understanding of how people use and interpret explanations. Specifically, prior work has characterized explanations for events as explanatory when they appeal to causal information or generalizations (e.g., Salmon, 1984; Tworek & Cimpian, 2016; Vasil & Lombrozo, 2020; Woodward, 1989, 2003). Ethical explanations for social change do not appear to fit this characterization since, at least on the surface, they appeal to nothing more than an ethical claim. Explaining that "women were given the right to vote because denying it was morally wrong" appears to contain no information about how or why slavery ended. Philosophers have noted that ethical explanations are at least not obviously good explanations, leading to disagreement over what role they play (if any) in explaining social change (e.g., Brink, 1989; Cohen, 1997; Harman, 1986, 2007; Leiter, 2001; Luco, 2019; Sturgeon, 1988). However, our findings suggest that if taken with the belief that denying women the right to vote is objectively morally wrong and that the world progresses towards what is morally right, a picture of something akin to a causal force emerges. Thus, our moral commitments not only influence which explanations we prefer, but more surprisingly, determine whether or not we believe something is an explanation at all.

Alongside the contribution to our understanding of explanations, these findings also have practical importance. In conversations, we evidently use the explanations that others provide to make second-order inferences about their moral commitments. This may be especially important when meeting someone new, since we have little information to learn from. Our studies suggest that if you meet a colleague who says, "the death penalty will be abolished because it is morally wrong," you are more likely to think that she believes the death penalty is morally wrong (i.e., a first-order inference about what the explanation-provider believes is true). But you may also judge that she believes in moral progress and moral principles (i.e., a second-order inference about what the explanation-provider believes about the nature of morality such that this claim is explanatory).

These findings introduce many interesting questions for future research. First, differences in the explanatory power of ethical explanations depend on individual variation in moral progress and moral principles beliefs. But where do these differences come from? When others provide ethical explanations, we infer their moral commitments – does repeated exposure to these explanations shape our own moral commitments? Future work can address whether the relationship between ethical explanations and moral commitments is cyclical, such that one increases the likelihood of the other and vice versa.

Second, we have provided evidence for our findings within the social changes that we tested in a United States cultural context. We chose the items in Studies 1 and 2 for their widely acknowledged historical

significance and we chose the items in Studies 3 and 4 (which were all possible future social changes) because approximately 40–60% of Americans support each change according to Pew Research Center as of 2021 (Pew Research Center, 2015, 2021a, 2021b, 2021c, 2022). However, there was variation across social changes in both studies, which presents opportunity for future research to investigate whether these results generalize to other types of social change and why the results might not hold (or hold more weakly) for some types of social change.

Relatedly, participants in these studies were Prolific users based in the United States. The relevance of the social changes we tested certainly varies across cultures, which might affect the degree to which these changes are typically described in ethical or non-ethical terms. For example, imagine a country that has legalized physician-assisted suicide and that did so partly through political campaigns imbued with ethical explanations. In this country, the base rate of using an ethical explanation when talking about physician-assisted suicide might be inflated, such that when an explanation-provider gives an ethical explanation, the listener is unlikely to presume any particular meta-ethical commitments. Thus, particular social changes and their cultural contexts may interact with the extent to which listeners make second-order inferences about explanation-providers.

Third, these studies explored ethical explanations at the level of social change – it is possible, but not clear, whether people would endorse ethical explanations for individual-level actions. For example, if a friend returns a lost wallet, is saying he did it because “it was the right thing to do” genuinely explanatory (Uttich, 2012)? If so, is a belief in moral progress or moral principles still necessary, or are the criteria for explanatory adequacy fulfilled in a different way? What inferences are made, if any, of those who explain a mundane action in terms of it being the “right thing to do”? We hope to address these possibilities in future work to better understand the use and interpretation of explanations.

While there are many ways to expand on this work, these studies take crucial first steps towards demonstrating that explanations need not fulfill explanatory criteria at a surface level; rather, moral commitments can supplement an ethical explanation for social changes such that people view the explanation as explanatory without explicitly appealing to causal content. Moreover, people recognize this, and use it to make inferences about others. Thus, for some, “women were given the right to vote because denying it was morally wrong” can be genuinely explanatory, and for many, it can be used to evaluate others’ moral commitments.

