Kill or Die: Moral Judgment Alters Linguistic Coding of Causality

Julian De Freitas
Harvard University

Peter DeScioli
Stony Brook University

Jason Nemirow
Harvard University

Maxim Massenkoff
University of California, Berkeley

Steven Pinker
Harvard University

What is the relationship between the language people use to describe an event and their moral judgments? We test the hypothesis that moral judgment and causative verbs rely on the same underlying mental model of people's actions. Experiment 1a finds that participants choose different verbs to describe the major variants of a moral dilemma, the trolley problem, mirroring differences in their wrongness judgments: they described direct harm with a single causative verb (*Adam killed the man*), and indirect harm with an intransitive verb in a periphrastic construction (*Adam caused the man to die*). Experiments 1b and 2 separate physical causality from moral valuation by varying whether the victim is a person or animal and whether the harmful action rescues people or inanimate objects. The results show that people's moral judgments lead them to portray a causal event as either more or less direct and intended, which in turn shapes their verb choices. Experiment 3 finds the same basic asymmetry in verb usage in a production task in which participants freely described what happened.

Keywords: moral cognition, moral psychology, causative verbs, trolley problem, argument structure

Supplemental materials: http://dx.doi.org/10.1037/xlm0000369.supp

In many languages, speakers use distinct verbs for killing and dying (Haspelmath, 1993). The English verb *kill*, for example, is transitive, requiring a subject and object, as in *Alice killed the lobster*, whereas the verb *die* is intransitive, requiring only a subject, as in *The lobster died*. These verbs are not interchangeable: Speakers do not say *Alice died the lobster* or *The lobster killed* (to mean that the lobster died). In contrast, there are many other verbs that appear in both forms, such as *boil* in *Alice boiled the lobster* and *The lobster boiled* (Pinker, 1989, 2007; Gergely & Bever, 1986; Goldberg, 2001; Levin, 1993; Spellman & Mandel, 1999).

Killing and dying are also very different categories in our moral judgments. Someone could blame Alice for killing a lobster but she is not necessarily to blame if the lobster died. Many real-life situations are not obligatorily expressed by one or the other con-struction, and also do not fit into either moral category. If Alice caught the lobster and sold it to a cook, people might disagree about whether she *killed it* or *caused it to die*. Similarly, if a general ordered a drone strike that also hit civilians as collateral damage, people could disagree about whether the general *killed the civilians* or *caused them to die*. In both cases, observers might also disagree about the agents' moral culpability.

Here we investigate whether the same considerations that shape people's choice of verb, such as killing or dying, also shape their moral judgments. We build on a theory that moral judgment and verbs rely on the same underlying mental model of people's actions (Pinker, 2007). Prior research on verbs found that participants tend to use a single causative verb (a *lexical causative*) when an actor affects an object intentionally and directly, that is, without an intermediate link in the causal chain consisting of a second actor or of a natural event with causal potency such as electricity, a vehicle, or the weather (Wolff, 2003). In contrast, participants use an intransitive verb as a complement to a generic causal verb such as *cause* or *make* (a periphrastic construction) when an actor affects an object via intervening causes. Thus a boy *pops a balloon* when he pricks it, but *causes the balloon to pop* when he allows it to graze a hot light bulb on the ceiling. This difference between verb constructions can be explained by an underlying mental model of *force dynamics* in which people imagine that an antagonist exerts force on an agonist through physical contact (Pinker, 2007; Wolff, 2007). Foreseeabilty and intention matter as well (De Freitas & Johnson, 2015; Malle, 2006): An

actor who deliberately lets a balloon graze a hot bulb with the intention that the balloon burst could be said to *pop the balloon.*

This previous work examined morally neutral events such as moving a marble, extinguishing a candle, and turning on a TV. In this article, we examine the connection between language and morality by using moral dilemmas as stimuli to simultaneously observe verb choice and moral judgment. Specifically, we study two common variants of the trolley problem: the footbridge case in which an actor pushes one person in front of a trolley to save five people, and the switch case in which an actor flips a switch to redirect the trolley toward one person in order to save five people (Foot, 1978). Despite the similar tradeoff, most participants judge pushing to be morally wrong but flipping the switch to be morally permissible (Hauser et al., 2007).

Researchers have proposed several theories to account for the footbridge-switch difference, and these theories continue to be debated (DeScioli & Kurzban, 2009, 2013; Greene, Sommerville, Nystrom, Darley, & Cohen, 2001; Kurzban et al., 2012; Mikhail, 2007). We suggest that the same force-dynamic mental model that governs the choice of causative constructions might contribute to the moral difference (Pinker, 2007). In the footbridge dilemma, the actor directly contacts the victim, fitting the prototype for a causal action. In the switch dilemma, the actor only indirectly affects the victim by flipping a switch that redirects the trolley, deviating from the prototype.

If force dynamics underlie the footbridge-switch difference, then participants will choose different verbs to describe these morally contrasting scenarios. In Experiment 1a, we measure participants' preference for the transitive causative verb in *killed the man* over the intransitive verb in *caused the man to die* in the footbridge and switch scenarios. A *pure causality hypothesis* predicts greater use of the lexical causative *kill* for the footbridge than the switch scenario, mirroring the usual difference in judgments of wrongness. This result would point to a common mental model of action underlying two seemingly different phenomena, verb choice and moral judgment.

Experiments 1b through 2 consider a second mechanism that could align moral judgments with verb choices: in addition to a mental model of causation driving moral judgment, the reverse might occur. That is, people who condemn an action might portray its causal dynamics as more direct and intentional in order to support their accusation, using causative verbs to convey this condemnation; we refer to this as the *moral–physical conflation hypothesis*. For example, people who want to condemn a drone attack might choose to say *the general killed the civilians*, whereas defenders of the attack might prefer *the general caused the civilians to die*. In the trolley problem, greater condemnation of the actor in the footbridge version could motivate greater use of the causative verb *kill*. In support of this two-way influence, previous research shows that although moral judgments are influenced by more basic assessments of causality, intentions, and harm, the reverse can also occur: People's moral condemnation can lead them to portray causality, intentions, and harm to support their moral judgment (Alicke, 1992; DeScioli et al., 2011; Haidt, 2001; Knobe, 2005). For instance, people commonly judge that negative side effects are more intentional than positive side effects (Knobe, 2005). However, it is unknown whether these effects extend to language, shaping the verbs people choose.

