What we owe to family:

The impact of special obligations on moral judgment

Ryan M. McManus and Liane Young

Department of Psychology,
Boston College, Chestnut Hill, MA 02467

Author contributions: R.M. McManus developed the study concept. Both R.M. McManus and L. Young contributed to the study design. Data were collected and analyzed by R.M. McManus. R.M. McManus drafted the manuscript and L. Young provided critical feedback. Both authors contributed to revisions and approved the final version of the manuscript for submission.
Abstract

Although impartiality is often moralized, it is unclear under what conditions favoring specific others is similarly moralized. Across four pre-registered studies (N = 1,011), people modulated their moral judgments according to whether a special obligation was perceived as present. In Studies 1 (N = 209) and 2 (N = 193), agents who helped strangers were judged more positively than those who helped kin, but agents who helped strangers instead of kin were judged less positively than those who helped kin instead of strangers. In Studies 3 (N = 304) and 4 (N = 305), agents who failed to help strangers were judged less negatively than agents who failed to help kin. Study 4 also demonstrated that perceived obligations being violated versus fulfilled corresponded to and predicted moral judgments. Overall, these findings suggest that moral judgments are sensitive to properties of special relationships, as in the case of kinship.
Introduction

Peter Singer, an inspiring altruist, is well recognized for his philosophical arguments that we have a moral obligation to value close loved ones and distant strangers similarly (or at least not as differently as most of us do). Singer currently donates 40% of his income to strangers through high-efficacy charities to reduce suffering around the world (Bearne, 2017). When his mother developed Alzheimer’s, though, he partially eschewed his philosophy and spent more money caring for her than his own moral arguments permitted. Singer remarked on his violation of principle, “Perhaps it is more difficult than I thought before, because it is different when it’s your mother” (MacFarquhar, 2015). On the one hand, Singer might have been judged positively for choosing to help his mother instead of continuing to help the same number of strangers because he was fulfilling a special obligation to family. On the other hand, Singer might be judged positively for helping strangers precisely because obligation is absent.

While special obligations have received some attention in philosophy (e.g., Jeske, 1998; Sommers, 1986), moral psychology has largely neglected that our day-to-day lives are mostly spent with close others rather than unrelated strangers (Bloom, 2011), and therefore has also lagged in characterizing perceptions of obligation (Bartels, Bauman, Cushman, Pizarro, & McGraw, 2016). Existing frameworks, like Moral Foundations Theory (Graham, Nosek, Haidt, Iyer, Koleva, & Ditto, 2011) and Relationship Regulation Theory (RRT; Rai & Fiske, 2011), have noted the importance of socio-functional dynamics in morality (e.g., in-group loyalty and unity). Importantly, supporting RRT, perceived moral relevance is higher when a relational component is present versus absent, such as burning someone else’s arm versus burning one’s own arm (Tepe & Aydinli-Karakulak, in press). However, to our knowledge, neither framework has measured perceived obligations to specific others or their relation to moral judgment. In this
paper, we provide evidence that, in everyday (non-life-threatening) contexts, people indeed recognize a positive obligation to help kin more than non-kin, and these perceptions influence moral judgments.

Research on helping intentions suggests that people generally favor kin over non-kin. In life-threatening situations, people more often report intending to help a sibling responsible for her plight over a non-responsible acquaintance (Greitemeyer, Rudolph, & Weiner, 2003). Moreover, when someone needs help, genetic relatedness between the potential helper and beneficiary increases helping intentions (Burnstein, Crandall, & Kitayama, 1994), and people will even sustain uncomfortable physical exercise for longer to reward close versus distant genetic relatives (Madsen et al., 2007). Relatedly, people forego hypothetical money for themselves to benefit others as social closeness increases (Strombach et al., 2015). People are even sensitive to these distinctions in the context of moral dilemmas, being more willing to sacrifice one brother to save five brothers than one stranger to save five strangers (Kurzban, DeScioli, & Fein, 2012). And, when contemplating hypothetical crimes, adults are less willing to report a responsible brother than a responsible stranger, regardless of the (un)ambiguity of their culpability (Lee & Holyoak, 2018). Even children and adolescents judge socially close perpetrators (their mothers) as less likely to recidivate than distant perpetrators (foreign visitors; Linke, 2012).

From a third-person perspective, however, it remains unclear, in helping contexts, whether (a) the relationship between helper and beneficiary influences moral judgments, and (b) variation in obligation perceptions explain those judgments. A relevant evolutionary perspective posits that “good” people are perceived as those who confer benefits at a great cost without any apparent return benefit to themselves (Duntley & Buss, 2016). Consider one scenario in which an
unrelated stranger needs help, and another in which kin needs help. Agents who help strangers should be judged more positively than agents who help kin, precisely because people have no special obligations to strangers. Conversely, when a choice must be made between two people (who differ in relatedness to the helper), because an obligation to kin will be present (Hughes, 2017), agents should be judged more positively for fulfilling this obligation by helping kin instead of stranger (versus stranger instead of kin). Last, consider two cases in which agents simply fail to help. Agents who fail to help kin should be judged less positively (or more negatively) than agents who fail to help strangers, as the former are violating a special obligation (see Haidt & Baron, 1996).

Across four studies, we tested these predictions by manipulating the helper-beneficiary relationship and whether another potential beneficiary (whose relationship to the helper differed) was present. Importantly, our final study also explicitly measured perceptions of obligations being violated versus fulfilled and tested whether these perceptions predicted differences in moral judgments.

**Study 1**

**Method**

**Participants**

Participants were 234 United States residents recruited and compensated via Amazon’s Mechanical Turk. This sample size was chosen to account for exclusions and reach an analyzable sample of $N = 200$. Participants who failed to correctly answer two scenario-relevant memory questions ($n = 25$) were excluded from analyses, resulting in a final $N = 209$ (40.2% female, $M_{Age} = 34.75$).

**Materials and Procedure**
Participants read eight stories in which prosocial agents engaged in helping behavior. We varied the relationship between helper and beneficiary (Relationship: Stranger vs. Kin) and whether another potential beneficiary could have been helped instead (Choice: No Choice vs. Choice). Participants read two stories that each corresponded to “Stranger, No Choice,” “Stranger, Choice,” and so on (see Table 1 for a shortened scenario example across conditions for all studies). Thus, the design of the study was a fully crossed 2 (Relationship) x 2 (Choice) within-subjects design in which participants saw all eight stories but never the same story across conditions (which was true across studies).

Table 1
Shortened scenario example across all conditions. Manipulations are underlined here but were not underlined for participants. In Studies 1-2, participants saw only No Choice and Choice conditions. In Studies 3-4, participants saw all six conditions.

<table>
<thead>
<tr>
<th></th>
<th>Stranger</th>
<th>Kin</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Choice</td>
<td>A new tenant is moving into an apartment down the hall from John. She is a stranger. John helps his new neighbor move their furniture in.</td>
<td>A new tenant is moving into an apartment down the hall from John. She is his cousin whom he has not seen or spoken to in years. John helps his cousin move their furniture in.</td>
</tr>
<tr>
<td>Choice</td>
<td>Two new tenants are moving into two separate apartments down the hall from John. One new tenant is a stranger. The other is his cousin whom he has not seen or spoken to in years. Rather than helping his cousin, John helps his other new neighbor move their furniture in.</td>
<td>Two new tenants are moving into two separate apartments down the hall from John. One new tenant is a stranger. The other is his cousin whom he has not seen or spoken to in years. Rather than helping the stranger, John helps his cousin move their furniture in.</td>
</tr>
<tr>
<td>Failed to Help</td>
<td>A new tenant is moving into an apartment down the hall from John. She is a stranger. Rather than helping the stranger move their furniture in, John stays inside his apartment and plays video games.</td>
<td>A new tenant is moving into an apartment down the hall from John. She is his cousin whom he has not seen or spoken to in years. Rather than helping his cousin move their furniture in, John stays inside his apartment and plays video games.</td>
</tr>
</tbody>
</table>

Note. Full (un-shortened) texts of all scenarios, across studies, are available at https://osf.io/bnwdv/.

Across all studies, genetic relatives ranged from 1.56% (e.g., second cousin’s child) to 12.50% related (e.g., cousin) to the target agent. To keep relatives otherwise stranger-like (isolating the effect of relatedness), they were always described as someone who had either never been met in person or had not been seen or spoken to in some time. Thus, the tested effects of
perceived obligation on moral judgment are arguably very conservative and may indeed be stronger if relatedness or familiarity were higher (e.g., comparing an agent-stranger interaction to a parent-offspring interaction). Full scenario texts, traditionally formatted data and output, mixed effects data, R code and output, and by-scenario/by-item breakdowns for all studies are available at https://osf.io/bnwdv/. The design, hypotheses, exclusion criteria, and analyses for Study 1 were all pre-registered at http://aspredicted.org/blind.php?x=f9au26.

After reading each scenario, participants answered questions assessing the agent’s moral goodness and trustworthiness (1 = “extremely bad/untrustworthy” to 7 = “extremely good/trustworthy”)\(^1\), as these characteristics appear most important in person perception (Brambilla, Carraro, Castelli, & Sacchi, 2019; Cottrell, Neuberg, & Li, 2007; Goodwin, 2015; Goodwin, Piazza, & Rozin, 2014).

**Results**

Descriptive statistics for morality and trustworthiness are shown in Figure 1 (for graphing purposes, we show “neither bad nor good; untrustworthy nor trustworthy,” originally 4s on the 1-7 scale, at the 0 midpoint). Traditional repeated-measures analyses (including effect sizes) for all variables, across studies, are reported in Supplemental Online Material (henceforth, SOM). As pre-registered, data were analyzed with linear mixed effects models, specified to predict moral judgments from the fixed effects of Relationship (Stranger vs. Kin) and Choice (No Choice vs. Choice), their interaction, and the random intercepts of Subject and Scenario. The random intercepts model allows for generalization of results to other stimuli and individuals (Judd, Westfall, & Kenny, 2012). All pre-registered comparisons and corresponding statistics, across Studies 1-2, are reported in Table 2 at the end of Study 2 Method/Results.
Moral Goodness. As predicted, there was a significant interaction between Relationship and Choice, $B = 0.640, SE_B = 0.094, t = 6.84, p < .001$; this interaction model provided a better fit to the data than a main effects-only model, $\chi^2(1) = 46.09, p < .001$. No means crossed below the mid-point into badness, and therefore, results represent “moral goodness.” In No Choice conditions, agents who helped strangers (versus kin) were judged as significantly more morally good, whereas in Choice conditions, agents helping kin instead of strangers (versus strangers instead of kin) were judged as significantly more morally good.

Trustworthiness. An identical pattern of judgments emerged for perceived trustworthiness. As predicted, there was a significant interaction between Relationship and Choice, $B = 0.772, SE_B = 0.095, t = 8.17, p < .001$; this interaction model provided a better fit to the data than a main effects-only model, $\chi^2(1) = 65.40, p < .001$.

Study 2

The purpose of Study 2 was to directly replicate the results of Study 1.

Method
Participants

Participants were 235 United States residents recruited and compensated via Amazon’s Mechanical Turk. This sample size was chosen to account for exclusions and reach an analyzable sample of \( N = 200 \). Participants who failed to correctly answer two scenario-relevant memory questions (\( n = 43 \)) were excluded from analyses, resulting in a final \( N = 193 \) (49.7% female, \( M_{\text{Age}} = 38.30 \)).

Materials and Procedure

The study design and procedure were identical to those of Study 1 and again pre-registered at http://aspredicted.org/blind.php?x=aa63pe. After reading each scenario, participants answered questions assessing the story’s agent. Because no means crossed below the mid-point in Study 1, moral goodness and trustworthiness were measured on unidirectional scales (1 = “not at all good/trustworthy” to 5 = “extremely good/trustworthy”).

Results

Descriptive statistics for moral goodness and trustworthiness are shown in Figure 2. Data were analyzed with identical mixed effects models as Study 1.
Moral Goodness. As predicted, there was a significant interaction between Relationship and Choice, $B = 0.532$, $SE_B = 0.073$, $t = 7.27$, $p < .001$; this interaction model provided a better fit to the data than a main effects-only model, $\chi^2(1) = 51.91$, $p < .001$. In No Choice conditions, agents who helped strangers (versus kin) were judged as significantly more morally good, whereas in Choice conditions, agents helping kin instead of strangers (versus strangers instead of kin) were judged as significantly more morally good.