Author note

All preregistrations, data, and materials are available online at https://osf.io/f9n3q/?view_only=aa2abbf8755c40e099ac4adaba0c3475.

CRedit authorship contribution statement

Casey Lewry: Writing – review & editing, Writing – original draft, Visualization, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **George Tsai:** Writing – review & editing, Conceptualization. **Tania Lombrozo:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Data availability

All preregistrations, data, and materials are available online at https://osf.io/f9n3q/?view_only=96fe9613ab244cf4962dfa0de06d49aa.

Acknowledgements

We would like to thank the members of the Concepts and Cognition

Lab for their valuable feedback on this project. This work was supported by the National Science Foundation Graduate Research Fellowship (DGE-2039656) awarded to Casey Lewry. We would also like to thank Kevin Uttich, whose dissertation (Uttich, 2012) offered valuable inspiration for this project.

References

- Amid a series of mass shootings in the U.S., gun policy remains deeply divisive. (2021, April 20). Pew Research Center - U.S. Politics & Policy. <https://www.pewresearch.org/politics/2021/04/20/amid-a-series-of-mass-shootings-in-the-u-s-gun-policy-remains-deeply-divisive/>.
- Brink, D. O. (1989). *Moral realism and the foundations of ethics*. Cambridge University Press.
- Budesheim, T. L., & Bonnelle, K. (1998). The use of abstract trait knowledge and behavioral exemplars in causal explanations of behavior. *Personality and Social Psychology Bulletin*, 24(6), 575–587. <https://doi.org/10.1177/0146167298246002>
- California legalizes assisted suicide amid growing support for such laws. (2015, October 5). Pew Research Center. <https://www.pewresearch.org/fact-tank/2015/10/05/california-legalizes-assisted-suicide-amid-growing-support-for-such-laws/>.
- Clayton, P. (2004). *Mind and emergence: From quantum to consciousness*. OUP Oxford.
- Cohen, J. (1997). The arc of the moral universe. *Philosophy & Public Affairs*, 26(2), 91–134. <https://doi.org/10.1111/j.1088-4963.1997.tb00078.x>
- Elsbach, K. D., & Elofson, G. (2000). How the packaging of decision explanations affects perceptions of trustworthiness. *Academy of Management Journal*, 43(1), 80–89. <https://doi.org/10.5465/1556387>
- Felton, M. K., & Kuhn, D. (2007). “How do I know?” the epistemological roots of critical thinking. *Journal of Museum Education*, 32(2), 101–110.
- Goodwin, G. P., & Darley, J. M. (2008). The psychology of meta-ethics: Exploring objectivism. *Cognition*, 106(3), 1339–1366. <https://doi.org/10.1016/j.cognition.2007.06.007>
- Goodwin, G. P., & Darley, J. M. (2012). Why are some moral beliefs perceived to be more objective than others? *Journal of Experimental Social Psychology*, 48(1), 250–256. <https://doi.org/10.1016/j.jesp.2011.08.006>
- Harman, G. (1986). Moral explanations of natural facts—Can moral claims be tested against moral reality? *The Southern Journal of Philosophy*, 24(S1), 57–68.
- Harman, G. (2007). *Ethics and observation* (pp. 36–41). Malden/Mass: Blackwell.
