Philosophical Transactions

Moral judgments reflect default representations of possibility

Abstract

Keywords: Morality; possibility; modality; modal cognition; moral psychology

Moral judgment requires representing what is possible: judging that someone ought to do something implies that they actually can do that thing. And if they cannot do that thing, then it’s not the case that they ought to have done it. Moral judgments are often made quickly and effortlessly, suggesting that they may rely on default, rather than deliberative, representations of what is possible. To investigate this possibility, we asked participants to make 10 different kinds of moral judgments either very fast or more slowly about 240 different actions across 12 contexts. We found that these moral judgments were more similar to one another when participants were forced to quickly assess the morality of immoral actions, suggesting a common default template for moral judgment that becomes more differentiated upon reflection. When making moral permissibility judgments quickly, participants were more likely to judge that improbable, irrational, and impossible actions were not permissible, indicating that default representations of permissibility may be reflecting default representations of possibility. A direct investigation revealed a close relationship between default representations of possibility and fast judgments of moral permissibility. These findings demonstrate the role of default representations of possibility in moral cognition.

Humans represent and reason about what’s possible, engaging in what is often called ‘modal’ thought. This capacity underlies many forms of high-level cognition, from causal reasoning to language comprehension (Phillips & Cushman, 2017; Lagnado, Gerstenberg, & Zultan, 2013; Shtulman & Tong, 2013; Shou, Olney, Smithson, & Song, 2020). Moreover, it plays a key role in moral cognition (Byrne, 2017; Byrne & Timmons, 2018; Tepe & Byrne, 2022). When reasoning about moral judgments of ‘permissibility,’ like whether an agent ‘ought to’ or ‘should’ do something, one must first consider what options are available to the agent. And in cases where a given action is not possible for that agent to do, then it is also not the case that they ‘should’ or ‘ought to’ do it (for more on the “Ought Implies Can” principle, see Sidgwick, 1884; Parfit, 1984; Zimmerman, 1996).

A notable feature of everyday moral judgments is that despite requiring us to reason about possibilities, we are often able to make moral judgments quickly and effortlessly. This suggests that they may recruit implicit or *default* (rather than deliberative) representations of what is possible. Previous work exploring such default representations of possibility finds that a signature feature of default possibility representations is that they often exclude events that are prescriptively or descriptively abnormal, i.e., immoral, irrational, and statistically improbable events (Phillips & Cushman, 2017; Phillips & Knobe, 2018; Phillips, Morris, Cushman, 2019). The present research extends this line of research by focusing specifically on moral judgments and asking whether and how they recruit default representations of possibility.

In the present studies, we ask participants to make 1 of 5 different moral permissibility judgments about 240 actions across 12 different scenarios. These judgments were either made under time pressure or with unlimited time to respond. Using the resulting data set of over 100,000 moral judgments, we examine the relationship between the different moral judgments, and between moral judgments and default representations of possibility. We find that permissibility judgments made under time pressure exhibit a signature feature of default representations of possibility, namely tending to exclude improbable and irrational events. We then go on to directly investigate the relationship between these moral judgments and speeded judgments of what is possible, finding that permissibility judgments, especially when made quickly, reflect whether these actions are regarded as possible by default.

Connections between modal and moral cognition

Prior research suggests points of connection between moral reasoning and reasoning about possibility. For example, people's judgments of morality influence their perceptions of the relevance of alternative possibilities, such that they think it is more relevant to consider morally good possibilities than morally bad ones (Phillips, Luguri, & Knobe, 2015). Moreover, prior work by Shtulman and Tong (2013), found that the more frequently participants judge extraordinary events as possible, the more often they judge extraordinary actions as permissible.

One approach that has been taken to studying default modal cognition is to ask people to make judgments of what is possible either quickly or only after taking time to reflect (Phillips & Cushman, 2017). This research found two pieces of evidence for a relationship between modal and moral judgments. First, when people were forced to make possibility judgments quickly, people increased their tendency to report that immoral events in particular were not possible. Second, when participants were forced to make a large range of modal judgments quickly (what someone ‘could’ do, ‘might’ do, ‘may’ do, etc.), all of these judgments became increasingly correlated with judgments of what someone ‘ought’ to do. These findings suggest that judgments of permissibility, especially moral permissibility, constrain default representations of possibility.

Moreover, there is evidence that this tendency appears early in human development. Studies with young children have established that children’s perceptions of possibility are related to, and even constrained by, morality. For example, recent developmental work demonstrates that young children’s judgments of what is possible or could be done are constrained by their ideas about what would be good to do (Kushnir, Gopnik, Chernyak, Seiver, & Wellman, 2015), and by moral value in particular (Phillips & Bloom, 2017; Shtulman & Phillips, 2018). Three- to 5-year-old children judge that immoral actions such as stealing are impossible and would require magic, and do so to approximately the same extent as violations of physical laws (Phillips & Bloom, 2017; Shtulman & Phillips, 2018). Collectively, this work suggests that moral judgments influence and even constrain representations of what is possible from an early age, and that this relationship persists into adulthood, at least when judgments of possibility are made under time pressure (Phillips & Cushman, 2017).

Critically, recent developmental work suggests that not only does morality constrain ideas about possibility, but the converse occurs as well: possibility constrains judgments of moral permissibility. For example, children infer prescriptive norms from simply learning that actions are normal (Roberts, Gelman, & Ho, 2016). Moreover, when playing a simple game, young children intervene when third parties act in ways that violate descriptive norms (Casler, Terziyan, & Greene, 2009; Rakoczy, Warneken, & Tomasello, 2008). Further, prior work found that young children tended to judge impossible and improbable events to be ‘wrong’ and deserve punishment, suggesting that they may not be clearly differentiating between an event's moral permissibility and its physical possibility (Shtulman & Phillips, 2018). These discoveries raise the question of whether this default way of reasoning about morality persists into adulthood. In the present research, we ask whether adults’ ideas about possibility influence their moral judgments, especially when they are made quickly.

Present research

The present research examines the relationship between representations of possibility and judgments of morality, and specifically asks whether moral judgments reflect the signature features of default representations of possibility. We build on a methodological approach introduced in Phillips and Cushman (2017). Participants were assigned to make 1 of 5 moral judgments (e.g., ‘Would it be morally acceptable for an agent to do the following action?’) about 240 different actions in 5 event category types (i.e., immoral events, irrational events, improbable events, ordinary events, and physically impossible events) across 12 contexts. Participants were either forced to respond under time pressure (<1550ms) or were asked to reflect before responding. We then examine whether these moral permissibility judgments converge on a common default representation by examining the similarity between them when participants are forced to make moral judgments quickly vs. slowly. Second, we then asked a new set of participants to judge the possibility of these events either quickly or slowly. Combining the resulting datasets allowed us to ask whether moral judgments made quickly reflect default representations of possibility.

Methods

For all the studies described below, English-speakers located in the United States were recruited via Amazon Mechanical Turk or Prolific. Eligible participants first consented and then answered a brief demographic questionnaire asking for age, gender, education level, and handedness. Subsequently, they were presented with 12 different vignettes describing different agents in a variety of scenarios in which they faced a problem. For example, a participant might be instructed to read the following vignette:

> For her son's birthday, Emily bakes a cake to serve to the children in her son's class. One child in the class is allergic to nuts, so the cake needs to be nut-free. However, right before serving the cake, Emily realizes the cake contains nuts.

After reading, participants were presented with 20 different actions that were designed to fall into 5 different categories. To illustrate in the above context, 4 of these actions were intended to be perceived as ordinary (e.g., 'serve ice cream instead'), 4 were meant to be impossible (e.g., 'zap the nuts from the cake'), 4 immoral (e.g., 'lie that the cake is nut-free'), 4 improbable (e.g., 'discover a nut-free cake in the house'), and 4 irrational (e.g., 'postpone the party for a month'). The same contexts and actions were used across all of the following studies except where specifically noted (see supplemental materials for the text of all 12 scenarios and 240 actions).