The pure causality and moral–physical conflation hypotheses diverge when the physical causal structure of a scenario is kept constant while the moral structure is changed. In such instances, the pure causality hypothesis predicts no change in verb choice, because the verbs will reflect only the physical structure. In contrast, the moral–physical conflation hypothesis predicts that verb choice will mirror a change in moral structure, such that a more condemnable action is more likely to be described by a single causative verb. If so, participants will be more likely to choose the causative verb in the footbridge than the switch dilemma; yet, this difference would not necessarily occur in physically equivalent scenarios that do not differ in moral wrongness.

In Experiment 1b through Experiment 2, we separate physical causality from morality by using the same description of the physical forces that drive an event while changing the level of moral wrongness. We do so by varying whether the victim is a person or an animal and whether the actor rescued five people or five inanimate objects. In Experiment 3, we test whether the basic footbridge-switch difference in verb choice also translates to natural language production by asking participants to explain in open-ended responses their moral judgments for the footbridge and switch dilemmas.

Experiment 1a

Experiment 1a examines the connection between moral judgment and verbs by observing whether participants choose different verb constructions to describe the footbridge and switch dilemmas. The pure causality hypothesis predicts that people will show greater use of the causative verb *kill* in the footbridge dilemma than in the switch dilemma, mirroring the contrast in their moral judgments.

In this and subsequent experiments, the main focus is participants' verb choice. Hence, participants first answered a verb choice question before providing their moral judgments. In the present context, participants' moral judgments are essentially a manipulation check to confirm that they show the same footbridge-switch moral asymmetry that has been replicated many times before (Greene, 2014) or, in Experiments 1b through 2b, that this moral asymmetry has been muted or amplified as intended by the modified scenarios.

The primary measure of verb choice is participants' forced-choice selection between two sentences to describe the actor's behavior: a sentence with the lexical causative *kill* or a sentence with a periphrastic causative consisting of a generic causative added to the intransitive verb *die*. Second, we ask participants to rate the accuracy of each sentence separately. Here we are mainly interested in ratings of the lexical causative *kill,* because previous research found that lexical causatives are more particular to unmediated and/or intentional actions, whereas periphrastic causatives show less specificity and can be applied to mediated or unmediated causes (Wolff, 2003). Last, we follow previous research by also measuring participants' perceptions of the number of events that occurred in the dilemma.

Method

We recruited participants using Amazon's Mechanical Turk, an online crowdsourcing platform (see, Berinsky, Huber, & Lenz, 2012;

Buhrmester, Kwang, & Gosling, 2011; Goodman, Cryder, & Cheema, 2012; Ipeirotis, 2010; Paolacci, Chandler, & Ipeirotis, 2010). We chose a sample size that provides sufficient power to detect a large effect size, and excluded 16 participants for incomplete responses, previous participation in similar studies, or incorrectly answering a comprehension question at the end of the study ($n = 2$, see the following text for question wording), yielding a sample of 104 participants ($M_{age} = 33$ years, 43% female).

Participants were assigned to either the switch or footbridge condition. in the switch condition, participants read the following:

> One day Adam was walking near some trolley tracks. Suddenly, a trolley was quickly approaching out of control. Adam saw that five people were standing on a crosswalk in the trolley's path. Adam was standing next to a switch that could divert the trolley onto a sidetrack, but there was one man standing on the sidetrack. The five people were too far away to hear Adam's warnings and there was no other way to help them. Adam decided to pull the switch to divert the trolley. As a result, the five people escaped unharmed but the one man who was standing on the sidetrack was killed by the trolley.

In the footbridge condition, participants read a similar vignette in which Sentences 4 onward were replaced with the following:

> . . . Adam was standing on a footbridge over the tracks next to a man wearing a large, heavy backpack. Adam realized that he could slow the trolley and save the five people if he pushed a heavy object in the path of the trolley. The only object that was heavy enough was the man with the backpack. The five people were too far away to hear Adam's warnings and there was no other way to help them. Adam decided to push the man onto the tracks in front of the trolley. As a result, the five people escaped unharmed but the one man who was pushed onto the tracks was killed by the trolley.

Participants rated the accuracy of the following two statements about the scenario on a 7-point scale (scale anchors: *very poor description* to *very accurate description*), beginning with a statement with a causative verb (*Adam killed the man*) followed by a statement with an intransitive verb in a periphrastic construction (*Adam caused the man to die*). Participants also answered a forced-choice item about which of these statements best described what happened (*Choose the sentence that best describes what happened in this scenario*; the answer options were *Adam killed the man, Adam caused the man to die,* or *neither*). Following previous research on causal chains and verbs (Wolff, 2003), we also asked participants whether they thought that one or multiple events occurred in the scenario, using the same wording from previous work (*How many events do you think occurred in this scenario? Answer this question casually as if someone asked you how many events occurred*; the answer options were *one event* or *multiple events.*)

To measure moral wrongness, participants then answered a forced-choice question about whether the actor's behavior was morally wrong (yes or no). They also rated how morally wrong the behavior was on a 7-point scale (scale anchors: *not at all morally wrong* to *extremely morally wrong*). Participants then explained in a textbox how they made their decisions for the entire study (*Please describe how you made your decisions in this study*), answered a comprehension question about the scenario, *Which, if any, of the following events occurred in the scenario that you read?* (answer options: A. *A TV was stolen,* B. *A car was broken into,* C. *A woman screamed for help,* D. *All of the above,* E. *None of the above*), and completed demographic items.

Results and Discussion

Ratings of verb accuracy. Participants rated the causative verb *kill* as more accurate in the footbridge condition ($M = 5.72$, $SD = 1.78$) than in the switch condition ($M = 3.72$, $SD = 2.11$), $t(102) = 5.20$, $p < .001$, $d = 1.03$. Participants also rated the statement with the intransitive verb *die* as more accurate in the footbridge condition ($M = 6.24$, $SD = 1.48$) than in the switch condition ($M = 5.54$, $SD = 1.72$), $t(102) = 2.22$, $p = .028$, $d = 0.44$.