Trustworthiness. An identical pattern of judgments emerged for perceived trustworthiness. As predicted, there was a significant interaction between Relationship and Choice, $B = 0.519$, $SE_B = 0.073$, $t = 7.06$, $p < .001$; this interaction model provided a better fit to the data than a main effects-only model, $\chi^2(1) = 49.02$, $p < .001$. 

Figure 2. Judgments of target agent in Study 2. Error bars represent +/- 1 SEM.
Study 3

Study 3 had three aims. First, it served as an additional replication of Studies 1-2. Second, we tested the prediction that simply failing to help kin would elicit harsher moral judgments than failing to help strangers. Third, our design allowed us to compare helping strangers instead of kin to simply failing to help kin.

Method

Participants

Participants were 330 United States residents recruited and compensated via Amazon’s Mechanical Turk. This sample size was chosen to account for exclusions and reach an analyzable sample of $N = 300$. Participants who failed to correctly answer at least two scenario-relevant memory questions ($n = 26$) were excluded from analyses, resulting in a final $N = 304$ (37.5% female, $M_{Age} = 36.88$).

Materials and Procedure

Participants read six of eight possible stories in which agents engaged or failed to engage in helping behavior. The study design and procedure were similar to Studies 1-2 with the addition of the Failed to Help scenarios. Thus, the design of the study was a fully crossed 2
(Relationship: Stranger vs. Kin) x 3 (Choice: No Choice vs. Choice vs. Failed to Help) within-subjects design in which participants saw six total stories, each corresponding to a different condition. Again, the study was pre-registered at http://aspredicted.org/blind.php?x=45sh9a. After reading each scenario, participants rated the agent’s moral goodness (1 = “not at all good” to 5 = “extremely good”).

**Results**

Descriptive statistics for moral goodness are shown in Figure 3. Data were analyzed with similar mixed effects models to Studies 1-2. Unlike Studies 1-2, though, which investigated one 2x2 interaction, Study 3 decomposed a 2x3 interaction into two 2x2 interactions to: (1) replicate Studies 1-2, and (2) investigate how Failed to Help scenarios relate to Choice scenarios. All pre-registered comparisons and corresponding statistics, across Studies 3-4, are reported in Tables 3-4 at the end of Study 4 Methods/Results.

![Figure 3](image-url) **Figure 3.** Judgments of target agent in Study 3. Error bars represent +/- 1 SEM.
Moral Goodness (No Choice vs. Choice). As predicted, there was a significant interaction between Relationship and Choice, $B = 0.518$, $SE_B = 0.090$, $t = 5.78$, $p < .001$; this interaction model provided a better fit to the data than a main effects-only model, $\chi^2(1) = 32.90$, $p < .001$. Unlike Studies 1-2, which demonstrated the predicted difference between strangers and kin, in No Choice conditions, here, agents who helped strangers (versus kin) were judged no more morally good. However, replicating Studies 1-2, in Choice conditions, agents helping kin instead of strangers (versus strangers instead of kin) were judged as significantly more morally good.

Moral Goodness (Choice vs. Failed). As predicted, there was a significant interaction between Relationship and Choice, $B = -0.603$, $SE_B = 0.108$, $t = -5.57$, $p < .001$; this interaction model provided a better fit to the data than a main effects-only model, $\chi^2(1) = 30.47$, $p < .001$. Agents who failed to help strangers (versus kin) were judged as significantly more morally good. Agents failing to help kin were also judged as significantly less morally good than agents helping strangers instead of kin, suggesting that neglecting kin but still helping someone is better than simply neglecting kin.

Study 4

Study 4 had three purposes: (1) replicate the results of Study 3, (2) examine explicit obligation judgments, and (3) test whether obligation judgments predict unique variance in moral judgments. Across No Choice and Choice conditions, helping kin should be judged as more of a fulfillment of obligation than helping strangers. Conversely, failing to help kin should be judged as more of a violation of obligation than failing to help strangers. We also expected that, between conditions of interest, obligation difference scores would predict moral judgment difference scores (though this was not pre-registered).

Method
Participants

Participants were 333 United States residents recruited and compensated via Amazon’s Mechanical Turk. This sample size was chosen to account for exclusions and reach an analyzable sample of \( N = 300 \). Participants who failed to correctly answer at least two scenario-relevant memory questions (\( n = 28 \)) were excluded from analyses, resulting in a final \( N = 305 \) (45.9% female, \( M_{\text{Age}} = 36.33 \)).

Materials and Procedure

The study design and procedure were similar to Study 3. Again, Study 4 was pre-registered at http://aspredicted.org/blind.php?x=vd2dk3. After reading each scenario, participants answered questions assessing the story’s agent. However, because the Failed to Help scenarios could be viewed as morally bad, all measures were evaluated on bidirectional scales to potentially capture negative judgments that pass below the mid-point. Specifically, participants judged the agent’s moral goodness and trustworthiness (1 = “extremely bad/untrustworthy” to 9 = “extremely good/trustworthy”), as well as if the agent violated or fulfilled an obligation (1 = “completely violated” to 9 = “completely fulfilled”). For secondary (non-pre-registered) analyses, participants also judged how likely it was that the typical person would do what the agent did (1 = “extremely unlikely” to 9 = “extremely likely”) and how predictable the agent’s future behavior was (1 = “extremely unpredictable” to 9 = “extremely predictable”).

Main Results

Descriptive statistics for moral goodness, trustworthiness, and obligation are shown in Figure 4 (for graphing purposes, we show “neither bad nor good; untrustworthy nor trustworthy; violated nor fulfilled an obligation,” originally 5s on the 1-9 scale, at the 0 midpoint). Data were analyzed with identical mixed effects models as Study 3.
Moral Goodness (No Choice vs. Choice). As predicted, there was a significant interaction between Relationship and Choice, $B = 0.652$, $SE_B = 0.151$, $t = 4.31$, $p < .001$; this interaction model provided a better fit to the data than a main effects-only model, $\chi^2(1) = 18.41$, $p < .001$. Like Study 3, in No Choice conditions, agents who helped strangers (versus kin) were judged as no more morally good. However, consistent with Studies 1-3, in Choice conditions, agents helping kin instead of strangers (versus strangers instead of kin) were judged as significantly more morally good.
Moral Goodness (Choice vs. Failed). As predicted, there was a significant interaction between Relationship and Choice, $B = -1.049, SE_B = 0.167, t = -6.27, p < .001$; this interaction model provided a better fit to the data than a main effects-only model, $\chi^2(1) = 38.59, p < .001$. Judgments passed below the mid-point for Failed to Help conditions. Specifically, agents who failed to help strangers (versus kin) were judged as significantly less morally bad (or more morally good). Replicating the results of Study 3, agents failing to help kin were judged as significantly less morally good than agents helping strangers instead of kin.

Trustworthiness (No Choice vs. Choice). Like moral goodness, an identical pattern of judgments emerged for perceived trustworthiness. As predicted, there was a significant interaction between Relationship and Choice, $B = 0.932, SE_B = 0.165, t = 5.66, p < .001$; this interaction model provided a better fit to the data than a main effects-only model, $\chi^2(1) = 31.49, p < .001$.

Trustworthiness (Choice vs. Failed). Like moral goodness, an identical pattern of judgments emerged for perceived trustworthiness. As predicted, there was a significant interaction between Relationship and Choice, $B = -1.598, SE_B = 0.179, t = -8.95, p < .001$; this interaction model provided a better fit to the data than a main effects-only model, $\chi^2(1) = 77.09, p < .001$.

Obligation (No Choice vs. Choice). As predicted, there was a significant interaction between Relationship and Choice, $B = 0.714, SE_B = 0.175, t = 4.08, p < .001$; this interaction model provided a better fit to the data than a main effects-only model, $\chi^2(1) = 16.51, p < .001$. In No Choice conditions, agents who helped strangers (versus kin) were judged significantly less as fulfilling an obligation, and, similarly, in Choice conditions, agents helping strangers instead of kin (versus kin instead of strangers) were judged significantly less as fulfilling an obligation.
Obligation (Choice vs. Failed). As predicted, there was a significant interaction between Relationship and Choice, $B = -1.649, SE_B = 0.185, t = -8.92, p < .001$; this interaction model provided a better fit to the data than a main effects-only model, $\chi^2(1) = 76.62, p < .001$.

Judgments for failing to help passed below the midpoint; specifically, agents who failed to help strangers (versus kin) were judged significantly less as violating an obligation. Agents failing to help kin were also judged significantly more as violating an obligation than agents helping strangers instead of kin.

### Table 3
Linear mixed effects means, 95% confidence intervals, t-statistics, and $p$-values for Studies 3-4 by Relationship and Choice.

<table>
<thead>
<tr>
<th></th>
<th>No Choice</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Kin</td>
<td>Stranger</td>
<td>$t$</td>
<td>$p$</td>
<td>Kin</td>
<td>Stranger</td>
<td>$t$</td>
<td>$p$</td>
<td>Kin</td>
<td>Stranger</td>
<td>$t$</td>
<td>$p$</td>
<td>Kin</td>
<td>Stranger</td>
<td>$t$</td>
</tr>
<tr>
<td>Study 3</td>
<td>Moral Goodness</td>
<td>4.14 (4.02, 4.26)</td>
<td>4.23 (4.11, 4.35)</td>
<td>1.46</td>
<td>.145</td>
<td>3.45 (3.33, 3.57)</td>
<td>3.02 (2.90, 3.14)</td>
<td>6.71</td>
<td>&lt; .001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study 4</td>
<td>Moral Goodness</td>
<td>7.62 (7.42, 7.81)</td>
<td>7.73 (7.53, 7.92)</td>
<td>1.03</td>
<td>.301</td>
<td>6.35 (6.16, 6.55)</td>
<td>5.81 (5.61, 6.01)</td>
<td>5.06</td>
<td>&lt; .001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Trustworthiness</td>
<td>7.29 (7.07, 7.50)</td>
<td>7.32 (7.11, 7.53)</td>
<td>0.27</td>
<td>.784</td>
<td>6.28 (6.07, 6.50)</td>
<td>5.38 (5.17, 5.60)</td>
<td>7.73</td>
<td>&lt; .001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Obligation</td>
<td>6.87 (6.63, 7.11)</td>
<td>6.52 (6.28, 6.76)</td>
<td>2.84</td>
<td>.005</td>
<td>6.27 (6.03, 6.51)</td>
<td>5.20 (4.96, 5.45)</td>
<td>8.61</td>
<td>&lt; .001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* In Study 3, moral goodness was rated on a 1-5 scale (1 = “not at all good” to “extremely good”). Study 3 total $N = 304$; total observations per variable = 1216. In Study 4, moral goodness and trustworthiness were rated on 1-9 scales (1 = “extremely bad/untrustworthy” to 5 = “neither bad nor good/untrustworthy nor trustworthy” to 9 = “extremely good/trustworthy”), and obligation was also rated on a 1-9 scale (1 = “completely violated” to 5 = “neither violated nor fulfilled” to 9 = “completely fulfilled”). Study 4 total $N = 305$; total observations per variable = 1220. $p$-values are unadjusted for multiple comparisons as all comparisons were pre-registered.

### Table 4
Linear mixed effects means, 95% confidence intervals, t-statistics, and $p$-values, for Studies 3-4 by Failed to Help comparisons.