The studies described differed in terms of what kind of judgment participants made about the actions and whether they were forced to respond quickly or given unlimited time to respond. Each study took approximately 15 minutes to complete. Upon successful completion of the study, participants were debriefed regarding the study methods and hypotheses, and were compensated $3.

The data, code, and all other supplementary materials for these studies are publicly available at https://doi.org/10.7910/DVN/UNNFSL.

Norming study.

We first conducted a study to confirm that participants perceived the 5 different kinds of events as intended.

Participants. Participants (N=240) were recruited through Amazon Mechanical Turk (108 females; 2 non-binary, M_{age} = 40.12, SD_{age} = 11.86).

Procedure and materials. After consenting and providing demographic information, participants responded to 20 events for 6 of the 12 scenarios (see supplemental materials for the text of all 12 scenarios). Participants were given unlimited time to respond. Scenarios were randomized across participants using a Latin Square Design. Participants made a single kind of judgment about all 120 actions they evaluated. These questions all took the form 'How [impossible/improbable/immoral/irrational] is it for [agent] to [action]?'. Participants responded on a 5-point scale with an additional option for 'not applicable' (1 = very [impossible/improbable/abnormal/immoral/irrational] to 5 = very [possible/probable/normal/moral/rational]; with an additional option: 6 = not applicable). This resulted in approximately 27 data points per judgment per action.

Morality and possibility judgments.

Procedure and materials. For the following two sets of studies, participants made judgments of the morality or possibility of different actions that participants rated in the prior study. Participants were randomly assigned to a condition in which they had to respond very quickly (in <1550ms from stimulus onset) or were asked to reflect before responding. In the “fast” condition, participants were instructed, ‘Please answer as quickly and accurately as you possibly can. You will only have about 1 second to respond to each event’ and the study automatically progressed after 1550ms. In the “slow” condition, participants were instructed, ‘Please take your time and carefully reflect on these questions. Make sure you do not answer too quickly or carelessly’ and the study did not progress until participants responded.

Morality judgments

Participants. Participants (N=584) were recruited through Amazon Mechanical Turk (266 females; 2 non-binary, M_{age} = 39.91, SD_{age} = 12.66). Participants were recruited in 10 separate groups that systematically varied both the moral judgment being made and the amount of time participants had to make that judgment. Because of an experimenter error in randomization, we collected more data than intended in two groups. In addition, 10 participants participated in more than one version of the study; in all analyses we include only data from the first time any participant took the study.

Procedure and materials. As part of our data collection process we collected judgments of permissibility and impermissibility, but for simplicity and brevity this paper only presents the analysis of the permissibility judgments.[1] Participants made 1 of 5 different moral judgments (judgments of whether the action was morally *acceptable*, should be *allowed*, would be *approved* of, would be *okay* to do, or *should* be done) about all of the 240 different actions we used. Three hundred and twenty-two participants were forced to respond under time pressure, while 262 were instructed to reflect before responding. On each trial, participants were asked to respond ‘yes’ or’ ‘no’ to the question of whether it would be [1 of the 5 moral permissibility terms] to do the action presented (see supplemental materials for the full text of all 5 moral judgments and all 240 different actions).

Possibility judgments

Participants. Participants (N=192) were recruited through Prolific and Amazon Mechanical Turk. Participants were assigned to make judgments about the possibility of the same 240 actions either under time pressure or after reflecting. Seven participants participated in more than one version of the study; in all analyses we include only data from the first time any participant took the study.

[1] A brief, analogous description of the results for impermissibility judgments can be found in the supplementary materials at https://doi.org/10.7910/DVN/UNNFSL. We would like to thank the editors and anonymous reviewers for suggesting this way of simplifying the results we present.

Procedure and materials. The procedure and materials for the possibility judgments study were the same as the moral judgments studies, except that participants were asked to respond 'yes' or 'no' to the possibility of the 240 actions, rather than the morality of those actions.

Data cleaning and coding

Exclusion criteria. Data were excluded at both the participant and trial level. We first calculated the average response time for each participant (excluding trials on which they took longer than 5 seconds to respond). We then excluded all data from any participant in the "slow" condition who responded on average in less than 1200 ms (51 participants in the moral judgment tasks; 7 participants in the possibility task), and all data from any participant in the "fast" condition who responded on average in less than 800ms (20 participants in the moral judgment task; 2 participants in the possibility task). At the trial level, we excluded data according to two criteria (1) all trials on which rt < 500ms were excluded as this was an insufficient amount of time to read the action presented (<1% of the data excluded in the moral judgment task; ~1% of the data in the possibility judgment task), and (2) all trials on which participants did not provide a response were excluded (<4% of the trials for the moral judgment tasks; ~ 4% of the trials for the possibility judgment tasks).

Coding. While participants in the "slow" condition were instructed to carefully reflect before responding, we found that they did not always do so on all trials. Accordingly, we coded responses on a trial-by-trial basis. All trials on which participants responded in less than 1550ms (the max time available in the "fast" condition) were coded as "fast" responses. All trials on which participants responded in 1550ms or more were coded as "slow" responses.[2]

Results

Norming judgments

We first analyzed participants' judgments of how abnormal, immoral, improbable, and irrational each action was to confirm that the actions we created fell into their intended action category. To do so, we created an average score for each action for all four judgment types and then compared these mean ratings for actions within one category to the mean ratings for actions in the other categories. This analysis confirmed that participants largely perceived the actions in each category as we intended. For example, participants rated the "ordinary" actions as less abnormal than actions in any other category ($t > 21.57$; $p < .001$, $d > 4.40$ for all comparisons). Participants also rated the "immoral" actions as more immoral than any other kind of action ($t > 13.09$; $p < .001$, $d > 2.67$ for all comparisons). They rated irrational actions as more irrational than all other kinds of actions, except for impossible ones ($t > 4.10$; $p < .001$, $d > 0.83$ for all other comparisons). And they rated improbable actions as more improbable than

[2] An alternative approach is to use the a priori condition labels instead of the posthoc coding and simply exclude the 'slow' trials on which participants took less than 1550ms. While this approach substantially reduces the proportion of data included in the analyses, it yields qualitatively similar patterns across all key analyses and does not affect the significance of any of the critical results.

ordinary actions (*t* = 22.47; *p* < .001, *d* = 4.59), somewhat more improbable than immoral actions (*t* = 1.84; *p* = .070, *d* = 0.38), but less improbable than irrational or impossible actions (*t* > 8.82; *p* < .001, *d* > 0.98 for both comparisons). See Figure 1 for the overall pattern for each action category. Having confirmed that the actions were broadly perceived as intended, we next analyzed participants' moral judgments of these actions.