Forced-choice of verb accuracy. In the footbridge dilemma, participants chose the sentence with the causative verb *kill* 46% of the time, the intransitive verb *die* 48%, and neither 6%, whereas in the switch dilemma they chose *kill* 9% of the time, *die* 61% of the time, and neither 30% of the time. Chi-square tests showed that participants chose *kill* significantly more often in the footbridge than in the switch condition, $\chi^2(1, N = 104) = 17.81$, $p < .001$, $\varphi = .41$ (see Figure 1).

Number of events. There was no significant difference in whether participants thought multiple events (rather than a single event) had occurred in the footbridge (78%) and switch (83%) conditions, $\chi^2(1, N = 104) = 0.48$, $p = .491$, $\varphi = .07$.

Moral judgment. Participants' moral judgments replicated previous work as expected. For ratings of moral wrongness, participants rated the actor's behavior as more wrong in the footbridge condition ($M = 4.88$, $SD = 1.61$) than in the switch condition ($M = 2.94$, $SD = 1.72$), $t(102) = 5.91$, $p < .001$, $d = 1.17$. For the forced-choice question about whether the actor's behavior was morally wrong (yes or no), participants judged that the actor's behavior was morally wrong more often in the footbridge condition (68%) than in the switch condition (19%), $\chi^2(1, N = 104) = 26.04$, $p < .001$, $\varphi = .50$.

In sum, we find that participants' verb choices mirrored their moral judgments: They were more likely to choose the causative verb *kill* in the footbridge scenario than the switch scenario (see Figure 1).

Experiment 1b

Experiment 1b distinguishes between the possible mechanisms discussed in the introduction that could align verbs with moral judgments in the footbridge and switch dilemmas. According to the *pure causality hypothesis*, speakers categorize an event by consulting an intuitive model of the physical force dynamics involved, and this categorization drives both their verb choice and their moral judgment. According to a *moral–physical conflation* hypothesis, people have a mental category that combines physical causation and moral responsibility (an agent foresees, desires, and directly brings about an outcome), so that a moral judgment can alter their verbal description of an event, holding constant their understanding of the physical dynamics.

We separate causality from morality by changing the victim in the trolley dilemmas from a person to a cow. This keeps physical causality the same as in Experiment 1a, while making the actor's behavior less morally wrong (at least according to the prevailing moral standard in the West in which the life of an animal is not morally sacrosanct in the way that the life of a human is). The pure
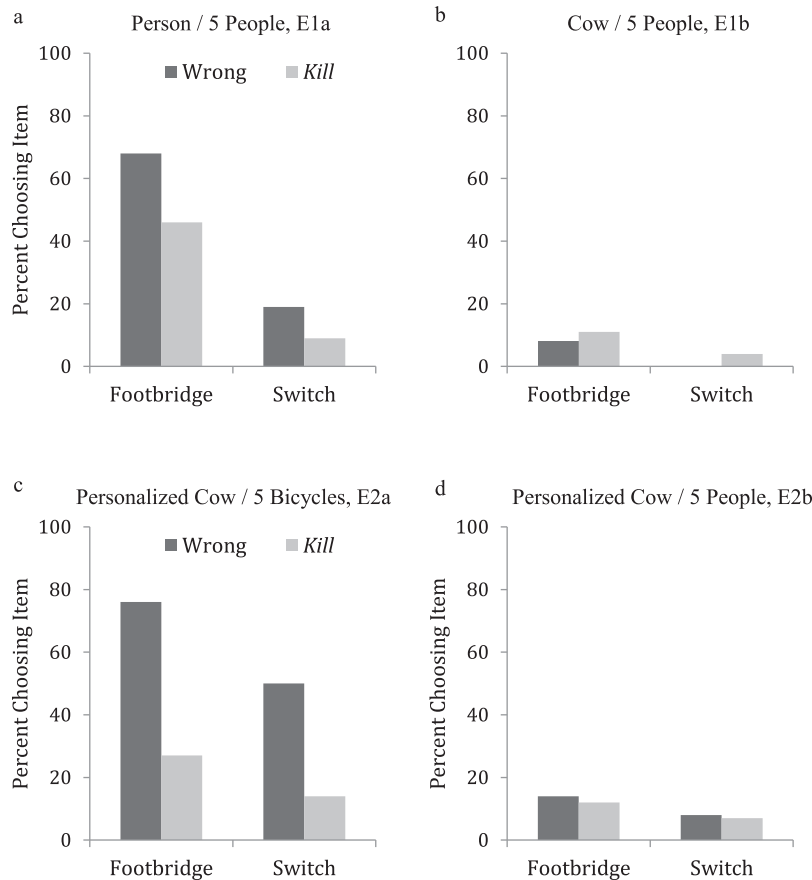
a

Person / 5 People, E1a

b

Cow / 5 People, E1b

c

Personalized Cow / 5 Bicycles, E2a

d

Personalized Cow / 5 People, E2b

Figure 1. Participants' moral judgments and choice of the verb *kill* (vs. *die* or neither) for Experiments 1a (a), 1b (b), 2a (c), and 2b (d).

causality hypothesis predicts that verb choice will show the same footbridge-switch difference as in Experiment 1a because the difference in causal structure is the same. Alternatively, the moral–physical conflation hypothesis predicts that participants will show no footbridge-switch difference in verb choice if there is no difference in moral wrongness.

Method

We recruited participants using Amazon's Mechanical Turk, choosing a sample size that provides sufficient power to detect a large effect size. We excluded 18 participants for incomplete responses, previous participation in similar studies, or incorrectly answering the same comprehension question as in Experiment 1a ($n = 4$), yielding a sample of 102 participants ($M_{age} = 34$ years, 53% female). The experimental design and stimuli were the same as in Experiment 1a, except the victim was changed from a man to a cow.

Results and Discussion

Ratings of verb accuracy. Participants rated the causative verb *kill* as more accurate in the footbridge condition ($M = 4.98$, $SD = 1.95$) than in the switch condition ($M = 3.94$, $SD = 2.32$), $t(100) = 2.46$, $p = .015$, $d = 0.49$. Participants' ratings for the

intransitive verb *die* did not differ between the footbridge condition ($M = 6.25$, $SD = 1.19$) and switch condition ($M = 5.96$, $SD = 1.61$), $t(100) = 1.03$, $p = .307$, $d = 0.21$.