<table>
<thead>
<tr>
<th></th>
<th>Failed to Help</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Kin</td>
<td>Stranger</td>
<td>$t$</td>
<td>$p$</td>
<td>Failed</td>
<td>Stranger</td>
<td>$t$</td>
<td>$p$</td>
<td>Neglected Kin</td>
<td>Failed</td>
<td>Stranger Instead</td>
<td>$t$</td>
<td>$p$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study 3</td>
<td>Moral Goodness</td>
<td>2.05 (1.86, 2.24)</td>
<td>2.21 (2.02, 2.40)</td>
<td>2.07</td>
<td>.039</td>
<td>2.05 (1.86, 2.24)</td>
<td>3.02 (2.83, 3.21)</td>
<td>12.59</td>
<td>&lt; .001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study 4</td>
<td>Moral Goodness</td>
<td>4.28 (3.98, 4.59)</td>
<td>4.79 (4.48, 5.10)</td>
<td>4.28</td>
<td>&lt; .001</td>
<td>4.28 (3.98, 4.59)</td>
<td>5.80 (5.49, 6.11)</td>
<td>12.81</td>
<td>&lt; .001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Trustworthiness</td>
<td>4.31 (4.07, 4.56)</td>
<td>5.01 (4.77, 5.25)</td>
<td>5.50</td>
<td>&lt; .001</td>
<td>4.31 (4.07, 4.56)</td>
<td>5.37 (5.13, 5.62)</td>
<td>8.41</td>
<td>&lt; .001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Obligation</td>
<td>4.30 (4.07, 4.53)</td>
<td>4.87 (4.65, 5.10)</td>
<td>4.41</td>
<td>&lt; .001</td>
<td>4.30 (4.07, 4.53)</td>
<td>5.20 (4.97, 5.42)</td>
<td>6.87</td>
<td>&lt; .001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* In Study 3, moral goodness was rated on a 1-5 scale (1 = “not at all good” to “extremely good”). Study 3 total $N = 304$; total observations per variable = 1216. In Study 4, moral goodness and trustworthiness were rated on 1-9 scales (1 = “extremely bad/untrustworthy” to 5 = “neither bad nor good/untrustworthy nor trustworthy” to 9 = “extremely good/trustworthy”), and obligation was also rated on a 1-9 scale (1 = “completely violated” to 5 = “neither violated nor fulfilled” to 9 = “completely fulfilled”). Study 4 total $N = 305$; total observations per variable = 1220. $p$-values are unadjusted for multiple comparisons as all comparisons were pre-registered.
Do Perceived Obligations Predict Moral Judgments?

Considering the perceived descriptive frequencies at which people tend to (1) help kin (versus strangers), (2) help kin instead of strangers (versus strangers instead of kin), and (3) fail to help kin (versus strangers), participants’ moral judgments could have been solely the result of the perceived typicality of the target agents’ behaviors or inferences about their future behavior (Mende-Siedlecki, Baron, & Todorov, 2013). To rule out the possibility that these variables were fully explaining our effects, we conducted hierarchical regressions using difference scores between conditions of interest. In the first model, participants’ typicality and future predictability judgment difference scores (e.g., Kin, No Choice typicality scores subtracted from Stranger, No Choice typicality scores) were entered as predictors, and, in the second model, obligation difference scores were entered as the predictor of interest. This allowed us to test whether obligation judgments accounted for unique variance in moral judgments, which would more directly support our hypotheses about the unique relationship between perceived obligations and moral judgments across studies.

Across all comparisons of interest, obligation difference scores were indeed significant (and always the strongest) predictors of moral judgment difference scores, suggesting that differences in judgments of obligations being violated or fulfilled drive differences in moral judgments (see Table 5). Additionally, using raw scores, we conducted the same hierarchical regressions within each individual condition. Again, within all conditions, obligation was a significant (and always the strongest) predictor of moral judgments (see SOM for full regression statistics). Within individual conditions, while controlling for the effects of typicality and future predictability on the relationship between obligation and moral goodness, all semi-partial correlations were significant, $r_s(301) \geq .41, ps < .001$. Similarly, while controlling for the effects
of the same variables on the relationship between obligation and trustworthiness, all semi-partial correlations were significant, \( rs(301) \geq .40, ps < .001 \).

### Table 5
Predictability, typicality, and obligation difference scores predicting moral judgment difference scores in Study 4.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>SNC – KNC</th>
<th>Moral Goodness</th>
<th>SNC – KNC</th>
<th>Moral Goodness</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model 1</td>
<td>Model 2</td>
<td></td>
<td>Model 2</td>
</tr>
<tr>
<td></td>
<td>b (SE)</td>
<td>( \beta )</td>
<td>sp-r</td>
<td>b (SE)</td>
</tr>
<tr>
<td>Predictability</td>
<td>.10 (.04)</td>
<td>.16</td>
<td>.15</td>
<td>.01</td>
</tr>
<tr>
<td>Typicality</td>
<td>.07 (.03)</td>
<td>.15</td>
<td>.13</td>
<td>.29</td>
</tr>
<tr>
<td>Obligation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F(2, 302)</td>
<td>4.15</td>
<td>.03</td>
<td>.017</td>
<td>3.01</td>
</tr>
<tr>
<td>R(^2)</td>
<td>.31</td>
<td></td>
<td>.015</td>
<td>.31</td>
</tr>
<tr>
<td>( \Delta R^2)</td>
<td>.01</td>
<td></td>
<td>.00</td>
<td></td>
</tr>
</tbody>
</table>

Note. SNC – KNC = Stranger, No Choice minus Kin, No Choice; KOS – SOK = Kin Over Stranger minus Stranger Over Kin; and SFH – KFH = Failed to Help Stranger minus Failed to Help Kin. Total \( N = 305 \).

### Discussion

Across four pre-registered studies, results suggest that special obligations influence moral judgments. In Studies 1-2, tentative evidence suggests that agents who help strangers are judged more positively than agents who help kin (though null effects were observed in Studies 3-4).

Conversely, in Studies 1-4, when given the choice between helping kin versus stranger, agents were judged more positively for helping kin instead of strangers (versus strangers instead of kin).

In Studies 3-4, when agents failed to help kin, they were judged more negatively than agents who...
failed to help strangers. Critically, in Study 4, when asked explicitly about obligations, perceptions aligned with the moral judgments made across studies. Participants perceived more fulfillment of an obligation for helping kin versus strangers (and for helping kin over strangers versus strangers over kin), more of a violation of obligation when agents failed to help kin versus failing to help strangers, and importantly, obligation judgment differences predicted moral judgment differences for all aforementioned comparisons.

Developmental research finds that young children expect others to behave prosocially toward ingroup members (Chalik & Dunham, 2018), and children even guide others to give more resources to family than to strangers (Olson & Spelke, 2008). That participants in our studies consistently judged agents helping kin over strangers (versus strangers over kin) as more morally good/trustworthy, and failing to help kin (versus strangers) as more morally bad/untrustworthy, suggests that adults believe others have obligations to help kin and make such contingency-based moral judgments. As people spend most of their time around those with whom relationships already exist, it is surprising that most prior work in moral psychology has investigated judgments about unrelated strangers (Bloom, 2011). However, a recent theory argues for a “morality-as-cooperation” account that invokes “helping kin” as an overlooked but important moral value, predicting (and demonstrating) that such help is judged as morally good (Curry, Chesters, & Van Lissa, 2019; Curry, Mullins, & Whitehouse, in press). Our results are consistent but add nuance to this view. Specifically, people may be judged even more positively when they help strangers (but less positively when helping strangers instead of kin), and when people fail to help kin, they are judged negatively. Somewhat inconsistent with our findings, Everett, Faber, Savulescu, and Crockett (2018) found that people who helped family over strangers were judged as no more moral than those who did the opposite. However, their scenarios contained a larger
number of strangers than family members (in which consequentialist or impartial reasoning would favor helping strangers), and helping behaviors were not matched across relationship. Our data suggest that when numbers and behaviors are matched, helping kin signals moral goodness. More generally, the Choice conditions’ results also suggest that moving from an agent-patient dyad (Schein & Gray, 2018) to a triad (or more) may meaningfully transform moral perception, an interesting avenue for future research.

While it has been argued and demonstrated that impartiality is moralized (e.g., DeScioli & Kurzban, 2009, 2013; Niemi, Wasserman, & Young, 2018), and valuing it can affect consequential ethical behavior (Waytz, Dungan, & Young, 2013), we have provided evidence that the opposite of impartiality, adherence to special obligations, is similarly moralized. Although favoring kin seemed appropriate across our scenarios, contexts may exist in which this behavior is inappropriate, such as when one is in a position to allocate valuable resources (Shaw & Olson, 2014). Exploratory analyses of one scenario in Studies 1-2 support this possibility. When a children’s soccer coach, after tryouts, provided private lessons to a cousin’s child instead of a stranger’s child, she was judged less positively than a coach who did the opposite, suggesting that special obligations and role-specific obligations (e.g., of a coach) can sometimes conflict.

One limitation of our work is that we tested exclusively U.S. participants (Henrich, Heine, & Norenzayan, 2010a). In No Choice conditions, more positive judgments when agents helped strangers may be due to WEIRD populations being more impersonally prosocial (Schulz, Bahrami-Rad, Beauchamp, & Henrich, 2018) and thus valuing this behavior in others (see Morgan, Mullen, & Skitka, 2010, for evidence of value-based judgment). In non-WEIRD cultures, however, helping strangers may be interpreted as a misuse of limited resources.
Moreover, in Choice and Failed to Help conditions, larger differences may exist in non-Western populations, who more strongly value loyalty (Graham et al., 2011). Another limitation is that none of our scenarios involved friends, who may (not) be perceived like kin; therefore, the generality or specificity of these effects is unknown (but see Marshall & Bloom, in prep, for evidence that, in similar situations involving children, older children and adults judge that family “has to” help more than friends). Future work would benefit from investigating how these effects vary developmentally and cross-culturally, and their boundary conditions.

**Conclusion**

In connecting our results to Everett and colleagues (2018), we suggest that people who are impartially prosocial are evaluated as less loyal and trustworthy precisely because they are perceived as not fulfilling, and perhaps not believing that they have, special obligations. While our work has focused on third-person perceptions of obligation, the findings may have practical implications for first-person prosociality and its promotion. For example, the more one donates to distant strangers, the less one has for family. Therefore, proponents of large-scale, impartial prosociality, like Singer, may need to do more than convince people that it is rational (and efficient) to be impersonally prosocial. The motivation people may need is to truly believe that they have obligations to strangers just as they do to family.
Footnotes

1. Participants also judged the wrongness of the act and whether the agent’s behavior was diagnostic of their future behavior. However, as our primary hypotheses were concerned with goodness/trustworthiness judgments (across studies), these results are not reported in the main text (see SOM).

2. Participants also judged the likelihood that others would behave similarly, how unpredictable the agent’s future behavior was, and how much harm, if any, occurred in the scenario (see SOM).

3. Adding the “Failed to Help” level to the Choice factor necessitated minor restructuring of some scenarios to increase the plausibility of failing to help. Entirely new scenarios were also created (see https://osf.io/bnwdv/).

4. Participants also judged the unpredictability of the agent’s future behavior (see SOM).

5. Participants also judged the rightness of the act, praise deserved, and fairness of the outcome, and four principles about obligations (see SOM).
Acknowledgments

We thank Arturo Balaguer-Townsend, Joshua Hirschfeld-Kroen, Minjae Kim, Gordon Kraft-Todd, Max Klieman-Weiner, Heather Maranges, Justin Martin, BoKyung Park, and Jordan Theriault for helpful feedback throughout this work, and the John Templeton Foundation for support.
References


Hughes, J.S. (2017). In a moral dilemma, choose the one you love: Impartial actors are seen as less moral than partial ones. *British Journal of Social Psychology, 56, 561-577.*


Supplemental Online Material

Study 1

Mixed Effects Models

Wrongness. There was a significant interaction between Relationship and Choice, $B = -0.844$, $SE_B = 0.105$, $t = -8.03$, $p < .001$; this interaction model provided a better fit to the data than a main effects-only model, $\chi^2(1) = 63.14$, $p < .001$. No means crossed above the mid-point into wrongness, and therefore, results represent “rightness”. In No Choice conditions, helping strangers (versus kin) was judged as significantly more right, whereas in Choice conditions, helping kin instead of strangers (versus strangers instead of kin) was judged as significantly more right (see SOM Table 1 for linear mixed effects means, 95% CIs, $t$-statistics, and $p$-values).

![Wrongness judgments in Study 1. Error bars represent +/- 1 SEM.](image)

SOM Figure 1a. Wrongness judgments in Study 1. Error bars represent +/- 1 SEM.