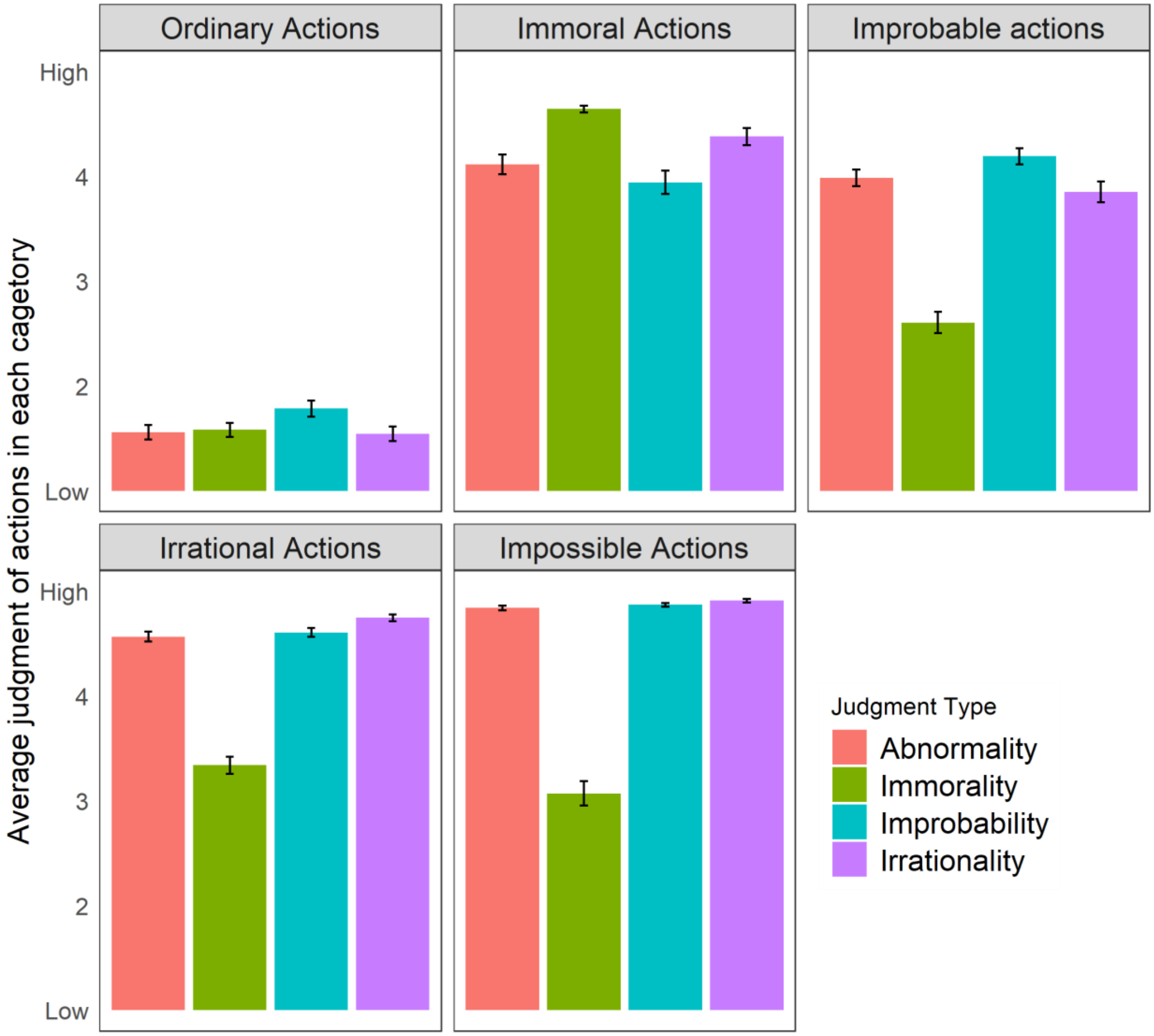

Figure 1. Averages of the judgment scores for each action within the five different categories of actions (depicted in separate panels) for each of the four different judgments (separate bars within each panel). Error bars depict +/- 1 SEM.

Default representations of permissibility

We first asked whether moral judgments became more similar to one another when participants were forced to respond quickly, which would provide evidence for a default form of moral judgment. We began by following the method in Phillips & Cushman (2017), in which we calculated the average judgment for each of the 5 different moral judgments for each of the 240 actions, doing so separately for fast and slow moral judgments. In other words, for each action, we calculated the proportion of participants who judged that that action *should* be done, or who judged that it was morally *acceptable*, and so on, for each of the 5 moral judgments, and separately calculated these scores when participants were making these judgments quickly vs. slowly. We can then estimate the similarity of different moral judgments to each other by simply calculating the item-wise correlation between pairs of moral judgments. To get a sense for why this is a good estimate of similarity, it should be obvious that this approach would result in a very high correlation between two samples of participants who were asked the *same* question. On average, they are very likely to judge the various actions in the same way, so one will observe a very high item-wise correlation between the two sets of judgments. Critically, for our purposes, this approach also allows us to answer a more interesting question of whether these judgments become more or less similar to one another when participants are answering quickly vs. slowly.

We first considered moral judgments of all actions across the 5 of the categories of actions that participants evaluated (immoral, impossible, improbable, ordinary, and irrational). We calculated the item-wise correlations between all ten unique, non-reflexive pairs of different moral judgments. Unsurprisingly, we found that the average level of correlation were high when participants were responding slowly (r_{Mean} = 0.885), but critically, we also found that it the correlations were significantly higher when participants were answering quickly (r_{Mean} = 0.970), $t(9) = 10.665$, $p < .001$, $d = 3.37$ see Fig. 2.

We next focused only on the category of immoral actions, and performed the same analysis, allowing us to ask whether this same pattern is obtained when only considering actions that are clearly immoral. We found that as a group, moral judgments were again quite similar to one another when these judgments were made quickly (r_{Mean} = 0.829); however, when participants reflected before responding, their moral judgments became much less similar to one another (r_{Mean} = 0.226), $t(9) = 9.742$, $p < .001$, $d = 3.08$. These findings suggest a common default form of moral judgment that is differentiated with additional processing. This interpretation is bolstered by the fact that the pairs of judgments that become most dissimilar on reflection are ones that are intuitively quite different from each other, e.g., whether an action is 'acceptable' and whether one 'should' do that action. Notably, however, when participants are forced to answer quickly, such judgments become markedly similar.