Forced-choice of verb accuracy. In the footbridge dilemma, participants chose the sentence with the causative verb *kill* 11% of the time, the intransitive verb *die* 79% of the time, and neither 9% of the time, whereas in the switch dilemma they chose *kill* 4% of the time, *die* 78% of the time, and neither 18% of the time; the proportion choosing *kill* did not differ between conditions ($p = .272$, Fisher's exact test; see Figure 1).

Number of events. There was no significant difference in whether participants thought multiple events (rather than a single event) had occurred in the footbridge (68%) and switch (71%) conditions, $\chi^2(1, N = 102) = 0.08$, $p = .779$, $\varphi = .03$.

Moral judgment. Participants' wrongness judgments confirmed that changing the sacrificed victim to a cow muted the moral differences between footbridge and switch scenarios. For ratings of moral wrongness, participants rated the actor's behavior as more wrong in the footbridge condition ($M = 1.87$, $SD = 1.13$) than in the switch condition ($M = 1.39$, $SD = 0.53$), $t(100) = 2.71$, $p = .008$, $d = 0.54$, although these ratings were much lower than in Experiment 1a (footbridge: 4.88, switch: 2.94). For the forced-choice question about whether the actor's behavior was morally wrong (yes or no), wrongness judgments did not differ between the footbridge

condition (8%) and switch condition (0%; $p = .119$, Fisher's exact test). These percentages were much lower than in Experiment 1a (footbridge: 68%, switch: 19%), confirming that most participants viewed sacrificing a cow to save five people as morally permissible.

In short, we found that simply changing the victim from "man" to "cow" was sufficient to alter which verbs participants chose to describe the physical event (see Figure 1). Specifically, a large majority of participants in both scenarios preferred the intransitive verb *die* over the causative verb *kill*. This shows that verb choice is sensitive not only to the causal structure of a scenario but also to the intuitive wrongness of the action, supporting the moral–physical conflation hypothesis, in which people who condemn an action portray its causal dynamics as more direct and intentional to support their accusation, using causative verbs to convey this condemnation. Recall that the physical difference between the footbridge and switch conditions— pushing the victim or pulling a switch—was identical to the original scenarios used in Experiment 1a. Yet participants did not tend to choose the causative verb *kill* for pushing the cow to save five people, whereas they did for pushing the man to save five people, consistent with the moral–physical hypothesis that diminishing wrongness also diminishes the choice of the causative verb.

Experiment 2a

Experiment 2a further tests the moral–physical conflation hypothesis by changing the circumstances of the dilemmas to make sacrificing the cow morally wrong. Specifically, we changed the rescuees from people to bicycles, and we personalized the cow. We expected participants to judge sacrificing a special cow to save five bicycles to be morally wrong, and more so for the footbridge case. Critically, the moral–physical conflation hypothesis predicts that the greater wrongness in the footbridge case will lead to a parallel increase in participants' choices of the causative verb *kill*.

Method

We recruited participants using Amazon's Mechanical Turk, choosing a sample size that provides sufficient power to detect a large effect size. We excluded 13 participants for incomplete responses, previous participation in similar studies, or incorrectly answering a comprehension question at the end of the study ($n = 2$), yielding a sample of 147 participants ($M_{age} = 32$ years, 35% female). The experimental design and stimuli were the same as in Experiment 1, except that the victim was a personalized cow named "Lucy," who was sacrificed to save five bicycles. We personalized the cow to increase empathy and the immorality of sacrificing the cow, motivated by recent public outrage about the killing of a lion named Cecil (McLaughlin, 2015).

Participants were assigned to either the switch or footbridge condition. In the switch condition, participants read the following:

> One day Adam was walking near some trolley tracks. Suddenly, a trolley was quickly approaching out of control. Adam saw that five bicycles were standing on a crosswalk in the trolley's path. Adam was standing next to a switch that could divert the trolley onto a sidetrack, but there was one cow, named Lucy, standing on the side track. Lucy was dearly loved by the community and was well-known for her gentle demeanor and for raising many of the young cows in the town. The five bicycles were too far away and there was no other way to save them. Adam decided to pull the switch to divert the trolley. As

a result, the five bicycles escaped undamaged but Lucy, who was standing on the sidetrack, was killed by the trolley.

In the footbridge condition, participants read the following:

> One day Adam was walking near some trolley tracks. Suddenly, a trolley was quickly approaching out of control. Adam saw that five bicycles were standing on a crosswalk in the trolley's path. Adam was standing on a footbridge over the tracks next to a large cow, named Lucy. Lucy was dearly loved by the community and was well-known for her gentle demeanor and for raising many of the young cows in the town. Adam realized that he could slow the trolley and save the five bicycles if he pushed a heavy object in the path of the trolley. The only object that was heavy enough was Lucy the cow. The five bicycles were too far away and there was no other way to save them. Adam decided to push Lucy onto the tracks in front of the trolley. As a result, the five bicycles escaped undamaged and Lucy, who was pushed onto the tracks, was killed by the trolley.

Results and Discussion

Ratings of verb accuracy. Participants rated the causative verb *kill* as more accurate in the footbridge condition ($M = 5.77$, $SD = 1.78$) than in the switch condition ($M = 4.83$, $SD = 2.14$), $t(145) = 2.90$, $p = .004$, $d = 0.48$. Participants also rated the intransitive verb *die* as more accurate in the footbridge condition ($M = 6.52$, $SD = 0.95$) than in the switch condition ($M = 5.82$, $SD = 1.68$), $t(145) = 3.13$, $p = .002$, $d = 0.52$.

Forced-choice of verb accuracy. In the footbridge dilemma, participants chose the sentence with the causative verb *kill* 27% of the time, the intransitive verb *die* 72% of the time, and neither 1% of the time, whereas in the switch dilemma they chose *kill* 14% of the time, *die* 69% of the time, and neither 17% of the time; participants chose *kill* significantly more often in the footbridge than switch condition, $\chi^2(1, N = 147) = 3.69$, $p = .054$, $\varphi = .14$ (see Figure 1).