Diagnosticity. There was a significant interaction between Relationship and Choice, $B = 0.453$, $SE_B = 0.086$, $t = 5.28$, $p < .001$; this interaction model provided a better fit to the data than a main effects-only model, $\chi^2(1) = 27.63$, $p < .001$. No means crossed below the mid-point into
disagreement, and therefore, results represent agreement that the agent’s behavior was diagnostic of their future behavior. In No Choice conditions, helping strangers (versus kin) was judged as significantly more diagnostic of future behavior, whereas in Choice conditions, helping kin instead of strangers (versus strangers instead of kin) was judged as significantly more diagnostic of future behavior (see SOM Table 1 for linear mixed effects means, 95% CIs, t-statistics, and p-values).

![Diagram showing diagnosticity judgments](image)

**SOM Figure 1b.** Diagnosticity judgments in Study 1. Error bars represent +/- 1 SEM.

**Traditional Repeated-Measures ANOVAs**

*Moral Goodness.* A 2 (Relationship) x 2 (Choice) within-subjects ANOVA was conducted yielded a main effect of Relationship, $F(1, 208) = 5.08, p = .025, \eta^2 = .02$, such that subjects agents as more morally good for having helped kin than a stranger; a main effect of Choice, $F(1, 208) = 296.52, p < .001, \eta^2 = .59$, such that agents were judged as more morally good when no other potential beneficiary was available to help than when there were two potential beneficiaries to help; and a significant Relationship x Choice interaction, $F(1, 208) =$
34.72, $p < .001$, $\eta^2 = .14$. No conditions crossed below the midpoint of “neither bad nor good,” and so all results are presented in terms of goodness (higher scores meaning more morally good). Follow-up two-tailed paired-samples t-tests demonstrated that in the No Choice conditions, helping a stranger was judged as significantly more good ($M = 6.28$, $SD = 0.81$) than helping kin ($M = 6.07$, $SD = 0.91$), $t(208) = 3.55$, $p < .001$, Cohen’s $d = 0.25$ (all within-subjects Cohen’s $d$ s are adjusted for correlations between conditions). In the Choice conditions, however, moral goodness was judged significantly higher when agents helped kin instead of a stranger ($M = 5.34$, $SD = 1.01$) compared to helping a stranger instead of kin ($M = 4.91$, $SD = 1.18$), $t(208) = 4.88$, $p < .001$, Cohen’s $d = 0.34$.

Trustworthiness. A 2 (Relationship) x 2 (Choice) within-subjects ANOVA showed a main effect of Relationship, $F(1, 208) = 10.63$, $p = .001$, $\eta^2 = .05$, such that agents were judged as more trustworthy when they helped kin than a stranger; a main effect of Choice, $F(1, 208) = 243.63$, $p < .001$, $\eta^2 = .54$, such that agents were judged as more trustworthy when no other potential beneficiary was available to help than when there were two potential beneficiaries to help; and a significant Relationship x Choice interaction, $F(1, 208) = 47.81$, $p < .001$, $\eta^2 = .19$. No conditions crossed the below the midpoint of “neither untrustworthy nor trustworthy,” and so all results are presented in terms of trustworthiness (higher scores meaning more trustworthy). Follow-up two-tailed paired-samples t-tests demonstrated that in the No Choice conditions, helping a stranger communicated significantly higher trustworthiness ($M = 6.19$, $SD = 0.89$) than helping kin ($M = 5.97$, $SD = 1.19$), $t(208) = 4.22$, $p < .001$, Cohen’s $d = 0.29$. In Choice conditions, however, helping kin instead of a stranger communicated significantly higher trustworthiness ($M = 5.36$, $SD = 1.06$) than helping a stranger instead of kin ($M = 4.81$, $SD = 1.19$), $t(208) = 5.88$, $p < .001$, Cohen’s $d = 0.41$. 

Wrongness. A 2 (Relationship) x 2 (Choice) within-subjects ANOVA yielded a main effect of Relationship, $F(1, 208) = 10.84, p = .001, \eta^2 = .05$, such that subjects judged it more right when agents helped kin than a stranger; a main effect of Choice, $F(1, 208) = 275.66, p < .001, \eta^2 = .57$, such that it was more right when no other potential beneficiary was available to help than when there were two potential beneficiaries to help; and a significant Relationship x Choice interaction, $F(1, 208) = 52.29, p < .001, \eta^2 = .20$. No conditions crossed below the midpoint of “neither right nor wrong,” and so all means are presented in terms of rightness (lower scores meaning more right). Follow-up two-tailed paired-samples t-tests demonstrated that in the No Choice conditions, helping a stranger was judged as significantly more right ($M = 1.82, SD = 1.01$) than helping kin ($M = 2.04, SD = 1.10$), $t(208) = 3.53, p = .001$, Cohen’s $d = 0.24$. In the Choice conditions, however, rightness was judged significantly higher when agents helped kin instead of a stranger ($M = 2.83, SD = 1.87$) compared to helping a stranger instead of kin ($M = 3.44, SD = 1.36$), $t(208) = 6.16, p = < .001$, Cohen’s $d = 0.43$.

Diagnosticity. A 2 (Relationship) x 2 (Choice) within-subjects ANOVA revealed no main effect of Relationship, $F(1, 208) = 0.27, p = .604, \eta^2 = .00$. There was, however, a main effect of Choice, $F(1, 208) = 158.95, p < .001, \eta^2 = .43$, such that agents’ actions were judged to be more diagnostic of future behavior when no alternate beneficiary was available to help than when there were two potential beneficiaries to help; and a significant Relationship x Choice interaction, $F(1, 208) = 25.65, p < .001, \eta^2 = .11$. Follow-up two-tailed paired-samples t-tests demonstrated that in the No Choice conditions, helping a stranger was judged as saying significantly more about an agent’s future behavior ($M = 6.16, SD = 0.90$) than helping kin ($M = 5.91, SD = 1.04$), $t(208) = 4.55, p < .001$, Cohen’s $d = 0.32$. In the Choice conditions, however, helping kin instead of a stranger was judged as significantly more diagnostic of an agent’s future behavior ($M = 5.40, SD$
= 1.12) than helping a stranger instead of kin (M = 5.19, SD = 1.14), t(208) = 2.93, p = .004, Cohen’s d = 0.20.

Dependent Variable Wording and Scales

*Moral Goodness.* “Please rate how morally bad or morally good [agent’s name] is as a person.”

(1 = “extremely bad”; 4 = “neither bad nor good”; 7 = “extremely good”)

*Trustworthiness.* “[Agent’s name] seems…”

(1 = “extremely untrustworthy”; 4 = “neither untrustworthy nor trustworthy”; 7 = “extremely trustworthy”)

*Wrongness.* “Please rate the moral rightness or wrongness of [agent’s name’s] behavior.”

(1 = “extremely right”; 4 = “neither right nor wrong”; 7 = “extremely wrong”)

*Diagnosticity.* “[Agent’s name’s] actions say a lot about how s/he will act in the future.”

(1 = “strongly disagree”; 4 = “neither disagree nor agree”; 7 = “strongly agree”)
Study 2

Mixed Effects Models

Typicality. There was a significant interaction between Relationship and Choice, $B = 0.388, SE_B = 0.084, t = 4.60, p < .001$; this interaction model provided a better fit to the data than a main effects-only model, $\chi^2(1) = 21.03, p < .001$. In No Choice conditions, helping strangers (versus kin) was judged as significantly less likely, and similarly, in Choice conditions, helping strangers instead of kin (versus helping kin instead of strangers) was also judged as significantly less typical (see SOM Table 1 for linear mixed effects means, 95% CIs, $t$-statistics, and $p$-values).

![Typicality Graph](image)

**SOM Figure 2a.** Frequency judgments in Study 2. Error bars represent +/- 1 SEM.

Unpredictability. There was a significant interaction between Relationship and Choice, $B = -0.706, SE_B = 0.089, t = -7.97, p < .001$; this interaction model provided a better fit to the data than a main effects-only model, $\chi^2(1) = 62.08, p < .001$. In No Choice conditions, helping strangers (versus kin) signaled significantly higher future unpredictability, and similarly, in
Choice conditions, helping strangers instead of kin (versus helping kin instead of strangers) also signaled significantly higher future unpredictability (see SOM Table 1 for linear mixed effects means, 95% CIs, \(t\)-statistics, and \(p\)-values).

SOM Figure 2b. Unpredictability judgments in Study 2. Error bars represent +/- 1 SEM.

_Harm._ There was a significant interaction between Relationship and Choice, \(B = -0.175, SE_B = 0.059, t = -2.94, p = .003\); this interaction model provided a better fit to the data than a main effects-only model, \(\chi^2(1) = 8.61, p = .003\). In No Choice conditions, helping strangers (versus kin) was judged as marginally less harmful, whereas in Choice conditions, helping strangers instead of kin (versus helping kin instead of strangers) was judged as significantly more harmful (see SOM Table 1 for linear mixed effects means, 95% CIs, \(t\)-statistics, and \(p\)-values).
SOM Figure 2c. Harm judgments in Study 2. Error bars represent +/- 1 SEM.

SOM Table 1
Linear mixed effects means, 95% confidence intervals, t-statistics, and p-values for Studies 1-2 by Relationship and Choice.

<table>
<thead>
<tr>
<th></th>
<th>Study 1</th>
<th>Study 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No Choice</td>
<td>Choice</td>
</tr>
<tr>
<td></td>
<td>Stranger</td>
<td>Kin</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Wrongness</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.04 (1.77, 2.30)</td>
<td>1.82 (1.55, 2.08)</td>
</tr>
<tr>
<td></td>
<td>5.91 (5.74, 6.08)</td>
<td>6.16 (5.99, 6.33)</td>
</tr>
<tr>
<td><strong>Diagnosticity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.74 (2.45, 3.04)</td>
<td>2.10 (1.81, 2.40)</td>
</tr>
<tr>
<td><strong>Typicality</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.18 (1.97, 2.39)</td>
<td>2.46 (2.25, 2.67)</td>
</tr>
<tr>
<td><strong>Unpredictability</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.27 (1.13, 1.40)</td>
<td>1.19 (1.06, 1.32)</td>
</tr>
<tr>
<td><strong>Harm</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.08 (1.77, 2.30)</td>
<td>1.82 (1.55, 2.08)</td>
</tr>
<tr>
<td></td>
<td>5.91 (5.74, 6.08)</td>
<td>6.16 (5.99, 6.33)</td>
</tr>
</tbody>
</table>

Note. In Study 1, wrongness and diagnosticity were rated on 1-7 scales: Wrongness: 1 = “extremely right” to 4 = “neither right nor wrong” to 7 = “extremely wrong”; and Diagnosticity: 1 = “strongly disagree” to 4 = “neither disagree nor agree” to 7 = “strongly agree.” Study 1 total N = 209; total observations per variable = 1672. In Study 2, typicality, unpredictability, and harm were rated on 1-5 scales. Typicality: 1 = “not at all likely” to 5 = “extremely likely”; Unpredictability: 1 = “not at all unpredictable” to 5 = “extremely unpredictable”; and Harm: 1 = “not at all” to 5 = “extremely.” Study 2 total N = 193; total observations per variable = 1544. p-values are unadjusted for multiple comparisons.

Traditional Repeated-Measures ANOVAs

Moral Goodness. A 2 (Relationship) x 2 (Choice) within-subjects ANOVA revealed no main effect of Relationship, $F(1, 192) = 1.03, p = .311, \eta^2 = .01$; a main effect of Choice, $F(1, 192) = 341.04, p < .001, \eta^2 = .64$, such that agents were judged as more morally good when no other potential beneficiary was available to help than when there were two potential beneficiaries; and a significant Relationship x Choice interaction, $F(1, 192) = 42.63, p < .001, \eta^2$
Follow-up two-tailed paired-samples t-tests demonstrated that in the No Choice conditions, agents helping a stranger were judged as significantly more morally good (\(M = 4.23, SD = 0.70\)) than when helping kin (\(M = 4.02, SD = 0.74\)), \(t(192) = 4.97, p < .001, \) Cohen’s \(d = 0.36\) (all within-subjects Cohen’s \(d\)s are adjusted for correlations between conditions). In the Choice conditions, however, moral goodness was judged significantly higher when agents helped kin instead of a stranger (\(M = 3.38, SD = 0.83\)) compared to helping a stranger instead of kin (\(M = 3.08, SD = 0.83\)), \(t(192) = 4.42, p = .001, \) Cohen’s \(d = 0.32\).