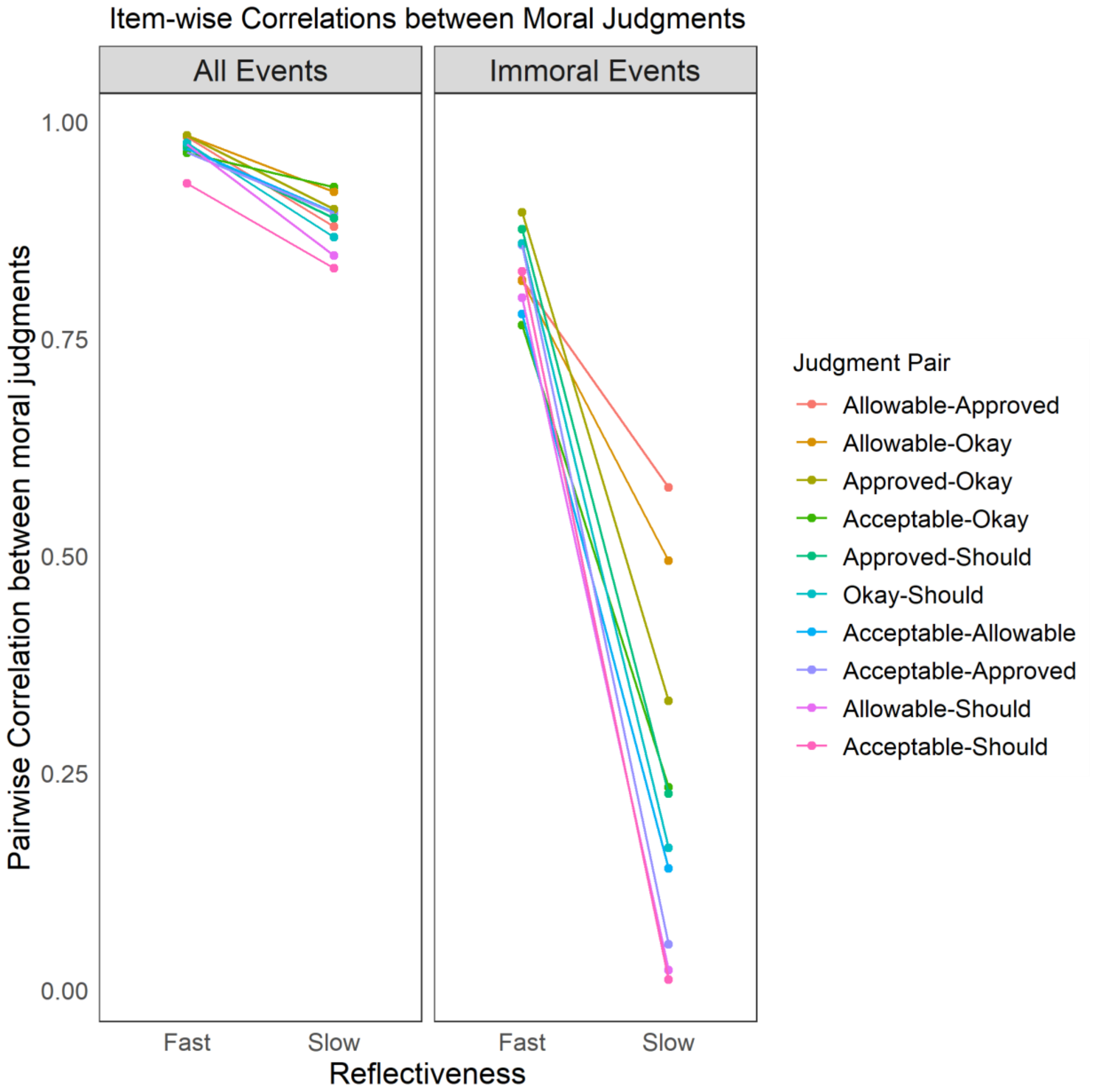

Figure 2. Pairwise correlations for all 10 unique, non-reflexive pairs of different moral judgments (depicted by colored lines) made either quickly (left points) or slowly (right points). The legend is arranged in descending order of the correlation for reflective judgments of immoral actions.

Characterizing default representations of permissibility

We next wanted to characterize how default representations of permissibility differ from more reflective representations. To do so, we began by calculating the average overall permissibility rating for each action both when participants responded quickly and when they responded slowly. To illustrate, consider the improbable action, ‘discover a nut-free cake in the house’ from the scenario in which Emily has accidentally baked a cake that contains nuts (see Methods for complete text). In the previous analyses, we calculated the average judgment for whether this action was morally *acceptable*, should be *allowed*, would be *approved* of, would be *okay* to do, or *should* be done when participants were answering quickly or slowly. In this analysis, we now take the average of these judgments when participants were answering quickly and when participants were answering slowly. These averages represent the overall fast and slow permissibility ratings for each action. These averages were calculated for each action in each action category separately (ordinary actions, immoral actions, irrational actions, impossible actions). Accordingly, this approach allows us to characterize how making participants respond quickly changed their judgments of different kinds of actions.

We observed clear effects of response speed on judgments of moral permissibility that differed for different action categories (see Figure 3). At a qualitative level, forcing participants to respond quickly mostly clearly changed their permissibility judgments by making them more inclined to judge that improbable, irrational, and physically impossible actions were *not* permissible. The differential signature effects of response speed on judgments of permissibility can be quantitatively demonstrated by asking whether we find an interaction between response speed (fast vs. slow) and action category. We investigated this possibility by comparing a series of linear mixed-effects models using the lme4 package in **R**, and using the anova function for model comparison (Bates, Maechler, Bolker, Walker, et al., 2014).[3] We found a highly significant interaction, $\chi^2(4) = 28.763$, $p < .001$. This interaction effect, which can be visually seen in Figure 3, is critical as it suggests that increasing response speed does not have an overall effect of simply increasing noise or simply increasing a tendency to judge events as impermissible.

[3] In this approach we build a model that includes the two-way interaction between response speed and action category and all main effects. The structure of the data does not allow us to use a fully maximal random effects structure (Barr, Levy, Scheepers, & Tily, 2013) as we have only 2 values for each action (an average fast response and an average slow response) and thus cannot estimate the random slopes for response speed for each action. Accordingly, we include a random intercept for each action. Using this same random effects structure, we next build a second model that does not include the two-way interaction, but does include all main effects. We then compare the fit of these two models as a way of testing for the significance of the two-way-interaction.

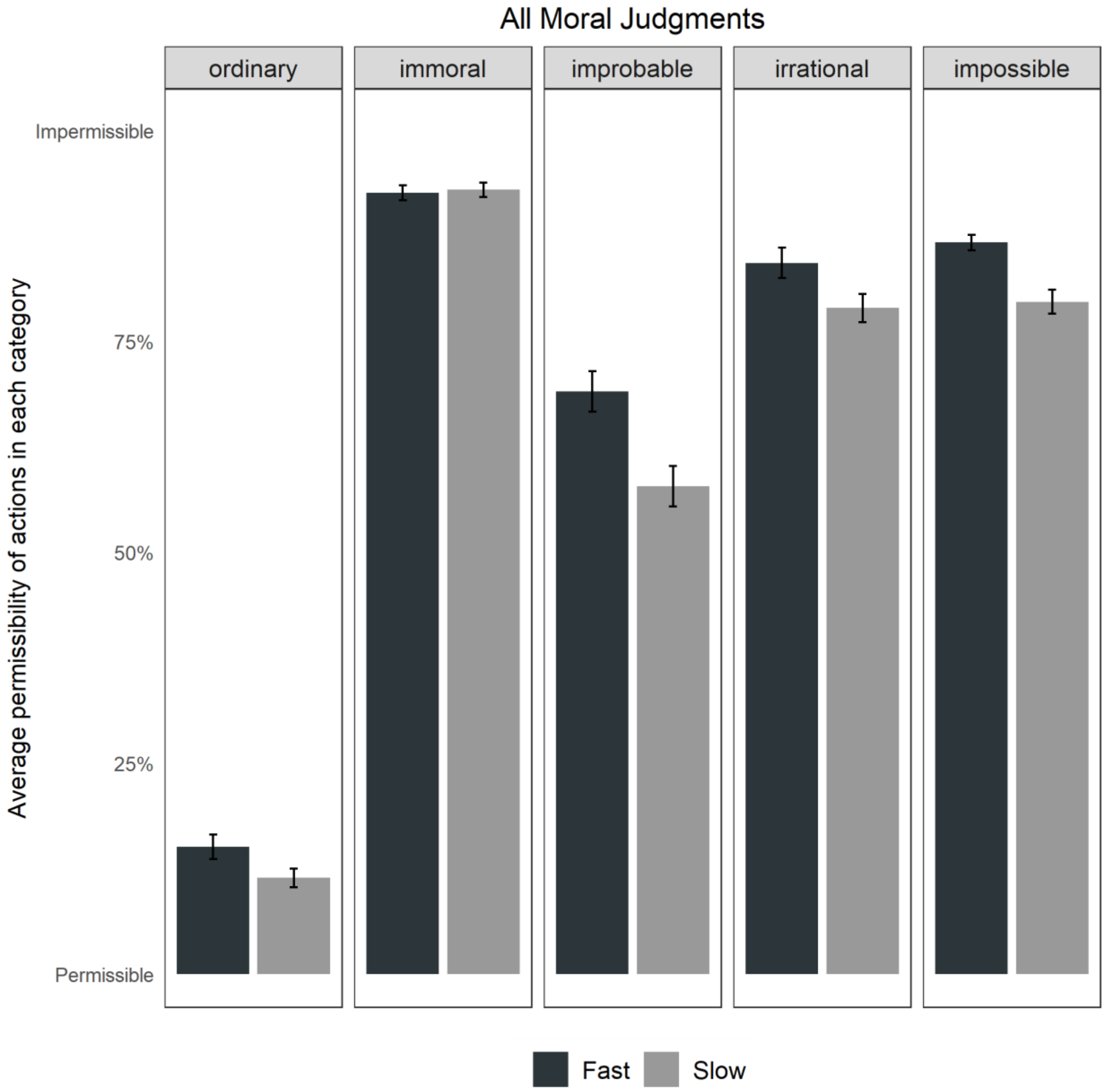

Figure 3. Average moral judgment for each action category (separated into different panels) when moral judgments were made quickly (black bars) or slowly (gray bars). Error bars depict +/-1 SEM.