Number of events. There was no significant difference in whether participants thought multiple events (rather than a single event) had occurred in the footbridge (60%) and switch (61%) conditions, $\chi^2(1, N = 147) = 0.00$, $p = 1$, $\varphi = .00$.

Moral judgment. Participants' moral judgments showed the footbridge-switch difference as expected. For ratings of moral wrongness, participants rated the actor's behavior as more wrong in the footbridge condition ($M = 5.05$, $SD = 2.05$) than in the switch condition ($M = 3.96$, $SD = 2.35$), $t(145) = 3.01$, $p = .003$, $d = 0.50$. For the forced-choice question about whether the actor's behavior was morally wrong (yes or no), participants judged the actor's behavior (sacrificing a personalized cow to save five bicycles) to be morally wrong more often in the footbridge condition (76%) than in the switch condition (50%), $\chi^2(1, N = 147) = 10.69$, $p = .001$, $\varphi = .26$.

In sum, we find that when sacrificing a cow is morally questionable, participants attribute greater moral wrongness in the footbridge than switch case, and critically, they show a parallel increase in their choice of the causative verb *kill* (see Figure 1).

Experiment 2b

In Experiment 2b, we further test whether people's moral judgments are conflated with causality by reducing the moral wrongness of killing a personalized cow. We use the same dilemmas as

in Experiment 2a but change the rescuees from bicycles back to people. We expect participants to view sacrificing the personalized cow as justified when saving five people, and if so, the moral–physical conflation hypothesis predicts reduced differences between footbridge and switch conditions in participants' verb choices. This test also allows us to confirm that the pattern of differences found in Experiment 2a was specifically due to the wrongness of killing a personalized cow to protect bicycles, rather than the wrongness of killing a personalizing cow alone. Hence, we expect to see a similar pattern of results with sacrificing a personalized cow to save five people as we saw with sacrificing a generic cow to save five people in Experiment 1b.

Method

We recruited participants using Amazon's Mechanical Turk, choosing a sample size that provides sufficient power to detect a large effect size. We excluded 12 participants for incomplete responses, previous participation in similar studies, or incorrectly answering a comprehension question at the end of the study ($n = 2$), yielding a sample of 148 participants ($M_{age} = 35$ years, 56% female). The experimental design and stimuli were the same as in Experiment 2a, except that the rescuees were people instead of bicycles.

Results and Discussion

Ratings of verb accuracy. Participants' ratings for the causative verb differed between the footbridge ($M = 4.84$, $SD = 1.97$) and switch conditions ($M = 3.93$, $SD = 2.03$), $t(146) = 2.75$, $p = .007$, $d = 0.45$, but the difference was smaller than in Experiment 2a, and the verb accuracy ratings were lower overall. Participants' ratings for the intransitive verb *die* did not differ between the footbridge condition ($M = 5.92$, $SD = 1.62$) and the switch condition ($M = 5.65$, $SD = 1.58$), $t(146) = 1.00$, $p = .317$, $d = 0.17$.

Forced-choice of verb accuracy. In the footbridge dilemma participants chose the sentence with the causative verb *kill* 12% of the time, the intransitive verb *die* 79% of the time, and neither 8% of the time, whereas in the switch dilemma they chose *kill* 7% of the time, *die* 79% of the time, and neither 15% of the time; the proportion choosing *kill* did not significantly differ between conditions, $\chi^2(1, N = 148) = 1.38$, $p = .239$, $\varphi = .07$ (see Figure 1).

Number of events. There was no significant difference in whether participants thought multiple events (rather than a single event) had occurred in the footbridge (60%) and switch (76%) conditions, $\chi^2(1, N = 148) = 3.53$, $p = .060$, $\varphi = .15$.

Moral judgment. Participants' wrongness judgments confirmed that changing the rescues from bicycles to people muted the moral differences between footbridge and switch scenarios. For ratings of moral wrongness, participants ratings did differ significantly between the footbridge condition ($M = 2.41$, $SD = 1.75$) and the switch condition ($M = 1.85$, $SD = 1.45$), $t(146) = 2.12$, $p = .036$, $d = 0.35$, though the ratings were much lower overall compared to when the cow was sacrificed to save bicycles in Experiment 2a. For the forced-choice question about whether the actor's behavior was morally wrong (yes or no), participants' wrongness judgments of the actor's behavior (sacrificing a personalized cow to save five people) did not differ significantly between the footbridge (14%) and switch conditions (8%), $\chi^2(1, N = 148) = 1.25$, $p = .264$, $\varphi = .07$ (see Figure 1). These

percentages confirmed that most participants viewed sacrificing a personalized cow to save five people as morally permissible.

In sum, we found that reducing the moral wrongness of the actor's behavior, in this case by sacrificing a personalized cow to save people (rather than bicycles), also muted differences in the verbs people used to describe the events. Taken together, Experiments 2a and 2b indicate that when people condemn an action, they construe the causal role of the actor as direct and intended, but when they want to condone an action, they choose verbs that convey indirect causality. Because, once again, the physical causal structure was the same as in Experiment 2a when bicycles (rather than people) were at stake, these data are not adequately accounted for by the pure causality hypothesis. Instead, they support the moral–physical conflation hypothesis, in which changes in moral wrongness affect the verbs people choose to describe the action.

Experiment 3

Experiment 3 tests whether the basic footbridge-switch difference in verb choice also appears in participants' spontaneous verb usage. Previous research on causality and verb choice typically presents participants with preset sentences which they rate or select (e.g., Wolff, 2003, 2007). Hence, we adopted the same methodology in Experiments 1 and 2. These methods offer greater experimental control than examining spontaneous speech because there are typically a large number of ways that a speaker can describe events. Previous research has commonly assumed that participants' evaluations of preset sentences will also capture patterns in spontaneous language use.

However, it is possible that the preset sentences might not reflect what participants would spontaneously produce. Hence, in Experiment 3 we examine verb choice in a more naturalistic production experiment, in which participants are asked to simply write a sentence justifying the moral judgment that they made. We expected that these sentences would be more rich in content and nuanced than the preset sentences. Even so, we can observe whether participants spontaneously produce sentences that are similar to those used in Experiments 1 and 2, with a single causative verb *kill* or a periphrastic construction. Further, we can test whether verb usage differs across footbridge and switch dilemmas, as in the previous experiments with preset sentences.