Trustworthiness. A 2 (Relationship) x 2 (Choice) within-subjects ANOVA showed a main effect of Relationship, \(F(1, 192) = 9.64, p = .002, \eta^2 = .05\), such that agents were judged as more trustworthy when they helped kin than a stranger; a main effect of Choice, \(F(1, 192) = 273.49, p < .001, \eta^2 = .59\), such that agents were judged as more trustworthy when no other potential beneficiary was available to help than when there were two potential beneficiaries; and a significant Relationship x Choice interaction, \(F(1, 192) = 37.26, p < .001, \eta^2 = .16\). Follow-up two-tailed paired-samples t-tests demonstrated that in the No Choice conditions, helping a stranger signaled significantly higher trustworthiness (\(M = 4.04, SD = 0.73\)) than helping kin (\(M = 3.92, SD = 0.73\)), \(t(192) = 2.94, p = .004, \) Cohen’s \(d = 0.21\). In the Choice conditions, however, helping kin instead of a stranger signaled significantly higher trustworthiness (\(M = 3.39, SD = 0.81\)) than helping a stranger instead of kin (\(M = 3.01, SD = 0.89\)), \(t(192) = 5.30, p < .001, \) Cohen’s \(d = 0.38\).

Typicality. A 2 (Relationship) x 2 (Choice) within-subjects ANOVA yielded a main effect of Relationship, \(F(1, 192) = 245.85, p < .001, \eta^2 = .56\), such that helping kin compared to strangers was judged as more likely; no main effect of Choice, \(F(1, 192) = 0.18, p = .671, \eta^2 = .00\); and a significant Relationship x Choice interaction, \(F(1, 192) = 15.29, p < .001, \eta^2 = .07\).
Follow-up two-tailed paired-samples t-tests demonstrated that in the No Choice conditions, helping a stranger was judged to be significantly less likely ($M = 2.07, SD = 0.80$) than helping kin ($M = 2.77, SD = 0.86$), $t(192) = 10.92, p < .001$, Cohen’s $d = 0.79$. Similarly, in Choice conditions, helping a stranger instead of kin was judged as significantly less likely ($M = 1.92, SD = 0.81$) than agents who helped kin instead of a stranger ($M = 2.96, SD = 0.85$), $t(192) = 13.60, p < .001$, Cohen’s $d = 0.98$.

**Unpredictability.** A 2 (Relationship) x 2 (Choice) within-subjects ANOVA resulted in a main effect of Relationship, $F(1, 192) = 142.29, p < .001$, $\eta^2 = .43$, such that agents who helped strangers (versus kin) signaled more unpredictable future behavior; a main effect of Choice $F(1, 192) = 18.44, p < .001$, $\eta^2 = .09$, such that when no alternative beneficiary was available to help, an agent’s future behavior was judged as less unpredictable than when there were two potential beneficiaries; and a significant Relationship x Choice interaction, $F(1, 192) = 56.90, p < .001$, $\eta^2 = .23$. Follow-up two-tailed paired-samples t-tests demonstrated that in the No Choice conditions, when agents helped a stranger, their future behavior was judged as significantly more unpredictable ($M = 2.49, SD = 0.92$) than when having helped kin ($M = 2.16, SD = 0.84$), $t(192) = 5.43, p < .001$, Cohen’s $d = 0.39$. In Choice conditions, again, agents who helped a stranger instead of kin were judged as significantly more unpredictable in the future ($M = 3.04, SD = 0.91$) than agents who helped kin instead of a stranger ($M = 2.05, SD = 0.79$), $t(192) = 12.38, p < .001$, Cohen’s $d = 0.89$.

**Harm.** A 2 (Relationship) x 2 (Choice) within-subjects ANOVA revealed no main effect of Relationship, $F(1, 192) = 0.42, p = .520$, $\eta^2 = .00$; a main effect of Choice, $F(1, 192) = 84.89, p < .001$, $\eta^2 = .31$, such that subjects perceived less harm when no other potential beneficiary was available to help than when there were two potential beneficiaries; and a significant Relationship
x Choice interaction, $F(1, 192) = 6.15, p = .014, \eta^2 = .03$. Follow-up two-tailed paired-samples t-tests demonstrated that in the No Choice conditions, participants perceived marginally less harm when agents helped a stranger ($M = 1.20, SD = 0.52$) than when they helped kin ($M = 1.26, SD = 0.55$), $t(192) = 1.89, p = .061$, Cohen’s $d = 0.14$. In the Choice conditions, however, participants perceived marginally more harm when agents helped a stranger instead of kin ($M = 1.62, SD = 0.73$) than when agents helped kin instead of a stranger ($M = 1.52, SD = 0.68$), $t(192) = 1.81, p = .072$, Cohen’s $d = 0.13$.

**Dependent Variable Wording and Scales**

*Moral Goodness.* “Please rate how morally good [agent’s name] is as a person.”

(1 = “not at all good”; 5 = “extremely good”)

*Trustworthiness.* “In terms of trustworthiness, [agent’s name] seems…”

(1 = “not at all trustworthy”; 5 = “extremely trustworthy”)

Typicality. “How likely is it that the average person would do what [agent’s name’s] did?”

(1 = “not at all likely”; 5 = “extremely likely”)

Unpredictability. “In terms of unpredictability, [agent’s name’s] future behavior seems…”

(1 = “not at all unpredictable”; 5 = “extremely unpredictable”)

**Harm.** “Please rate how much someone in this scenario is being harmed.”

(1 = “not at all”; 5 = “extremely”)

Study 3

Mixed Effects Models

Unpredictability (No Choice vs. Choice). There was a significant interaction between Relationship and Choice, $B = -0.749$, $SE_B = 0.106$, $t = -7.23$, $p < .001$; this interaction model provided a better fit to the data than a main effects-only model, $\chi^2(1) = 50.82$, $p < .001$. In No Choice conditions, helping strangers (versus kin) signaled significantly higher future unpredictability, and similarly, in Choice conditions, helping strangers instead of kin (versus helping kin instead of strangers) also signaled significantly higher future unpredictability (see SOM Table 2 for linear mixed effects means, 95% CIs, $t$-statistics, and $p$-values).

Unpredictability (Choice vs. Failed). There was a significant interaction between Relationship and Choice, $B = 1.299$, $SE_B = 0.110$, $t = 11.78$, $p < .001$; this interaction model provided a better fit to the data than a main effects-only model, $\chi^2(1) = 128.95$, $p < .001$. For agents who failed to help strangers (versus kin), their future behavior was judged to be significantly less unpredictable. For agents who failed to help kin, their future behavior was judged to be significantly less unpredictable than agents who helped strangers instead of kin, suggesting that neglecting kin to help someone else signals more unpredictable future behavior. (see SOM Table 3 for linear mixed effects means, 95% CIs, $t$-statistics, and $p$-values).
Traditional Repeated-Measures ANOVAs

*Moral Goodness (No Choice vs. Choice).* A 2 (Relationship) x 2 (Choice) within-subjects ANOVA was conducted and revealed a main effect of Relationship, $F(1, 303) = 14.58, p < .001$, $\eta^2 = .05$, such that agents were judged as more morally good when they helped kin versus a stranger; a main effect of Choice, $F(1, 303) = 357.70, p < .001$, $\eta^2 = .54$, such that agents were judged as more morally good when no other potential beneficiary was available to help than when there were two potential beneficiaries; and a significant Relationship x Choice interaction, $F(1, 303) = 20.27, p < .001$, $\eta^2 = .12$. Follow-up two-tailed paired-samples t-tests demonstrated that in the No Choice conditions, agents who helped a stranger were judged as significantly more morally good ($M = 4.24, SD = 0.81$) than agents who helped kin ($M = 4.14, SD = 0.79$), $t(303) = 2.20, p = .029$, Cohen’s $d = 0.13$ (all within-subjects Cohen’s $d$s are adjusted for correlations between conditions). In the Choice conditions, however, moral goodness was judged significantly higher when agents helped kin instead of a stranger ($M = 3.45, SD = 0.93$)
compared to helping a stranger instead of kin ($M = 3.02$, $SD = 1.04$), $t(303) = 5.85$, $p = < .001$, Cohen’s $d = 0.34$.

*Moral Goodness (Choice vs. Failed).* A 2 (Relationship) x 2 (Choice) within-subjects ANOVA was conducted while adding the Failed to Help conditions and leaving out the No Choice conditions. This revealed a main effect of Relationship, $F(1, 303) = 8.28$, $p = .004$, $\eta^2 = .03$, such that agents were judged as more morally good when interacted with kin versus a stranger; a main effect of Choice, $F(1, 303) = 297.95$, $p < .001$, $\eta^2 = .50$, such that agents were judged as more morally good when they helped rather than failed to help; and a significant Relationship x Choice interaction, $F(1, 303) = 31.90$, $p < .001$, $\eta^2 = .10$. Follow-up two-tailed paired-samples t-tests demonstrated that in the Failed to Help conditions, agents who failed to help a stranger were judged as significantly more morally good ($M = 2.19$, $SD = 1.05$) than those who failed to help kin ($M = 2.06$, $SD = 1.10$), $t(303) = 1.97$, $p = .050$, Cohen’s $d = 0.11$. Agents who failed to help kin were also seen as less morally good ($M = 2.06$, $SD = 1.10$) than agents who chose to help a stranger instead of kin ($M = 3.02$, $SD = 1.04$), $t(303) = 11.26$, $p = < .001$, Cohen’s $d = 0.65$.

*Unpredictability (No Choice vs. Choice).* A 2 (Relationship) x 2 (Choice) within-subjects ANOVA yielded a main effect of Relationship, $F(1, 303) = 196.62$, $p < .001$, $\eta^2 = .39$, such that agents who helped strangers (versus kin) signaled more unpredictable future behavior; a main effect of Choice, $F(1, 303) = 78.95$, $p < .001$, $\eta^2 = .21$, such that when no alternative beneficiary was available to help, an agent’s future behavior was judged as less unpredictable than when there were two potential beneficiaries; and a significant Relationship x Choice interaction, $F(1, 303) = 52.73$, $p < .001$, $\eta^2 = .15$. Follow-up two-tailed paired-samples t-tests demonstrated that in the No Choice conditions, when agents helped a stranger, their future behavior was judged as
significantly more unpredictable \( (M = 2.40, SD = 1.16) \) than when having helped kin \( (M = 2.01, SD = 1.00) \), \( t(303) = 5.81, p < .001 \), Cohen’s \( d = 0.24 \). In the Choice conditions, again, for agents who helped a stranger instead of kin, their future behavior was judged as significantly more unpredictable \( (M = 3.26, SD = 1.09) \) than the future behavior of agents who helped kin instead of a stranger \( (M = 2.15, SD = 1.02) \), \( t(303) = 14.42, p < .001 \), Cohen’s \( d = 0.83 \).

Unpredictability (Choice vs. Failed). A 2 (Relationship) x 2 (Choice) within-subjects ANOVA was conducted while adding the Failed to Help conditions and leaving out the No Choice conditions. This analysis demonstrated a main effect of Relationship, \( F(1, 303) = 100.27, p < .001 \), \( \eta^2 = .25 \), such that agents interacting with strangers (versus kin) signaled more unpredictable future behavior; a main effect of Choice, \( F(1, 303) = 6.28, p = .013 \), \( \eta^2 = .02 \), such that an agent’s future behavior was judged as less unpredictable when they helped than when they failed to help; and a significant Relationship x Choice interaction, \( F(1, 303) = 127.70, p < .001 \), \( \eta^2 = .30 \). Follow-up two-tailed paired-samples t-tests demonstrated that in the Failed to Help conditions, when agents failed to help a stranger, their future behavior was judged as significantly less unpredictable \( (M = 2.47, SD = 1.17) \) than when failing to help kin \( (M = 2.62, SD = 1.15) \), \( t(303) = 2.16, p = .032 \), Cohen’s \( d = 0.12 \). For agents who failed to help kin, their future behavior was also judged as significantly less unpredictable \( (M = 2.62, SD = 1.15) \) than agents who helped a stranger instead of kin \( (M = 3.26, SD = 1.08) \), \( t(303) = 7.74, p < .001 \), Cohen’s \( d = 0.44 \).