We next decomposed this interaction through a series of pairwise comparisons using the emmeans package (Lenth, 2019) to quantitatively capture the qualitative patterns noted above and depicted in Figure 3.[4] We found that faster response speed increased the tendency to judge ordinary actions as *not* permissible to a small extent, *t-ratio*(*df*=233) = 2.351, *p* = .020. There was no significant impact on judgments of immoral events, *t-ratio*(*df*=233) = -0.230, *p* = .818.

[4] Here, we make pairwise comparisons based on response speed for each action category, i.e., pairwise ~ response_speed | action_category.

However, faster response speed increased the tendency to judge improbable, irrational, and physically impossible actions as *not* permissible (improbable actions *t-ratio*(*df*=233) = 7.173, $p < .001$; irrational actions *t-ratio*(*df*=233) = 3.436, $p < .001$; impossible actions *t-ratio*(*df*=233) = 4.511, $p < .001$). In sum, we found clear evidence that fast judgments of an action's permissibility reveal a default representation that tends to exclude improbable, irrational, and impossible actions.

Recall that prior work on default representations of possibility suggests that improbable, irrational, immoral, and physically impossible actions all tend to be treated as impossible by default (Phillips & Cushman, 2017; Phillips & Knobe, 2018; Shtulman & Phillips, 2018). Here, given that we see the same factors influencing default judgments of permissibility, a potential explanation for the pattern uncovered above is that default representations permissibility may reflect default representations of whether that action is possible. This is the hypothesis we pursue more directly next.

Default representations of possibility

Intuitively, for it to be the case that you should do an action, it also must be the case that it is *possible* for you to do that action. And thus, if by default you treat an action as *not* possible, then by default that action should not be treated as one that should be done. We can leverage this connection between what is possible and what should be done to investigate whether the change in moral judgments we observed when participants were forced to respond quickly was due to an increased reliance on default representations of possibility. That is, we can ask whether the observed changes observed in fast moral judgments could be explained by an increased reliance on default representations of possibility.

To characterize default representations of possibility, we calculated an average *possibility* rating for each of the 240 actions when participants responded quickly (<1550 ms) vs. slowly (>1550 ms). As an initial step, we then asked whether we replicate the patterns observed in Phillips and Cushman (2017). As in that prior work, we found a highly significant interaction between the speed with which possibility judgments were made and the kind of action being judged, $\chi^2(4) = 45.82$, $p < .001$. We decomposed this interaction with pairwise comparisons, and also found that we replicated the prior finding that faster response speed increased the tendency to judge immoral and irrational actions as *not* possible (immoral actions *t-ratio*(*df*=235) = 5.873, $p < .001$; irrational actions *t-ratio*(*df*=235) = 5.307, $p < .001$) (see Figure 4).

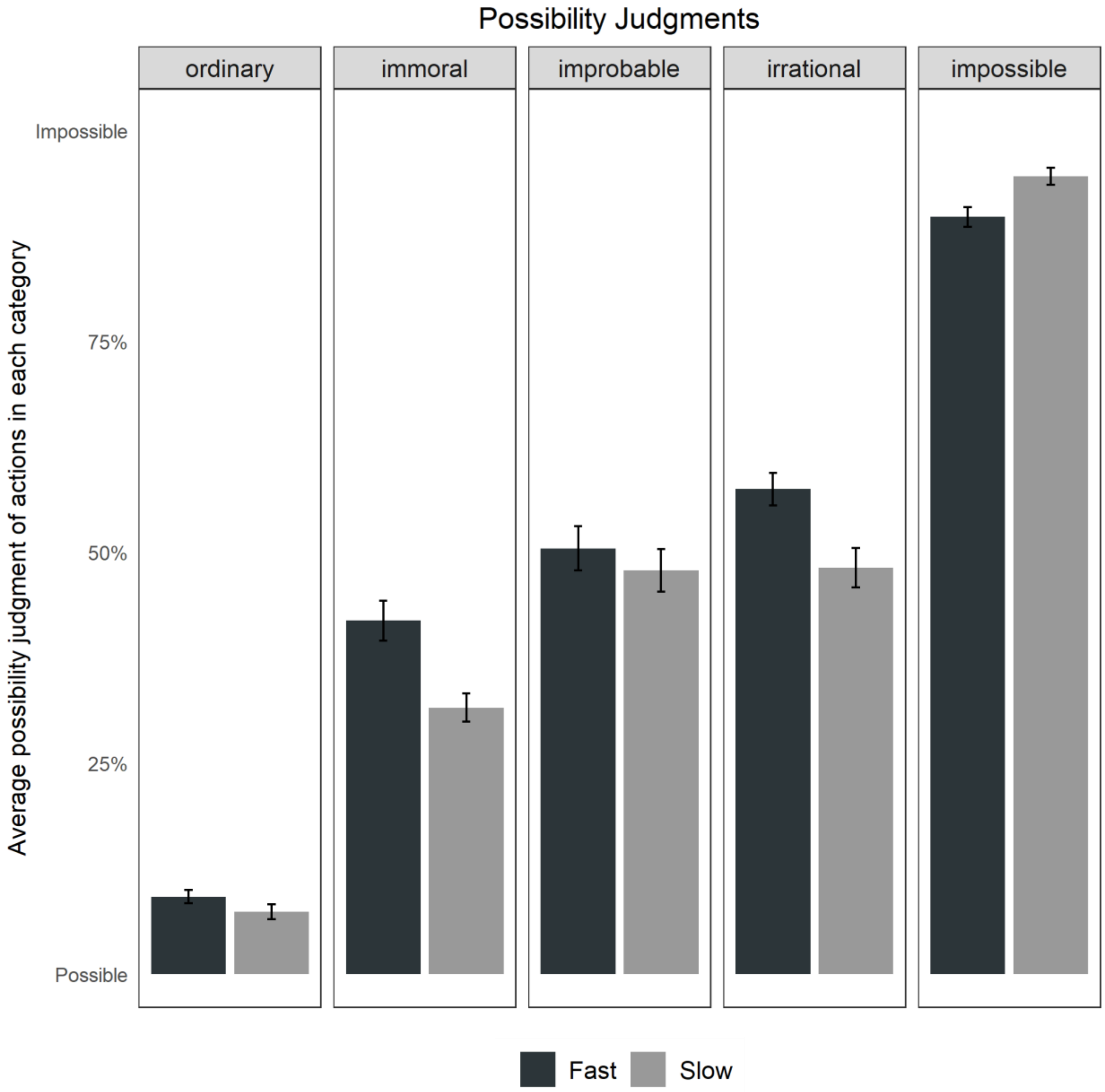

Figure 4. Average possibility judgment of actions in each category (separated into different panels) both when possibility judgments were made quickly (black bars) or slowly (gray bars). Error bars depict +/-1 SEM.

Relationship between permissibility and possibility

We next investigated the correlation between these possibility judgments and participants' prior moral judgments about these same actions. We found that all moral judgments, both those made quickly and those made slowly, were more highly correlated with *speeded* judgments of possibility than reflective judgments of possibility, *t*(9) =11.055, *p* < .001, *d* = 3.50. Second, focusing specifically on the relationship with fast possibility judgments, we found that all five fast moral judgments were more highly correlated with default representations of possibility (0.676 < *r* < 0.797) than slow permissibility judgments (0.528 < *r* < 0.708), see Figure 5.