Method

We recruited participants using Amazon's Mechanical Turk. We excluded 9 participants for incomplete responses, previous participation in similar studies, or incorrectly answering a comprehension question at the end of the study ($n = 4$), yielding a sample of 111 participants ($M_{age} = 38$ years, 58% female).

The experimental vignettes were the same as in Experiment 1a. Participants received either the footbridge or switch dilemma. Participants answered whether they thought the agent was guilty of murder (*Is Adam guilty of murder?* yes or no) and then explained their response (*Please explain your answer in one short sentence.*). These questions were designed to elicit participants' spontaneous descriptions of the relevant events in a natural way. Last, participants answered the same comprehension and demographic items as in previous experiments.

Two independent coders, who were blind to condition and the hypotheses, categorized participants' sentences according to a rubric

(see the Appendix for the complete coding instructions). Participants' responses were combined across conditions and randomly resorted for coding. The coders used the following five categories: one category for a causative verb *kill*, three categories for different kinds of indirect verbs (the intransitive verb *die* with and without a separate causative verb, and a clause with an inanimate subject like "trolley"), and a last category for sentences that did not fit the main categories. (Coders could select more than one category if they applied to different parts of a sentence but they only agreed on one such sentence.) Coders received instructive feedback on the first 20 sentences from one of the authors, who was blind to condition for these sentences. Intercoder reliability for the full set of sentences was high, coders selected the same category for 92 of the 111 responses, giving a Cohen's kappa value of .74. For analysis, we excluded the 19 cases in which coders chose different categories, yielding 92 final responses for analysis. The full coded dataset is available in the online supplemental material.

Results and Discussion

Forced-choice of the actor's guilt. As expected, participants judged that the actor was guilty of murder more often in the footbridge condition (90%) than in the switch condition (23%), $\chi^2(1, N = 92) = 42.02$, $p < .001$, $\varphi = .65$.

Production of direct verbs. For verb choice, we found that participants' sentences used a lexical causative 58% of the time in the footbridge dilemma versus 4% in the switch dilemma. Chi-square tests showed that participants chose *kill* significantly more often in the footbridge than in the switch condition, $\chi^2(1, N = 92) = 30.22$, $p < .001$, $\varphi = .55$ (see Figure 2). Participants wrote sentences like the following: "He willfully, knowingly murdered him. The motive doesn't matter" and "He still killed a man regardless of saving five lives."

Production of indirect verbs. In contrast, we did not find significant differences across conditions for the three indirect verb categories: intransitive verb with a separate causative verb, 15% (footbridge) versus 11% (switch), $\chi^2(1, N = 92) = 0.21$, $p = .647$, $\varphi = .02$; intransitive verb without a separate causative, 2% versus 0%, $\chi^2(1, N = 92) = 0.93$, $p = .336$, $\varphi = 0.00$; and clause with

inanimate subject, 0% versus 5%, $\chi^2(1, N = 92) = 2.23$, $p = .135$, $\varphi = .08$. There was also no significant difference when we grouped these three categories into a single indirect verb category, 17% versus 16%, $\chi^2(1, N = 92) = 0.01$, $p = .922$, $\varphi = .00$. Some examples of sentences that used indirect verbs include, "He took positive action to cause the death of another" and "He caused a man to die." Last, participants were less likely to write sentences in the "other" category for the footbridge (27%) than switch (80%) condition, $\chi^2(1, N = 92) = 25.32$, $p < .001$, $\varphi = .50$. Evidently, sentences in the switch condition were much more likely to be described without referring to killing or death at all. Many of these sentences tended to describe the difficulty of the agent's decision, emphasize that the agent did not intend to kill, and/or exonerate the agent in some way. For example, "He saved four people, made a very tough decision," "This is a blurry line, but he did not intend to kill someone. He tried to save others' lives," and "People were going to die anyway. He reduced the fatalities."

To summarize, participants spontaneously produced sentences using verbs that reflected the same distinction between direct and indirect causality, and these verb choices mirrored participants' moral judgments. Supporting the methods used in Experiments 1 and 2, we found that nontrivial proportions of participants spontaneously used both verb constructions that were used as stimuli in Experiments 1 and 2. Moreover, we observed greater use of a single causative verb *kill* in the footbridge dilemma than the switch dilemma, replicating this basic effect with a language production task.

General Discussion

The cumulative results of Experiments 1 through 3 show that the causal dynamics of a moral situation can influence the verbs people use to describe it. Yet these verb choices are also sensitive to other determinants of the moral gravity of the action. Roughly half of participants chose *kill* to describe the actor's behavior in the footbridge dilemma, but after we changed the victim from a person to a cow, only 11% of participants chose *kill* (Experiment 1). Experiment 2 further confirmed that participants were unlikely to use the causative verb *kill* if they believed the actor's behavior was morally justifiable. In these cases, even pushing the victim off the footbridge was described as though it was an indirect cause of death, perhaps because the impact of the trolley was now construed as the more immediate cause. Yet, when participants viewed the same pushing act as immoral, they were more likely to choose the causative verb *kill* to match their condemnation. These effects were not merely an artifact of providing preset sentences to participants: Experiment 3 showed that participants' own spontaneously written sentences showed the same asymmetry in verb choice based on moral judgment.

This pattern of results points to the existence of event categories that combine elements of physical causation (relatively direct, without intervening potent links) with elements of moral responsibility (foreseeing and intending a momentous outcome). A causative verb such as *kill* efficiently picks out moral transgressions from accidental, incidental, or inconsequential events, and communicating this choice of event category can be used rhetorically and tendentiously to influence how others behave (Pinker, 2007). When expressing a public protest, weaving a narrative before a jury, crafting a newspaper headline, or gossiping about everyday offenses, people can choose a
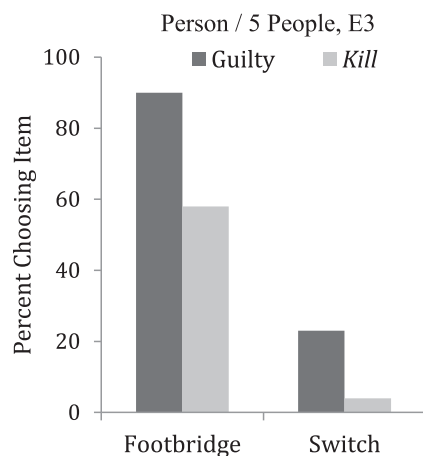
Figure 2. Participants' judgments about whether the agent was guilty of murder and their usage of a single causative verb for *kill* to describe the agent's action in Experiment 3.