- Heiphetz, L., & Young, L. L. (2017). Can only one person be right? The development of objectivism and social preferences regarding widely shared and controversial moral beliefs. *Cognition*, 167, 78–90. <https://doi.org/10.1016/j.cognition.2016.05.014>
- Hopster, J. (2019). The meta-ethical significance of experiments about folk moral objectivism. *Philosophical Psychology*, 32(6), 831–852. <https://doi.org/ezproxy.princeton.edu/10.1080/09515089.2019.1627304>.
- Huiders, I., Voet, M., & De Wever, B. (2021). Students’ arguing skills: Impact of teaching beliefs and students’ epistemological understanding. In *In 19th biennial EARLI conference* (p. 120). EARLI.
- Ionescu, A., Furdui, R., Gavreliuc, A., Greenfield, P. M., & Weinstock, M. (2023). The effects of sociocultural changes on epistemic thinking across three generations in Romania. *PLoS One*, 18(3), Article e0281785.
- Kalish, C. (1998). Reasons and causes: Children’s understanding of conformity to social rules and physical laws. *Child Development*, 69(3), 706–720. <https://psycnet.apa.org/doi/10.2307/1132199>.
- Kirfel, L., Icard, T., & Gerstenberg, T. (2022). Inference from explanation. *Journal of Experimental Psychology: General*, 151(7), 1481. <https://psycnet.apa.org/doi/10.1037/xge0001151>.
- Klopp, E., & Stark, R. (2022). How to change epistemological beliefs? Effects of scientific controversies, epistemological sensitization, and critical thinking instructions on epistemological change. *Education in Science*, 12(7), 499.
- Kuhn, D. (1991). *The skills of argument*. Cambridge University Press.
- Kuhn, D. (2000). Metacognitive development. *Current directions in psychological science*, 9(5), 178–181. <https://doi-org.ezproxy.princeton.edu/10.1111/1467-8721.00088>.
- Leiter, B. (2001). Moral facts and best explanations. *Social Philosophy and Policy*, 18(2), 79–101. <https://doi.org/10.1017/S0265052500002910>
- Lewry, C., Asifriyaz, S., & Lombrozo, T. (2023). *Intuitive theories of moral progress* [Unpublished manuscript]. Department of Psychology, Princeton University
- Liquin, E. G., & Lombrozo, T. (2020). A functional approach to explanation-seeking curiosity. *Cognitive Psychology*, 119, 101276. <https://doi-org.ezproxy.princeton.edu/10.1016/j.cogpsych.2020.101276>.
- Liquin, E. G., & Lombrozo, T. (2022). Motivated to learn: An account of explanatory satisfaction. *Cognitive Psychology*, 132, 101453. <https://doi.org/10.1016/j.cogpsych.2021.101453>
- Lombrozo, T. (2012). *Explanation and abductive inference* (pp. 260–276). Oxford handbook of thinking and reasoning.
- Lombrozo, T., & Liquin, E. G. (2023). Explanation Is Effective Because It Is Selective. *Current Directions in Psychological Science*, 32(3), 212–219. <https://doi-org.ezproxy.princeton.edu/10.1177/09637214231156106>.
- Lombrozo, T., & Vasilyeva, N. (2017). *Causal explanation* 22 (p. 415). The Oxford handbook of causal reasoning.
- Luco, A. (2019). How moral facts cause moral progress. *Journal of the American Philosophical Association*, 5(4), 429–448. <https://doi.org/10.1017/apa.2019.23>
- Malle, B. F. (1999). How people explain behavior: A new theoretical framework. *Personality and Social Psychology Review*, 3(1), 23–48. https://doi.org/10.1207/s15327957pspr0301_2