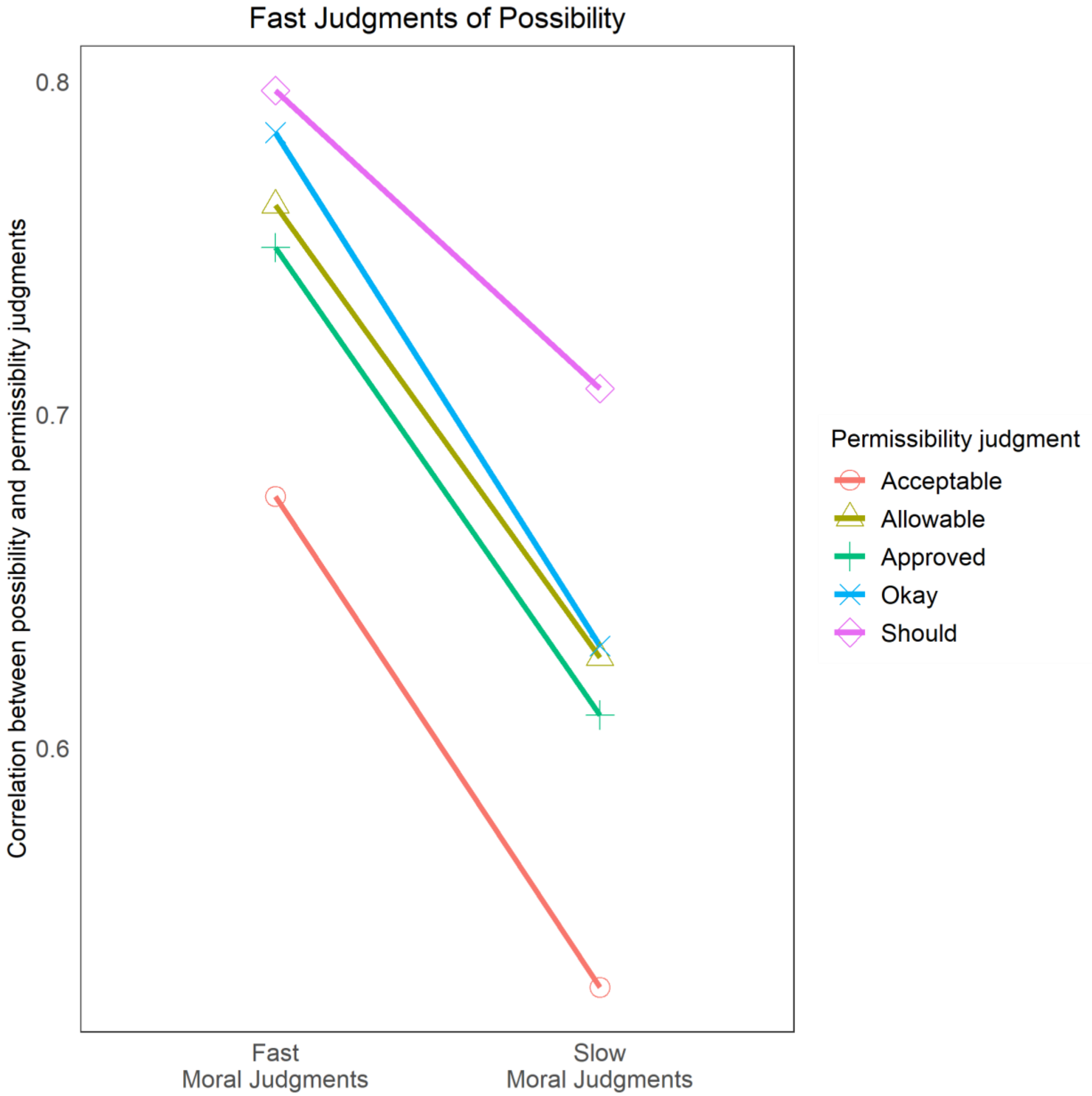

Figure 5. Item-wise correlations between fast judgments of possibility and judgments of moral permissibility (differently colored lines) when these judgments were made quickly (left shapes) or slowly (right shapes).

To get a sense for the robustness of this relationship within different action categories, we then calculated the correlation between speeded possibility judgments and each moral judgment (‘should’, ‘okay’, etc.) independently for each action type (ordinary, immoral, etc.). Figure 6, for example, depicts the relationship between fast judgments of whether an action would *not* be 'okay' to do and fast judgments of whether that action is impossible both overall and for each action category separately.

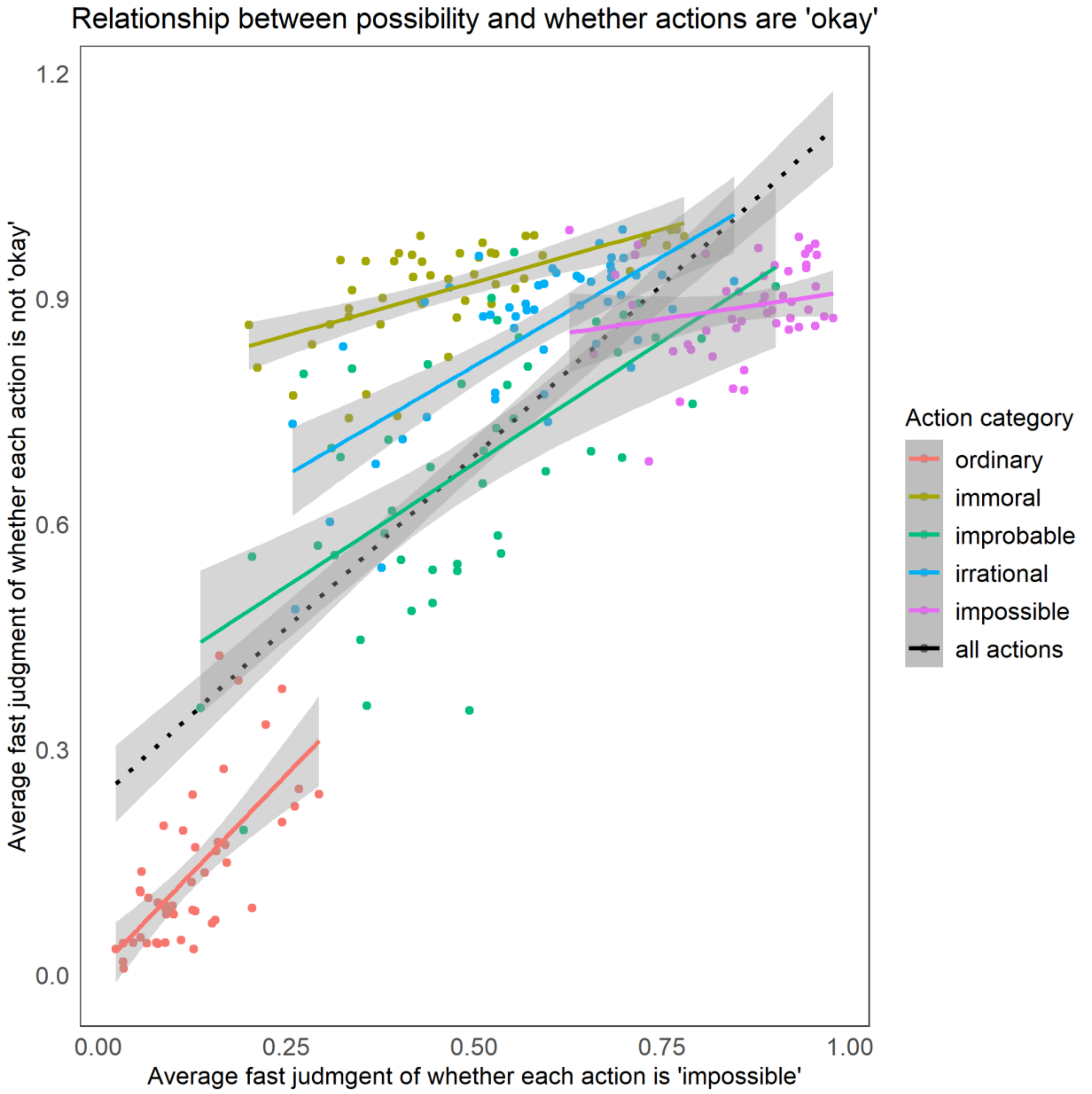

Figure 6. Relationship between speeded judgments of whether an action is not 'okay' and fast judgments of whether an action is impossible both within each separate category of actions (solid-colored lines) and overall across categories (dotted black line).