Dependent Variable Wording and Scales

Moral Goodness. “How morally good is [agent’s name] as a person?”

\( 1 = \) “not at all good”; \( 5 = \) “extremely good”

Unpredictability. “How unpredictable is [agent’s name’s] future behavior?”
(1 = “not at all unpredictable”; 5 = “extremely unpredictable”)
Study 4

Mixed Effects Models

Rightness (No Choice vs. Choice). There was a significant interaction between Relationship and Choice, $B = 0.967$, $SE_B = 0.176$, $t = 5.50$, $p < .001$; this interaction model provided a better fit to the data than a main effects-only model, $\chi^2(1) = 29.78$, $p < .001$. No means crossed below the mid-point into wrongness, and therefore, results represent “rightness”. In No Choice conditions, helping strangers (versus kin) was judged as no more right, whereas in Choice conditions, helping kin instead of strangers (versus strangers instead of kin) was judged as significantly more right (see SOM Table 2 for linear mixed effects means, 95% CIs, $t$-statistics, and $p$-values).

Rightness (Choice vs. Failed). There was a significant interaction between Relationship and Choice, $B = -1.681$, $SE_B = 0.189$, $t = -8.90$, $p < .001$; this interaction model provided a better fit to the data than a main effects-only model, $\chi^2(1) = 76.20$, $p < .001$. Failed to help conditions crossed below the midpoint into wrongness. In Failed to Help conditions, failing to help strangers (versus kin) was judged as significantly less wrong (or more right). Failing to help kin was also judged as significantly more wrong than helping a stranger instead of kin. (see SOM Table 3 for linear mixed effects means, 95% CIs, $t$-statistics, and $p$-values).
Praise (No Choice vs. Choice). There was a significant interaction between Relationship and Choice, $B = 0.690, SE_B = 0.170$, $t = 4.06, p < .001$; this interaction model provided a better fit to the data than a main effects-only model, $\chi^2(1) = 16.29, p < .001$. No means crossed below the midpoint into blame, and therefore, results represent “praise”. In No Choice conditions, helping strangers (versus kin) was judged as no more praiseworthy, whereas in Choice conditions, helping kin instead of strangers (versus strangers instead of kin) was judged as significantly more praiseworthy (see SOM Table 2 for linear mixed effects means, 95% CIs, $t$-statistics, and $p$-values).

Praise (Choice vs. Failed). There was a significant interaction between Relationship and Choice, $B = -1.056, SE_B = 0.177$, $t = -5.97, p < .001$; this interaction model provided a better fit to the data than a main effects-only model, $\chi^2(1) = 35.14, p < .001$. Failed to help conditions crossed below the midpoint into blame. In Failed to Help conditions, failing to help strangers (versus kin) was judged as significantly less blameworthy (or more praiseworthy). Failing to help
kin was also judged as significantly more blameworthy than helping a stranger instead of kin. (see SOM Table 3 for linear mixed effects means, 95% CIs, t-statistics, and p-values).

**SOM Figure 4b.** Praise judgments in Study 4. Error bars represent +/- 1 SEM.

*Fairness (No Choice vs. Choice).* There was a significant interaction between Relationship and Choice, $B = 1.055$, $SE_B = 0.181$, $t = 5.84$, $p < .001$; this interaction model provided a better fit to the data than a main effects-only model, $\chi^2(1) = 33.50$, $p < .001$. No means crossed below the midpoint into unfair, and therefore, results represent “fairness”. In No Choice conditions, strangers (versus kin) being helped was judged as no more fair, whereas in Choice conditions, kin being helped instead of strangers (versus strangers being helped instead of kin) was judged as significantly more fair (see SOM Table 2 for linear mixed effects means, 95% CIs, t-statistics, and p-values).

*Fairness (Choice vs. Failed).* There was a significant interaction between Relationship and Choice, $B = -1.828$, $SE_B = 0.207$, $t = -8.81$, $p < .001$; this interaction model provided a better fit to the data than a main effects-only model, $\chi^2(1) = 74.83$, $p < .001$. Some Failed to Help...
judgments crossed below the midpoint into unfair. Failing to help strangers (versus kin) was judged as significantly more fair (or less unfair). Failing to help kin was also judged as significantly less fair than a stranger being helped instead of kin. (see SOM Table 3 for linear mixed effects means, 95% CIs, t-statistics, and p-values).

**SOM Figure 4c.** Fairness judgments in Study 4. Error bars represent +/- 1 SEM.

*Predictability (No Choice vs. Choice).* There was a significant interaction between Relationship and Choice, $B = 1.262$, $SE_B = 0.186$, $t = 6.77$, $p < .001$; this interaction model provided a better fit to the data than a main effects-only model, $\chi^2(1) = 44.80$, $p < .001$. In No Choice conditions, helping strangers (versus kin) signaled significantly lower future predictability, and similarly, in Choice conditions, helping strangers instead of kin (versus helping kin instead of strangers) also signaled significantly lower future predictability (and in fact, this measure crossed below the midpoint such that helping strangers instead of kin signaled unpredictability; see SOM Table 2 for linear mixed effects means, 95% CIs, t-statistics, and p-values).
Predictability (Choice vs. Failed). There was a significant interaction between Relationship and Choice, \( B = -2.138, SE_B = 0.191, t = -11.22, p < .001; \) this interaction model provided a better fit to the data than a main effects-only model, \( \chi^2(1) = 118.39, p < .001.\) For agents who failed to help strangers (versus kin), their future behavior was judged to be significantly more predictable. For agents who failed to help kin, judgments fell below the midpoint into unpredictable. Their future behavior was judged to be significantly more unpredictable (or less predictable) than agents who helped strangers instead of kin, suggesting that neglecting kin to help someone else signals more unpredictable future behavior (see SOM Table 3 for linear mixed effects means, 95% CIs, \( t \)-statistics, and \( p \)-values).

\[
\begin{array}{ccc}
\text{Predictability} \\
(1 = \text{extremely unpredictable}; 5 = \text{neither}; 9 = \text{extremely predictable})
\end{array}
\]

\[
\begin{array}{ccc}
\text{No Choice} & \text{Choice} & \text{Failed} \\
\hline
\text{Stranger} & \text{Kin} \\
\end{array}
\]

SOM Figure 4d. Predictability judgments in Study 4. Error bars represent +/- 1 SEM.

Typicality (No Choice vs. Choice). There was a significant interaction between Relationship and Choice, \( B = 1.198, SE_B = 0.206, t = 5.82, p < .001; \) this interaction model provided a better fit to the data than a main effects-only model, \( \chi^2(1) = 33.40, p < .001.\) In No Choice conditions, helping strangers (versus kin) was judged as significantly less typical (and in
fact, it was judged as somewhat unlikely, crossing below the midpoint), and similarly, in Choice conditions, helping strangers instead of kin (versus helping kin instead of strangers) was also judged as significantly less likely (and similar to No Choice conditions, it was judged as somewhat unlikely, crossing the midpoint; see SOM Table 1 for linear mixed effects means, 95% CIs, $t$-statistics, and $p$-values).

Typicality (Choice vs. Failed). There was a significant interaction between Relationship and Choice, $B = -3.761$, $SE_B = 0.222$, $t = -16.94$, $p < .001$; this interaction model provided a better fit to the data than a main effects-only model, $\chi^2(1) = 250.54$, $p < .001$. For agents who failed to help strangers (versus kin), this was judged as significantly more likely (and in fact, failing to help kin was judged as somewhat unlikely, crossing below the midpoint). Failing to help kin was also judged as more likely (or less unlikely) than helping a stranger instead of helping kin (both judgments crossed below the midpoint; see SOM Table 3 for linear mixed effects means, 95% CIs, $t$-statistics, and $p$-values).

SOM Figure 4e. Predictability judgments in Study 4. Error bars represent +/- 1 SEM.
Running head: WHAT WE OWE TO FAMILY

SOM Table 2
Linear mixed effects means, 95% confidence intervals, t-statistics, and p-values, for Studies 3-4 by Relationship and Choice.

<table>
<thead>
<tr>
<th></th>
<th>No Choice</th>
<th></th>
<th></th>
<th>Choice</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Kin</td>
<td>Stranger</td>
<td>t</td>
<td>p</td>
<td>Kin</td>
<td>Stranger</td>
</tr>
<tr>
<td>Study 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unpredictability</td>
<td>2.03 (1.82, 2.23)</td>
<td>2.38 (2.18, 2.59)</td>
<td>4.88</td>
<td>&lt; .001</td>
<td>2.14 (1.93, 2.34)</td>
<td>3.24 (3.04, 3.45)</td>
</tr>
<tr>
<td>Study 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rightness</td>
<td>7.66 (7.41, 7.91)</td>
<td>7.72 (7.47, 7.97)</td>
<td>0.55</td>
<td>.582</td>
<td>6.34 (6.09, 6.59)</td>
<td>5.44 (5.19, 5.69)</td>
</tr>
<tr>
<td>Praise</td>
<td>7.70 (7.50, 7.89)</td>
<td>7.84 (7.65, 8.03)</td>
<td>1.19</td>
<td>.234</td>
<td>6.14 (5.95, 6.34)</td>
<td>5.60 (5.40, 5.79)</td>
</tr>
<tr>
<td>Fairness</td>
<td>7.51 (7.26, 7.75)</td>
<td>7.53 (7.29, 7.77)</td>
<td>0.14</td>
<td>.893</td>
<td>6.41 (6.17, 6.65)</td>
<td>5.37 (5.13, 5.61)</td>
</tr>
<tr>
<td>Predictability</td>
<td>6.13 (5.73, 6.54)</td>
<td>5.80 (5.40, 6.20)</td>
<td>2.54</td>
<td>.011</td>
<td>6.05 (5.64, 6.45)</td>
<td>4.45 (4.05, 4.85)</td>
</tr>
<tr>
<td>Typicality</td>
<td>5.77 (5.18, 6.37)</td>
<td>4.68 (4.08, 5.27)</td>
<td>7.35</td>
<td>&lt; .001</td>
<td>6.19 (5.59, 6.78)</td>
<td>3.89 (3.30, 4.49)</td>
</tr>
</tbody>
</table>

Note. In Study 3, unpredictability was rated on a 1-5 scale. Unpredictability: 1 = “not at all unpredictable” to 5 = “extremely unpredictable.” Study 3 total N = 304; total observations per variable = 1216. In Study 4, rightness, praise, fairness, predictability, and typicality were rated on 1-9 scales. Rightness: 1 = “extremely wrong”; 5 = “neither”; 9 = “extremely right”; Praise: 1 = “extremely blameworthy”; 5 = “neither”; 9 = “extremely praiseworthy”; Fairness: 1 = “completely unfair”; 5 = “neither”; 9 = “completely fair”; Predictability: 1 = “extremely unpredictable”; 5 = “neither”; 9 = “extremely predictable”; and Typicality: 1 = “extremely unlikely”; 5 = “neither”; 9 = “extremely likely.” Study 4 total N = 305; total observations per variable = 1220. p-values are unadjusted for multiple comparisons.

SOM Table 3
Linear mixed effects means, 95% confidence intervals, t-statistics, and p-values, for Studies 3-4 by Failed to Help comparisons.