- Malle, B. F. (2011). Time to give up the dogmas of attribution: An alternative theory of behavior explanation. In , Vol. 44. *Advances in experimental social psychology* (pp. 297–352). Academic Press. <https://doi.org/10.1016/B978-0-12-385522-0.00006-8>.
- Mason, L., Boldrin, A., & Zurlo, G. (2006). Epistemological understanding in different judgment domains: Relationships with gender, grade level, and curriculum. *International Journal of Educational Research*, 45(1–2), 43–56.
- Metz, S. E., Liquin, E. G., & Lombrozo, T. (2023). Distinct Profiles for Beliefs About Religion Versus Science. *Cognitive science*, 47(11), Article e13370. <https://doi.org.ezproxy.princeton.edu/10.1111/cogs.13370>.
- Moss, D., Montealegre, A., Bush, L. S., Caviola, L., & Pizarro, D. (2023). *Signaling (in) tolerance: The reputational implications of metaethical relativism and objectivism*. <https://doi.org/10.31234/osf.io/62z8c>
- Pölzler, T., & Wright, J. C. (2019). Empirical research on folk moral objectivism. *Philosophy Compass*, 14(5), Article e12589. <https://doi.org/10.1111/phc3.12589>
- Pölzler, T., & Wright, J. C. (2020). Anti-realist pluralism: A new approach to folk metaethics. *Review of Philosophy and Psychology*, 11, 53–82. <https://doi.org/10.1007/s13164-019-00447-8>
- Prasada, S., & Dillingham, E. M. (2006). Principled and statistical connections in common sense conception. *Cognition*, 99(1), 73–112.
- R Core Team. (2022). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. URL <https://www.R-project.org/>.
- Read, S. J., & Marcus-Newhall, A. (1993). Explanatory coherence in social explanations: A parallel distributed processing account. *Journal of Personality and Social Psychology*, 65(3), 429. <https://psycnet.apa.org/doi/10.1037/0022-3514.65.3.429>.
- Rosman, T., Peter, J., Mayer, A. K., & Krampen, G. (2018). Conceptions of scientific knowledge influence learning of academic skills: Epistemic beliefs and the efficacy of information literacy instruction. *Studies in Higher Education*, 43(1), 96–113. <https://doi-org.ezproxy.princeton.edu/10.1080/03075079.2016.1156666>.
- Rutjens, B. T., van der Pligt, J., & van Harreveld, F. (2009). Things will get better: The anxiety-buffering qualities of progressive hope. *Personality and Social Psychology Bulletin*, 35(5), 535–543. <https://doi.org/10.1177/0146167208331252>
- Rutjens, B. T., van Harreveld, F., & van der Pligt, J. (2010). Yes we can: Belief in progress as compensatory control. *Social Psychological and Personality Science*, 1(3), 246–252. <https://doi.org/10.1177/1948550610361782>
- Rutjens, B. T., van Harreveld, F., van Der Pligt, J., van Elk, M., & Pyszczynski, T. (2016). A march to a better world? Religiosity and the existential function of belief in social-moral progress. *The International Journal for the Psychology of Religion*, 26(1), 1–18. <https://doi.org/10.1080/10508619.2014.990345>
- Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world*. Princeton University Press.
- Sarkissian, H., Park, J., Tien, D., Wright, J. C., & Knobe, J. (2011). Folk moral relativism. *Mind & Language*, 26(4), 482–505. <https://doi.org/10.1111/j.1468-0017.2011.01428.x>
- Stanley, M. L., Marsh, E. J., & Kay, A. C. (2020). Structure-seeking as a psychological antecedent of beliefs about morality. *Journal of Experimental Psychology: General*, 149(10), 1908.
- Strevens, M. (2018). *The Mathematical Route to Causal. Explanation beyond causation: philosophical perspectives on non-causal explanations*, 96.
- Sturgeon, N. (1988). Moral explanations. *Essays on Moral Realism*, 229–255.
- Sulik, J., van Paridon, J., & Lupyan, G. (2023). Explanations in the wild. *Cognition*, 237, Article 105464. <https://doi.org/10.1016/j.cognition.2023.105464>
- Tworek, C. M., & Cimpian, A. (2016). Why do people tend to infer “ought” from “is”? The role of biases in explanation. *Psychological Science*, 27(8), 1109–1122. <https://doi.org/10.1177/0956797616650875>
- Uttich, K., Tsai, G., & Lombrozo, T. (2014). Exploring metaethical commitments: Moral objectivity and moral progress. In H. Sarkissian, & J. Cole Wright (Eds.), *Advances in moral psychology (experimental philosophy series)* (pp. 188–208). Continuum Press.
- Uttich, K. T. (2012). *The relationship between social cognition and moral reasoning*. Berkeley: University of California.
- Vasil, N., & Lombrozo, T. (2020). Structural thinking about social categories: Evidence from formal explanations, generics, and generalization. *Cognition*, 204, Article 104383. <https://doi.org/10.1016/j.cognition.2020.104383>
- Weinstock, M., Israel, V., Tabak, I., & Harari, Y. (2020). Young schoolchildren’s epistemic development: A longitudinal qualitative study. *Frontiers in Psychology*, 11, Article 531651.
- Woodward, J. (1989). The causal mechanical model of explanation. *Minnesota Studies in the Philosophy of Science*, 13.
- Woodward, J. (2003). *Scientific explanation*. The Stanford.
- Wright, C. (2023). *Essays on Relativism: 2001–2021*. Oxford University Press.
- Zijlstra, L. (2019). Folk moral objectivism and its measurement. *Journal of Experimental Social Psychology*, 84, Article 103807. <https://doi.org/10.1016/j.jesp.2019.04.005>
- Zijlstra, L. (2023). The psychology of Metaethics: Evidence for and against folk moral objectivism. In *Experiments in moral and political philosophy* (pp. 83–103). Routledge.