lexical causative verb that implicates intentional, direct causality to bolster their moral accusation, as in *That surgeon killed my husband* or the famous 1960s Vietnam protest chant, "Hey, hey, LBJ, how many kids did you kill today?" Conversely, they will opt for periphrastic constructions containing generic causative verbs such as *make, lead to,* or *result in* when they seek to conceal agency and hence responsibility, as in the notorious euphemism *caused collateral damage* and the evasive *mistakes were made.* We conclude that the potency of verbs in such incidences and in moral reasoning more generally is due to the existence of cognitive categories which are essential in social discourse by singling out events for which speakers feel they have a right to hold the agents morally responsible, namely events whose outcomes are foreseen, intended, and directly caused without an intervening causal force.

This influence of morality on verb choice is predicted by the moral–physical conflation hypothesis. The pure causality hypothesis, in contrast, emphasizes only the physics of the interaction, and so cannot account for the large effect of moral judgment on verb choice. This is not to say that physics does not also play an important role. Previous research already showed that physical causality affects verb choice (e.g., Wolff, 2003, 2007). Indeed, the fact that participants chose verbs that portray a more direct and hence culpable physical role for the agent implies that they expect listeners' moral judgments to be swayed by physical causality. The novelty of the current contribution can thus be summarized in two mutually reinforcing discoveries: (a) moral judgments influence verb choice, and (b) verb choice, in turn, can reveal how moral judgments interact with causal cognition.

One interesting question about the present results is why participants tended to judge the footbridge actor as wrong more often, overall, than they chose *kill* over *cause to die*. That is, why did some participants condemn the agent but still show some resistance to saying he killed the victim? This reluctance might reflect how people grapple with the dilemma. Although they judge that killing a man is wrong, they also understand that the intention behind it was to save five lives. The forced-choice wrongness measure did not allow participants to express these mixed feelings, whereas some who judged the action as wrong might have still felt that the *cause to die* sentence better captured their mixture of judgment and understanding of good intentions. Furthermore, because *cause to die* can be an acceptable description for both direct and indirect causation (Wolff, 2003), participants did not need to choose *kill* unless they wanted to emphasize direct causation.

One question for future work is how people choose verbs in the midst of communication, debate, and negotiation about moral wrongdoing. The present experiments used simple, controlled tasks in which participants evaluated sentences or explained their judgments in a single statement. Even in this minimal social context, moral judgment influenced verb choice. However, the interactions between verbs, causality, and moral judgment are likely to be even more complex in back-and-forth dialogue about an agent's behavior. For instance, a speaker might exaggerate direct verbs to threaten condemnation or to recruit others against the accused. Alternatively, a speaker could offer indirect verbs to convey a willingness to compromise and assign only moderate responsibility to the agent. Finally, future research can examine other morally relevant verbs in addition to *kill* and *die.* Although there are not many mutually exclusive verb pairs like *kill* and *die,* there are many examples of lexical versus periphrastic causatives with the same verbs, sometimes in the passive or intransitive

form, which may be contrasted for rhetorical or moral effect, for example, *Mike injured him* versus *Mike caused/allowed him to be injured* and *Sally broke the bicycle* versus *Sally caused the bicycle to break.* More generally, we believe that the intersection between moral psychology and psycholinguistics can explain many word choices in language production, and holds the promise of illuminating issues in the study of composition, rhetoric, persuasion, and style.

References

Alicke, M. D. (1992). Culpable causation. *Journal of Personality and Social Psychology, 63,* 368–378. http://dx.doi.org/10.1037/0022-3514 .63.3.368

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's Mechanical turk. *Political Analysis, 20,* 351–368. http://dx.doi.org/10.1093/pan/mpr057

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality data? *Perspectives on Psychological Science, 6,* 3–5. http://dx.doi.org/10.1177/ 1745691610393980

De Freitas, J., & Johnson, S. G. B. (2015). Behaviorist thinking in judgments of wrongness, punishment, and blame. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), In *Proceedings of the 37th annual conference of the Cognitive Science Society* (pp. 524–529). Austin, TX: Cognitive Science Society.

DeScioli, P., Bruening, R., & Kurzban, R. (2011). The omission effect in moral cognition: Toward a functional explanation. *Evolution and Human Behavior, 32,* 204–215. http://dx.doi.org/10.1016/j.evolhumbehav.2011 .01.003

DeScioli, P., & Kurzban, R. (2009). Mysteries of morality. *Cognition, 112,* 281–299. http://dx.doi.org/10.1016/j.cognition.2009.05.008

DeScioli, P., & Kurzban, R. (2013). A solution to the mysteries of morality. *Psychological Bulletin, 139,* 477–496. http://dx.doi.org/10.1037/ a0029065

Foot, P. (1978). *Virtues and vices and other essays in moral philosophy.* Berkeley, CA: University of California Press.

Gergely, G., & Bever, T. G. (1986). Related intuitions and the mental representation of causative verbs in adults and children. *Cognition, 23,* 211–277.

Goldberg, A. E. (2001). Patient arguments of causative verbs can be omitted: The role of information structure in argument distribution. *Language Sciences, 23,* 503–524.

Goodman, J. K., Cryder, C. E., & Cheema, A. (2012). Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral Decision Making, 26,* 213–224. http://dx.doi.org/ 10.1002/bdm.1753

Greene, J. (2014). *Moral tribes: Emotion, reason and the gap between us and them.* New York, NY: Penguin.