These analyses help illustrate the more general trend that *fast* permissibility judgments closely reflect default representations of whether an action is possible. In comparison, slow moral judgments much less closely reflect possibility judgments within each action type; Figure 7 depicts this difference. Within different action categories, all five fast moral judgments were almost always significantly correlated with speeded possibility judgments (21 of 25 judgment-type-by-action-category conditions), and only did not exhibit a significant correlation for actions that were physically impossible. By contrast, slow moral judgments were typically not correlated with possibility judgments within different action categories (11 of 25 judgment-type-by-action-category conditions), and never exhibited a significant correlation for actions that were categorized as improbable or irrational. This difference in the number of significant relationships with speeded possibility for fast vs. slow moral judgments was statistically significant, $\chi^2(1) = 7.03$, $p = .008$, $V = 0.375$.

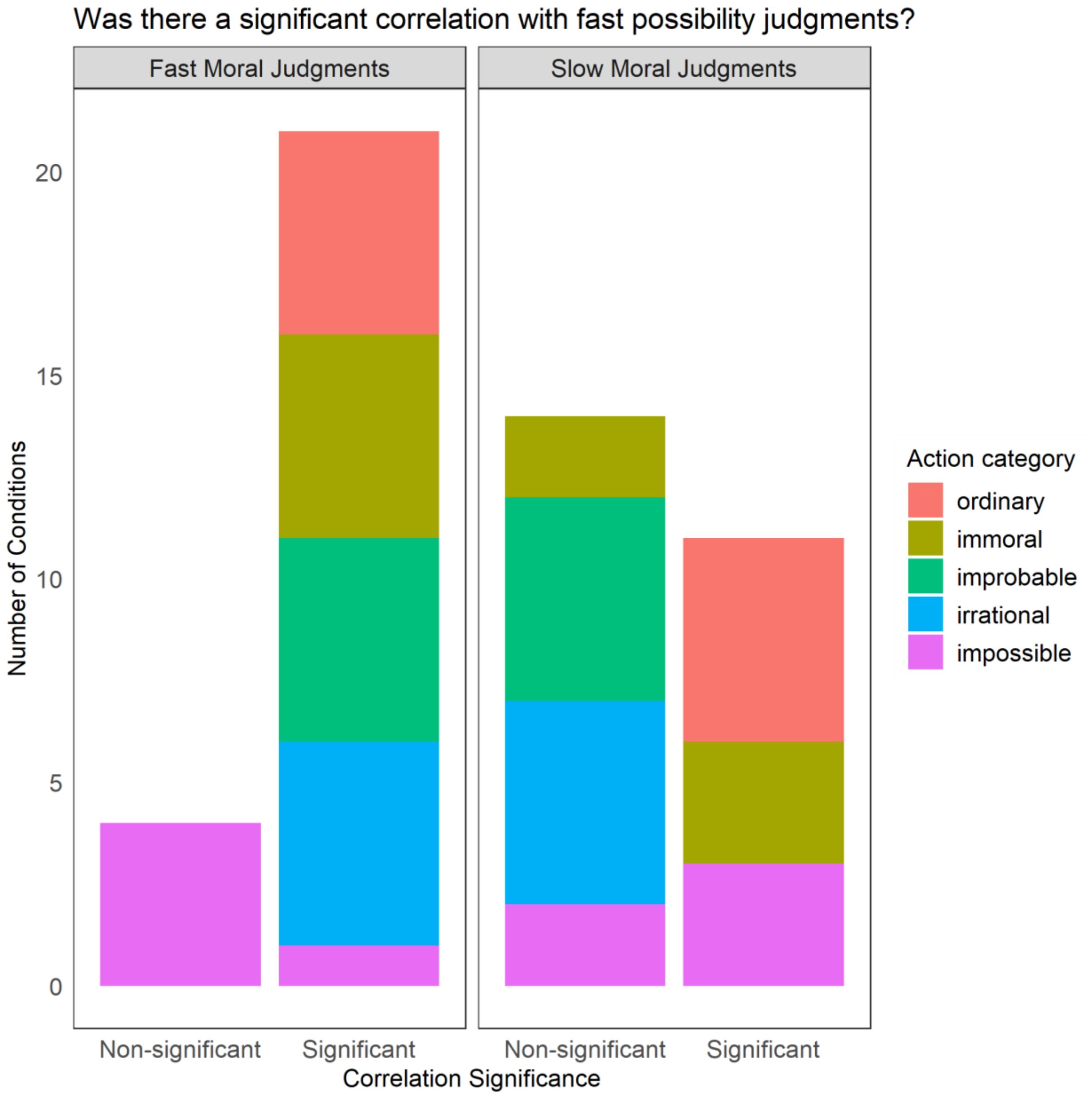

Figure 7. Stacked bar graph depicting the number of moral-judgments by action-category conditions with significant vs. non-significant correlations with speeded possibility judgments,

both for fast moral judgments (left panel), and slow moral judgments (right panel). Differently colored portions of the bars indicate different action categories.

In brief summary, we find clear evidence that default representations of permissibility reflect default representations of possibility, but as additional time is taken before making a moral judgment, we see a shift away from this shared default and toward the individual meanings of each permissibility judgments, which are each more moderately related to default representations of what is possible.

Discussion.

It has long been noted that there is a connection between judgments of whether someone ought to do a given action and judgments of whether they can do that action. When deciding whether an agent 'ought to' or 'should' do an action, one must first consider what options are available to the agent. And for actions that are not possible for that agent to do, it is also not the case that they 'should' or 'ought to' do them (Sidgwick, 1884; Parfit, 1984; Zimmerman, 1996).

Although moral judgments, such as whether someone should do an action, require representing and reasoning over possibilities, we are often able to make moral judgments quickly and seemingly effortlessly. This suggests that moral judgments may recruit default representations of possibility rather than deliberative assessments of what it is possible for an agent to do in a given situation. The present research investigated this possibility by comparing moral judgments that are made quickly vs. slowly, and then asking whether and how they recruit default representations of possibility. We found that when moral permissibility judgments are made quickly, they converge on a common default representation, and that this default representation tends to exclude improbable, irrational, and physically impossible actions. We also found that speeded moral judgments in particular reflected default representations of what is possible.

An important, but thus far unaddressed, question is whether the patterns we observed could be accounted for by some simpler explanation. For example, could it be that our time pressure manipulation simply resulted in an increased tendency either (i) for participants to respond randomly or (ii) for participants to respond in the negative (e.g., saying that an action is *not* allowed or *not* possible)?[5] While we would not want to argue that such effects could not be occurring in our task, we do think it is unlikely that they could explain the key patterns on which we base our argument concerning the relationship between judgments of permissibility and possibility. First, consider the effect of time pressure on the overall agreement ratings that various kinds of actions are permissible (Fig. 3) or possible (Fig. 4). If time pressure simply increases the randomness of participants' responses, we should expect that fast responses will uniformly move closer toward the midpoint of the y-axis. On the other hand, if time pressure simply increases the tendency to respond negatively, we should expect that fast responses will uniformly move toward the top of the y-axes. We don't find clear evidence for either pattern. In the case of permissibility judgments, for example, we find that increasing time pressure actually moves responses further from the midpoint, not closer. Moreover, for both permissibility and

[5] We'd like to thank Reviewer 2 for raising the second of these concerns.

possibility judgments, the effect of time pressure is not uniform across different categories of actions. Instead of uniformly increasing the tendency to judge events as impermissible or impossible, we find that time pressure selectively increases this tendency for some kinds of actions but not others: immoral actions are not judged to be more unacceptable under time pressure, and physically impossible events are actually judged to be *more* possible under time pressure (see Phillips & Cushman, 2017 for a similar pattern). Thus, the mean changes observed under time pressure, and especially the interaction effects between time pressure and action category which are central for our conclusions, are unlikely to be fully accounted for by these alternative explanations.