<table>
<thead>
<tr>
<th></th>
<th>Failed to Help</th>
<th></th>
<th></th>
<th>Neglected Kin</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Kin</td>
<td>Stranger</td>
<td>t</td>
<td>p</td>
<td>Failed</td>
<td>Stranger Instead</td>
</tr>
<tr>
<td>Study 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unpredictability</td>
<td>2.63 (2.45, 2.82)</td>
<td>2.46 (2.28, 2.64)</td>
<td>2.23</td>
<td>.026</td>
<td>2.63 (2.45, 2.82)</td>
<td>3.26 (3.08, 3.45)</td>
</tr>
<tr>
<td>Study 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rightness</td>
<td>4.13 (3.80, 4.45)</td>
<td>4.91 (4.58, 5.23)</td>
<td>5.85</td>
<td>&lt; .001</td>
<td>4.13 (3.80, 4.45)</td>
<td>5.43 (5.10, 5.76)</td>
</tr>
<tr>
<td>Praise</td>
<td>4.04 (3.76, 4.32)</td>
<td>4.55 (4.27, 4.83)</td>
<td>4.03</td>
<td>&lt; .001</td>
<td>4.04 (3.76, 4.32)</td>
<td>5.59 (5.31, 5.87)</td>
</tr>
<tr>
<td>Fairness</td>
<td>4.47 (4.08, 4.86)</td>
<td>5.27 (4.87, 5.66)</td>
<td>5.44</td>
<td>&lt; .001</td>
<td>4.47 (4.08, 4.86)</td>
<td>5.36 (4.97, 5.76)</td>
</tr>
<tr>
<td>Predictability</td>
<td>5.24 (4.94, 5.54)</td>
<td>5.78 (5.48, 6.07)</td>
<td>3.99</td>
<td>&lt; .001</td>
<td>5.24 (4.94, 5.54)</td>
<td>4.45 (4.15, 4.75)</td>
</tr>
<tr>
<td>Typicality</td>
<td>4.80 (4.43, 5.16)</td>
<td>6.26 (5.89, 6.63)</td>
<td>9.32</td>
<td>&lt; .001</td>
<td>4.80 (4.43, 5.16)</td>
<td>3.89 (3.52, 4.26)</td>
</tr>
</tbody>
</table>

Note. In Study 3, unpredictability was rated on a 1-5 scale. Unpredictability: 1 = “not at all unpredictable” to 5 = “extremely unpredictable.” Study 3 total N = 304; total observations per variable = 1216. In Study 4, rightness, praise, fairness, predictability, and typicality were rated on 1-9 scales. Rightness: 1 = “extremely wrong”; 5 = “neither”; 9 = “extremely right”; Praise: 1 = “extremely blameworthy”; 5 = “neither”; 9 = “extremely praiseworthy”; Fairness: 1 = “completely unfair”; 5 = “neither”; 9 = “completely fair”; Predictability: 1 = “extremely unpredictable”; 5 = “neither”; 9 = “extremely predictable”; and Typicality: 1 = “extremely unlikely”; 5 = “neither”; 9 = “extremely likely.” Study 4 total N = 305; total observations per variable = 1220. p-values are unadjusted for multiple comparisons.

Traditional Repeated-Measures ANOVAs

Moral Goodness (No Choice vs. Choice). A 2 (Relationship) x 2 (Choice) within-subjects ANOVA revealed a main effect of Relationship, F(1, 304) = 8.56, p = .027, η² = .03, such that agents were judged as more morally good when they helped kin than a stranger; a main effect of
Choice, $F(1, 304) = 370.23, p < .001, \eta^2 = .55$, such that agents were judged as more morally good when no other potential beneficiary was available to help than when there were two potential beneficiaries; and a significant Relationship x Choice interaction, $F(1, 304) = 21.41, p < .001, \eta^2 = .07$. No conditions in this 2x2 crossed below the midpoint, and so all analyses are presented in terms of “goodness” with higher scores meaning more good. Follow-up two-tailed paired-samples t-tests demonstrated that in the No Choice conditions, agents who helped a stranger were judged as no more morally good ($M = 7.72, SD = 1.29$) than agents who helped kin ($M = 7.62, SD = 1.27$), $t(304) = 1.57, p = .118$, Cohen’s $d = 0.09$ (all within-subjects Cohen’s $d$s are adjusted for correlations between conditions). In the Choice conditions, however, moral goodness was judged significantly higher when agents helped kin instead of a stranger ($M = 6.35, SD = 1.60$) compared to helping a stranger instead of kin ($M = 5.81, SD = 1.69$), $t(304) = 4.18, p < .001$, Cohen’s $d = 0.24$.

Moral Goodness (Choice vs. Failed). A 2 (Relationship) x 2 (Choice) within-subjects ANOVA revealed no main effect of Relationship, $F(1, 304) = 0.01, p = .938, \eta^2 = .00$; a main effect of Choice, $F(1, 304) = 287.29, p < .001, \eta^2 = .49$, such that agents were judged as more morally good when they helped rather than failed to help; and a significant Relationship x Choice interaction, $F(1, 304) = 42.79, p < .001, \eta^2 = .12$. Failure to help judgments were below the midpoint indicating that failing to help reflected moral badness. Follow-up two-tailed paired-samples t-tests demonstrated that in the Failed to Help conditions, agents who failed to help a stranger were judged as significantly less morally bad ($M = 4.81, SD = 1.53$) than those who failed to help kin ($M = 4.28, SD = 1.75$), $t(304) = 5.12, p < .001$, Cohen’s $d = 0.29$. Agents who failed to help kin were also seen as more morally bad (or less morally good) ($M = 4.28, SD = $
1.75) than agents who helped a stranger instead of kin \((M = 5.81, SD = 1.69), t(304) = 11.68, p < .001, \text{Cohen’s } d = 0.67.\)

**Trustworthiness (No Choice vs. Choice).** A 2 (Relationship) x 2 (Choice) within-subjects ANOVA showed a main effect of Relationship, \(F(1, 304) = 29.66, p < .001, \eta^2 = .09,\) such that agents were judged as more trustworthy when they helped kin than a stranger; a main effect of Choice, \(F(1, 304) = 294.95, p < .001, \eta^2 = .49,\) such that agents were judged as more trustworthy when no other potential beneficiary was available to help than when there were two potential beneficiaries; and a significant Relationship x Choice interaction, \(F(1, 304) = 32.44, p < .001, \eta^2 = .10.\) No conditions in this 2x2 crossed below the midpoint into untrustworthiness, and so all results are presented in terms of “trustworthiness” with higher scores meaning more trustworthy. Follow-up two-tailed paired-samples t-tests demonstrated that in the No Choice conditions, helping a stranger signaled no more trustworthiness \((M = 7.31, SD = 1.48)\) than helping kin \((M = 7.30, SD = 1.51), t(304) = 0.25, p = .800, \text{Cohen’s } d = 0.02.\) In the Choice conditions, however, helping kin instead of a stranger signaled significantly higher trustworthiness \((M = 6.28, SD = 1.65)\) than helping a stranger instead of kin \((M = 5.38, SD = 1.76), t(304) = 6.35, p < .001, \text{Cohen’s } d = 0.36.\)

**Trustworthiness (Choice vs. Failed).** A 2 (Relationship) x 2 (Choice) within-subjects ANOVA yielded no main effect of Relationship, \(F(1, 304) = 1.34, p = .249, \eta^2 = .00;\) a main effect of Choice, \(F(1, 304) = 413.20, p < .001, \eta^2 = .38,\) such that agents were judged as more trustworthy when they helped rather than failed to help; and a significant Relationship x Choice interaction, \(F(1, 304) = 72.15, p < .001, \eta^2 = .19.\) Follow-up two-tailed paired-samples t-tests demonstrated that in the Failed to Help conditions, agents who failed to help a stranger were judged as significantly more trustworthy \((M = 5.02, SD = 1.51)\) than those who failed to help kin
(who were seen as somewhat untrustworthy, crossing below the midpoint) ($M = 4.31, SD = 1.78$), $t(304) = 6.38, p < .001$, Cohen’s $d = 0.37$. Agents who failed to help kin were also seen as less trustworthy (or more untrustworthy) ($M = 4.31, SD = 1.78$) than agents who chose to help a stranger instead of kin ($M = 5.38, SD = 1.76$), $t(304) = 8.25, p = < .001$, Cohen’s $d = 0.47$.

**Obligation (No Choice vs. Choice).** A 2 (Relationship) x 2 (Choice) within-subjects ANOVA revealed a main effect of Relationship, $F(1, 304) = 66.21, p < .001$, $\eta^2 = .18$, such that agents were judged as fulfilling an obligation more for helping kin than for helping a stranger; a main effect of Choice, $F(1, 304) = 106.34, p < .001$, $\eta^2 = .26$, such that agents were judged as fulfilling an obligation more when no other potential beneficiary was available to help than when there were two potential beneficiaries; and a significant Relationship x Choice interaction, $F(1, 304) = 18.65, p < .001$, $\eta^2 = .06$. No conditions in this 2x2 crossed below the midpoint into violating an obligation, and so all results are presented in terms of “fulfilling obligations” with higher scores meaning higher belief in agents fulfilling obligations. Follow-up two-tailed paired-samples t-tests demonstrated that in the No Choice conditions, helping strangers was judged as fulfilling less of an obligation ($M = 6.51, SD = 1.67$) than helping kin ($M = 6.89, SD = 1.73$), $t(304) = 3.96, p < .001$, Cohen’s $d = 0.23$. In the Choice conditions, again, helping strangers instead of helping kin was judged as fulfilling less of an obligation ($M = 5.20, SD = 1.82$) than helping kin instead of helping strangers ($M = 6.27, SD = 1.71$), $t(304) = 7.57, p < .001$, Cohen’s $d = 0.43$.

**Obligation (Choice vs. Failed).** A 2 (Relationship) x 2 (Choice) within-subjects ANOVA demonstrated a main effect of Relationship, $F(1, 304) = 7.26, p = .007$, $\eta^2 = .02$, such that judgments of fulfilling obligation were higher for kin compared to a stranger; a main effect of Choice, $F(1, 304) = 139.66, p < .001$, $\eta^2 = .32$, such that helping was judged as fulfilling more of
an obligation than failing to help; and a significant Relationship x Choice interaction, \( F(1, 304) = 80.67, p < .001, \eta^2 = .21 \). Failure to help judgments were below the midpoint indicating that failing to help is judged as violating an obligation. Follow-up two-tailed paired-samples t-tests demonstrated that failing to help a stranger was judged as violating an obligation significantly less (\( M = 4.89, SD = 1.54 \)) than failing to help kin (\( M = 4.30, SD = 1.82 \)), \( t(304) = 5.13, p < .001, \) Cohen’s \( d = 0.29 \). Failing to help kin was also seen as violating an obligation significantly more (or fulfilling an obligation less) (\( M = 4.30, SD = 1.82 \)) than helping a stranger instead of kin (\( M = 5.20, SD = 1.82 \)), \( t(304) = 6.78, p = < .001, \) Cohen’s \( d = 0.39 \).

**Rightness (No Choice vs. Choice).** A 2 (Relationship) x 2 (Choice) within-subjects ANOVA yielded a main effect of Relationship, \( F(1, 304) = 23.89, p < .001, \eta^2 = .07 \), such that subjects judged it as more right when agents helped kin than a stranger; a main effect of Choice, \( F(1, 304) = 342.21, p < .001, \eta^2 = .53 \), such that it was more right when no other potential beneficiary was available to help than when there were two potential beneficiaries; and a significant Relationship x Choice interaction, \( F(1, 304) = 35.67, p < .001, \eta^2 = .11 \). No conditions in this 2x2 crossed below the midpoint into wrongness, and so all results are presented in terms of “rightness” with higher scores meaning more right. Follow-up two-tailed paired-samples t-tests demonstrated that in the No Choice conditions, helping a stranger was judged as no more right (\( M = 7.72, SD = 1.46 \)) than helping kin (\( M = 7.67, SD = 1.55 \)), \( t(304) = 0.61, p = .542, \) Cohen’s \( d = 0.04 \). In the Choice conditions, however, rightness was judged significantly higher when agents helped kin instead of a stranger (\( M = 6.34, SD = 1.79 \)) compared to when agents helped a stranger instead of kin (\( M = 5.44, SD = 1.90 \)), \( t(304) = 6.29, p = < .001, \) Cohen’s \( d = 0.36 \).
**Rightness (Choice vs. Failed).** A 2 (Relationship) x 2 (Choice) within-subjects ANOVA demonstrated no main effect of Relationship, \(F(1, 304) = 0.31, p = .578, \eta^2 = .00\); a main effect of Choice, \(F(1, 304) = 177.60, p < .001, \eta^2 = .37\), such that an agent’s behavior was judged as more right when they helped rather than failed to help; and a significant Relationship x Choice interaction, \(F(1, 304) = 78.89, p < .001, \eta^2 = .21\). Failure to help judgments were below the midpoint indicating that failing to help was perceived as wrong. Follow-up two-tailed paired-samples t-tests demonstrated that failing to help a stranger was judged as significantly less wrong \((M = 4.92, SD = 1.84)\) than failing to help kin \((M = 4.12, SD = 1.94)\), \(t(304) = 6.95, p < .001, \) Cohen’s \(d = 0.40\). Failing to help kin was also seen as more wrong \((M = 4.12, SD = 1.94)\) than helping a stranger instead of kin \((M = 5.44, SD = 1.90)\), \(t(304) = 9.50, p = < .001, \) Cohen’s \(d = 0.54\).