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science, 293,* 2105–2108. http://dx.doi.org/10.1126/science .1062872

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108,* 814–834. http://dx.doi.org/10.1037/0033-295X.108.4.814

Haspelmath, M. (1993). More on the typology of inchoative/causative verb alternations. In B. Comrie & M. Polinsky (Eds.), *Causatives and transitivity.* Amsterdam, the Netherlands: John Benjamins. http://dx.doi.org/ 10.1075/slcs.23.05has

Hauser, M., Cushman, F., Young, L., Kang-Xing Jin, R., & Mikhail, J. (2007). A dissociation between moral judgments and justifications. *Mind & Language, 22,* 1–21. http://dx.doi.org/10.1111/j.1468-0017.2006 .00297.x

Ipeirotis, P. (2010). *The new demographics of Mechanical Turk*. Retrieved from http://www.behind-the-enemy-lines.com/2010/03/new-demographics-of-mechanical-turk.html

Knobe, J. (2005). Theory of mind and moral cognition: Exploring the connections. *Trends in Cognitive Sciences, 9,* 357–359. http://dx.doi.org/10.1016/j.tics.2005.06.011

Kurzban, R., DeScioli, P., & Fein, D. (2012). Hamilton vs. Kant: Pitting adaptations for altruism against adaptations for moral judgment. *Evolution and Human Behavior, 33,* 323–333. http://dx.doi.org/10.1016/j.evolhumbehav.2011.11.002

Levin, B. (1993) *English verb classes and alternations: A preliminary investigation*. Chicago, IL: University of Chicago Press.

Malle, B. F. (2006). Intentionality, morality, and their relationship in human judgment. *Journal of Cognition and Culture, 6,* 87–112. http://dx.doi.org/10.1163/156853706776931358

McLaughlin, E. C. (2015, October 12). *Zimbabwe won't press charges against Cecil the Lion's killer*. Retrieved from http://www.cnn.com/2015/10/12/africa/zimbabwe-cecil-lion-walter-palmer-no-charges/

Mikhail, J. (2007). Universal moral grammar: Theory, evidence and the future. *Trends in Cognitive Sciences, 11,* 143–152. http://dx.doi.org/10.1016/j.tics.2006.12.007

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making, 5,* 411–419.

Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure*. Cambridge, MA: MIT Press.

Pinker, S. (2007). *The stuff of thought: Language as a window into human nature*. Penguin.

Spellman, B. A., & Mandel, D. R. (1999). When possibility informs reality: Counterfactual thinking as a cue to causality. *Current Directions in Psychological Science, 8,* 120–123.

Wolff, P. (2003). Direct causation in the linguistic coding and individuation of causal events. *Cognition, 88,* 1–48. http://dx.doi.org/10.1016/S0010-0277(03)00004-0

Wolff, P. (2007). Representing causation. *Journal of Experimental Psychology: General, 136,* 82–111. http://dx.doi.org/10.1037/0096-3445.136.1.82

Appendix

Instructions Given to Coders in Experiment 3

The following instructions were provided independently to two hypothesis-blind coders, who used the instructions to code the sentences produced by participants in Experiment 3.

In this task, you will read responses written by study participants and code the grammar that participants used in their response.

The Study

Participants read some scenarios about Adam, who faced a dilemma involving five people, one man, and an oncoming trolley. If Adam acted one way, five people would die; if Adam acted another way, one man would die. Adam made the choice, and one man died. (We are leaving out the specifics of the scenarios so they do not influence coding.)

Participants then answered the questions:

1. Is Adam guilty of murder? (yes or no)

2. Please explain your answer in one short sentence.

You will read participants' responses to Question 2 and code the grammar that they used to refer to the man's death.

Grammar Primer

Here are a few basic grammatical distinctions to help you code the grammar of participants' responses. [1]

- **Transitive verb.** A transitive verb takes the form subject + verb + object, such as *Alice kicked the ball*, or *Alice killed the spider*.
- **Intransitive verb.** An intransitive verb takes the form subject + verb, without requiring an object, such as *Alice jumped* or *The spider died*.
- **Passive voice.** A verb in the passive voice appears in the participle form (usually with an —*en* or –*ed* suffix), and has the done-to or acted-upon as the subject, while either omitting the doer or actor or expressing it in a *by-phrase* as in *The ball was kicked, The cake was eaten by the boy,* or *The woman was hit by the baby*.
- **Causal verb.** A causal verb is about causality, including caused/let/made/allowed, and is often followed by another verb that is intransitive or in the passive voice. For example, *Alice made the ball roll*, *Alice caused the spider to die*, or *Alice allowed the spider to be killed*.

(*Appendix continues*)

Coding Instructions

Read each participant's sentence(s). Some responses might be long and wordy, but focus on the part(s) of the sentence that refers to the man's death. We want you to code the grammatical structure of that part by using the rubric below.

Record whether that part most closely fits one of the categories below by putting a "1" in the categories' column within the excel sheet. Each response will typically fit best in exactly one category, except in rare cases in which two different parts of the sentence fit different categories; in these rare cases, you can select multiple categories.

Coding Rubric

(1) Adam killed the man.

The sentence includes a single clause with the subject Adam, a transitive verb meaning *kill*, and the man as the direct object of the verb.

Examples of acceptable variations: Adam/He killed/murdered/sacrificed the man. Adam/He Committed murder/manslaughter.

(2) Adam caused the man to die.

The sentence includes a main clause with the subject Adam, a verb about causality, the man, and a subordinate clause with a separate verb (either an intransitive verb in the active voice, such as *die*, or a transitive verb in the passive voice, such as *be killed* or *get run over*), for the man's death.

Examples of acceptable variations: Adam/he caused/let/made/ allowed the man to die/be killed/get run over. Adam's/his action led to the death of the man.

(3) Adam . . . and the man died.

This category is the same as (2), except that the clause does not involve a causative verb like caused/let/made/allowed/led.

Examples of acceptable variations: Adam/he . . . and the man died/was killed/was run over as a result/as a consequence.

(4) the trolley killed the man.

The sentence includes a clause with the trolley or another inanimate thing as the subject rather than Adam, and the man as the direct object of the verb in that clause. That clause can also be a subordinate clause in a sentence in which the main clause refers to what Adam did.

Examples of acceptable variations: Adam/he . . . caused the trolley to kill the man; Adam/he . . . and the trolley killed the man.

(5) Other

Pick this category if the sentence does not fall into any of the first four categories. The sentence does not refer to the man's death. It might restate that Adam made a tough choice, without specifically saying the man died. Or it might say that Adam pushed the man or flipped a switch, also without saying the man died. Also included in this category are idioms, such as "pushed the man to his death" and "took the man's life."