Importantly, more unequivocal evidence against these alternative explanations can be seen in the effect of time pressure on the similarity between different kinds of moral and modal judgments. If time pressure increases the randomness of responding, one should expect an overall decrease in the correlations between judgments, as the true correlation between two completely random variables is, obviously, 0. On the other hand, if time pressure increases the tendency to answer negatively, one should expect no change in the correlation between different kinds of moral and modal judgments, as such a change will not have a systematic effect on the level of covariation across judgments but simply the mean level of agreement or disagreement of each judgment. In clear contrast to these predictions, we find that time pressure (i) *increases* the correlation between all permissibility judgments (Fig. 2), (ii) *increases* the correlation between judgments of possibility and each moral acceptability judgment (Fig. 5), and (iii) even increases the correlation between possibility and permissibility judgments within almost all action-category-moral-judgment pairs (Fig. 7). In short, it is not clear how the key correlational results we use to support our conclusion can be accounted for by either an increased tendency to respond randomly or in the negative.

Stepping back, our findings contribute to the growing literature that has demonstrated a relationship between moral and modal cognition. Most of this prior work has demonstrated the way in which moral judgments affect how people reason about what is possible. For example, prior work has shown that moral judgments influence perceptions of the relevance of alternative possibilities, such that people think it is more relevant to consider morally good possibilities than morally bad ones (Phillips, Luguri, & Knobe, 2015). And work on default representations of possibility finds that a signature feature of default possibility representations is that they tend to exclude immoral actions (Phillips & Cushman, 2017; Phillips & Knobe, 2018; Phillips, Morris, Cushman, 2019). Indeed, this relationship has been found to exist early in development, as young children's judgments of what is possible or could be done are constrained by value (Kushnir, Gopnik, Chernyak, Seiver, & Wellman, 2015), and morality (Phillips & Bloom, 2017; Shtulman & Phillips, 2018).

In contrast to this prior work, we instead investigated whether default moral judgments are constrained by perceptions of what is possible. We found that permissibility judgments made under time pressure do reflect the signature features of default representations of possibility. For example, people exhibit an increased tendency to judge that it is immoral to do actions that are unlikely to be done. Further, we found that moral permissibility judgments, especially when made quickly, reflect whether or not these actions are regarded as possible by default. Along with prior developmental work (Shtulman & Phillips, 2018), these findings demonstrate that our representations of what is morally permissible are constrained by our default ideas about what is possible.

We hope that these findings provide a useful first step toward understanding the role of default representations of possibility in moral cognition, and spur continued research on the connection between default representations of morality and possibility.

Acknowledgements: We would like to thank Annie Politi and Isabel Morris for their help in the collection of data and conceptualization of the studies. We'd also like to thank members of Phillab for discussions throughout the project.

References

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language,* 68(3), 255-278.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bicchieri, & Xiao, E. (2009). Do the right thing: but only if others do so. *Journal of Behavioral Decision Making*, *22*(2), 191–208. https://doi.org/10.1002/bdm.621

Byrne, & Timmons, S. (2018). Moral hindsight for good actions and the effects of imagined alternatives to reality. *Cognition*, *178*, 82–91. https://doi.org/10.1016/j.cognition.2018.05.010

Byrne, R. M. J. (2017). Counterfactual Thinking: From Logic to Morality. Current Directions in Psychological Science, 26(4), 314–322. https://doi.org/10.1177/0963721417695617

Casler, Terziyan, T., & Greene, K. (2009). Toddlers view artifact function normatively. *Cognitive Development*, *24*(3), 240–247. https://doi.org/10.1016/j.cogdev.2009.03.005

Cialdini, Reno, R. R., & Kallgren, C. A. (1990). A Focus Theory of Normative Conduct: Recycling the Concept of Norms to Reduce Littering in Public Places. *Journal of Personality and Social Psychology*, *58*(6), 1015–1026. https://doi.org/10.1037/0022-3514.58.6.1015

Kant, I. (1781/1929). *Critique of Pure Reason*. (N. Kemp Smith, Trans.). London: Macmillan.

Kushnir, T., Gopnik, A., Chernyak, N., Seiver, E., & Wellman, H. M. (2015). Developing intuitions about free will between ages four and six. Cognition, 138, 79-101.

Lenth, R. V. (2019). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.3.3. Retrieved from https://CRAN.R-project.org/package=emmeans

Lagnado, D. A., Gerstenberg, T., & Zultan, R. (2013). Causal responsibility and counterfactuals. *Cognitive Science*, *37*(6), 1036–1073. https://doi-org.dartmouth.idm.oclc.org/10.1111/cogs.12054

Peysakhovich, & Rand, D. G. (2016). Habits of Virtue: Creating Norms of Cooperation and Defection in the Laboratory. *Management Science*, *62*(3), 631–647. https://doi.org/10.1287/mnsc.2015.2168

Parfit, D. (1984). *Reasons and Persons*. Oxford: OUP.

Phillips, J., & Cushman, F. (2017). Morality constrains the default representation of what is possible. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, *114*(18), 4649–4654. https://doi-org.dartmouth.idm.oclc.org/10.1073/pnas.1619717114

Phillips, J., & Knobe, J. (2018). The psychological representation of modality. *Mind & Language*, *33*(1), 65–94. https://doi-org.dartmouth.idm.oclc.org/10.1111/mila.12165

Phillips, J., Morris, A., & Cushman, F. (2019). How we know what not to think. *Trends in Cognitive Sciences*, *23*(12), 1026–1040. https://doi-org.dartmouth.idm.oclc.org/10.1016/j.tics.2019.09.007

Phillips, J. & Bloom, P. (2017). Do children believe immoral events are impossible? Unpublished Manuscript, Dartmouth College.

Phillips, Luguri, J. B., & Knobe, J. (2015). Unifying morality’s influence on non-moral judgments: The relevance of alternative possibilities. *Cognition*, *145*, 30–42. https://doi.org/10.1016/j.cognition.2015.08.001

Rakoczy, Warneken, F., & Tomasello, M. (2008). The Sources of Normativity: Young Children’s Awareness of the Normative Structure of Games. *Developmental Psychology*, *44*(3), 875–881. https://doi.org/10.1037/0012-1649.44.3.875

Roberts, Gelman, S. A., & Ho, A. K. (2017). So It Is, So It Shall Be: Group Regularities License Children’s Prescriptive Judgments. *Cognitive Science*, *41*(Suppl 3), 576–600.

https://doi.org/10.1111/cogs.12443

Shou, Y., Olney, J., Smithson, M., & Song, F. (2020). Impact of uncertainty and ambiguous outcome phrasing on moral decision-making. *PLoS ONE*, *15*(5). https://doi-org.dartmouth.idm.oclc.org/10.1371/journal.pone.0233127

Shtulman, A., & Phillips, J. (2018). Differentiating "could" from "should": Developmental changes in modal cognition. *Journal of Experimental Child Psychology*, *165*, 161–182. https://doi-org.dartmouth.idm.oclc.org/10.1016/j.jecp.2017.05.012

Shtulman, A., & Tong, L. (2013). Cognitive parallels between moral judgment and modal judgment. *Psychonomic Bulletin & Review*, *20*(6), 1327–1335. https://doi-org.dartmouth.idm.oclc.org/10.3758/s13423-013-0429-9

Sidgwick, H. (1884). *The Methods of Ethics*. London: Macmillan and Co.

Tepe, B., & Byrne, R. (2022). Cognitive processes in imaginative moral shifts: How judgments of morally unacceptable actions change. *Memory & cognition*, *50*(5), 1103–1123. https://doi.org/10.3758/s13421-022-01315-0

Zimmerman, M. (1996). *The Concept of Moral Obligation*. Cambridge: Cambridge University Press.