**Praise (No Choice vs. Choice).** A 2 (Relationship) x 2 (Choice) within-subjects ANOVA yielded a main effect of Relationship, \(F(1, 304) = 6.31, p = .013, \eta^2 = .02\), such that helping kin was judged more praiseworthy than helping a stranger; a main effect of Choice, \(F(1, 304) = 419.90, p < .001, \eta^2 = .58\), such that praise was higher when no other potential beneficiary was available to help than when there were two potential beneficiaries; and a significant Relationship x Choice interaction, \(F(1, 304) = 18.26, p < .001, \eta^2 = .06\). No conditions in this 2x2 crossed the below the midpoint into blame, and so all results are presented in terms of “praise” with higher scores meaning more praise. Follow-up two-tailed paired-samples t-tests demonstrated that in the No Alternative conditions, agents were praised marginally more when helping strangers \((M = 7.84, SD = 1.38)\) compared to kin \((M = 7.70, SD = 1.52)\), \(t(304) = 1.67, p = .095, \) Cohen’s \(d = 0.10\). In the Choice conditions, however, agent were praised significantly more for helping kin
instead of a stranger \((M = 6.14, SD = 1.74)\) compared to helping a stranger instead of kin \((M = 5.59, SD = 1.85)\), \(t(304) = 3.93, p < .001\), Cohen’s \(d = 0.23\).

Praise (Choice vs. Failed). A 2 (Relationship) x 2 (Choice) within-subjects ANOVA revealed no main effect of Relationship, \(F(1, 304) = 0.02, p = .878, \eta^2 = .00\); a main effect of Choice, \(F(1, 304) = 278.43, p < .001, \eta^2 = .48\), such that agents were praised more when they helped rather than failed to help; and a significant Relationship x Choice interaction, \(F(1, 304) = 36.41, p < .001, \eta^2 = .11\). Failure to help judgments were below the midpoint indicating that failing to help was perceived as blameworthy. Follow-up two-tailed paired-samples t-tests demonstrated that failing to help a stranger was judged as significantly less blameworthy \((M = 4.56, SD = 1.47)\) than failing to help kin \((M = 4.04, SD = 1.94)\), \(t(304) = 4.99, p < .001\), Cohen’s \(d = 0.29\). Failure to help kin was also judged as more blameworthy \((M = 4.04, SD = 1.94)\) than helping a stranger instead of kin \((M = 5.59, SD = 1.85)\), \(t(304) = 11.45, p = < .001\), Cohen’s \(d = 0.66\).

Fairness (No Choice vs. Choice). A 2 (Relationship) x 2 (Choice) within-subjects ANOVA yielded a main effect of Relationship, \(F(1, 304) = 38.03, p < .001, \eta^2 = .11\), such that it was judged as fairer when agents helped kin than a stranger; a main effect of Choice, \(F(1, 304) = 249.19, p < .001, \eta^2 = .45\), such that it was judged as fairer when no other potential beneficiary was available to help than when there were two potential beneficiaries; and a significant Relationship x Choice interaction, \(F(1, 304) = 39.48, p < .001, \eta^2 = .12\). No conditions in this 2x2 crossed below the midpoint into unfair, and so all results are presented in terms of “fairness” with higher scores meaning fairer. Follow-up two-tailed paired-samples t-tests demonstrated that in the No Alternative conditions, the outcome was judged as no fairer when agents helped a stranger \((M = 7.52, SD = 1.53)\) than when agents helped kin \((M = 7.52, SD = 1.53)\), \(t(304) = \)
0.04, \( p = .967 \), Cohen’s \( d = 0.00 \). In Choice conditions, however, helping kin instead of a stranger was judged as significantly fairer (\( M = 6.41, SD = 1.92 \)) than helping a stranger instead of kin (\( M = 5.37, SD = 1.98 \)), \( t(304) = 7.05, p < .001 \), Cohen’s \( d = 0.40 \).

**Fairness (Choice vs. Failed).** A 2 (Relationship) x 2 (Choice) within-subjects ANOVA resulted in no main effect of Relationship, \( F(1, 304) = 1.12, p = .292, \eta^2 = .00 \); a main effect of Choice, \( F(1, 304) = 77.91, p < .001, \eta^2 = .20 \), such that it was judged as fairer when agents helped rather than failed to help; and a significant Relationship x Choice interaction, \( F(1, 304) = 83.85, p < .001, \eta^2 = .22 \). Some failure to help judgments were below the midpoint indicating that, depending on relationship, failing to help was judged as somewhat unfair. Follow-up two-tailed paired-samples t-tests demonstrated that failing to help a stranger was judged as significantly fairer (\( M = 5.29, SD = 2.13 \)) than failing to help kin (\( M = 4.46, SD = 2.17 \)), \( t(304) = 6.08, p < .001, \) Cohen’s \( d = 0.35 \). Failing to help kin was also judged as more unfair (or less fair) (\( M = 4.46, SD = 2.17 \)) than when agents helped a stranger instead of kin (\( M = 5.37, SD = 1.98 \)), \( t(304) = 5.85, p = < .001, \) Cohen’s \( d = 0.34 \).

**Predictability (No Choice vs. Choice).** A 2 (Relationship) x 2 (Choice) within-subjects ANOVA demonstrated a main effect of Relationship, \( F(1, 304) = 96.06, p < .001, \eta^2 = .24 \), such that the future behavior of agents who helped strangers was judged as less predictable than the future behavior of agents who helped kin; a main effect of Choice, \( F(1, 304) = 54.19, p < .001, \eta^2 = .15 \), such that when no alternative beneficiary was available to help, an agent’s future behavior was judged as more predictable than when there were two potential beneficiaries; and a significant Relationship x Choice interaction, \( F(1, 304) = 51.80, p < .001, \eta^2 = .15 \). Follow-up two-tailed paired-samples t-tests demonstrated that in the No Alternative conditions, when agents helped a stranger, their future behavior was judged as significantly less predictable (\( M = 5.81, \)
\(SD = 2.00\) than when agents helped kin \((M = 6.17, SD = 1.85), t(304) = 3.26, p < .001, \text{Cohen’s } d = 0.19\). In the Choice conditions, again, the future behavior of agents who helped a stranger instead of kin were judged as significantly less predictable (or more unpredictable) \((M = 4.45, SD = 2.09)\) than the future behavior of agents who helped kin instead of a stranger \((M = 6.06, SD = 1.77), t(304) = 10.64, p < .001, \text{Cohen’s } d = 0.61\).

**Predictability (Choice vs. Failed).** A 2 (Relationship) x 2 (Choice) within-subjects ANOVA yielded a main effect of Relationship, \(F(1, 304) = 31.02, p < .001, \eta^2 = .09\), such that judgments of future predictability were higher when agents interacted with kin compared to a stranger; a main effect of Choice, \(F(1, 304) = 6.32, p = .012, \eta^2 = .02\), such that helping signaled higher future predictability than failing to help; and a significant Relationship x Choice interaction, \(F(1, 304) = 127.50, p < .001, \eta^2 = .30\). Follow-up two-tailed paired-samples t-tests demonstrated that the future behavior of agents failing to help a stranger was judged as more predictable \((M = 5.78, SD = 1.75)\) than the future behavior of agents failing to help kin \((M = 5.22, SD = 1.88), t(304) = 4.92, p < .001, \text{Cohen’s } d = 0.28\). Failure to help kin was also seen as signaling more predictable behavior in the future \((M = 5.22, SD = 1.88)\) than helping a stranger instead of kin \((M = 4.45, SD = 2.09), t(304) = 5.47, p = < .001, \text{Cohen’s } d = 0.31\).

**Typicality (No Choice vs. Choice).** A 2 (Relationship) x 2 (Choice) within-subjects ANOVA resulted in a main effect of Relationship, \(F(1, 304) = 246.70, p < .001, \eta^2 = .49\), such that helping kin compared to strangers was judged as more likely; a marginal main effect of Choice, \(F(1, 304) = 3.68, p = .056, \eta^2 = .01\), such that helping someone when no other potential beneficiary was available to help was judged as more likely than helping someone over another; and a significant Relationship x Choice interaction, \(F(1, 304) = 32.15, p < .001, \eta^2 = .10\). Follow-up two-tailed paired-samples t-tests demonstrated that in the No Alternative conditions, helping a
stranger was judged to be significantly less likely (or more unlikely) \((M = 4.69, SD = 2.15)\) than helping kin \((M = 5.84, SD = 2.14)\), \(t(304) = 8.30, p < .001\), Cohen’s \(d = 0.48\). In the Choice conditions, again, helping a stranger instead of kin was judged as significantly less likely (or more unlikely) \((M = 3.88, SD = 2.14)\) than helping kin instead of a stranger \((M = 6.20, SD = 1.95)\), \(t(304) = 14.23, p < .001\), Cohen’s \(d = 0.82\).

**Typicality (Choice vs. Failed).** A 2 (Relationship) x 2 (Choice) within-subjects ANOVA showed a main effect of Relationship, \(F(1, 304) = 14.12, p < .001\), \(\eta^2 = .04\), such that judgments of likelihood were higher when agents interacted with kin compared to a stranger; a main effect of Choice, \(F(1, 304) = 16.77, p < .001\), \(\eta^2 = .05\), such that helping was judged as less likely than failing to help; and a significant Relationship x Choice interaction, \(F(1, 304) = 295.33, p < .001\), \(\eta^2 = .49\). Some failure to help judgments were below the midpoint indicating that, depending on relationship, failing to help was judged as somewhat unlikely. Follow-up two-tailed paired-samples t-tests demonstrated that failing to help a stranger was judged as more likely \((M = 6.28, SD = 1.97)\) than failing to help kin \((M = 4.77, SD = 2.12)\), \(t(304) = 10.24, p < .001\), Cohen’s \(d = 0.59\). Failing to help kin was also judged as more likely (or less unlikely) \((M = 4.77, SD = 2.12)\) than helping a stranger instead of kin \((M = 3.88, SD = 2.14)\), \(t(304) = 5.72, p = < .001\), Cohen’s \(d = 0.33\).

**Do Obligation Judgments Predict Unique Variance in Moral Judgments within Conditions?**

To rule out the possibility that moral judgments were solely the result of behavior frequency and agent future predictability (Mende-Siedlecki, Baron, & Todorov, 2013), we conducted hierarchical regressions. In the first models, participants’ typicality and future predictability judgments were entered as predictors, and in the second model, obligation
judgments were entered as the predictor of interest. This allowed us to test if obligation judgments accounted for unique variance in moral judgments. Within all conditions, obligation was a significant (and always the strongest) predictor of moral judgments (see SOM Tables 4-5 for full regression statistics).

**SOM Table 4**
Hierarchical regressions predicting moral goodness within individual conditions for Study 4.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Model 1</th>
<th></th>
<th></th>
<th>Model 2</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b (SE)</td>
<td>sp-r</td>
<td>p</td>
<td>b (SE)</td>
<td>sp-r</td>
<td>p</td>
</tr>
<tr>
<td>SNC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predictability</td>
<td>.10 (.04)</td>
<td>.15 .13</td>
<td>2.35 .019</td>
<td>.02 (.04)</td>
<td>.03 .03</td>
<td>0.56 .574</td>
</tr>
<tr>
<td>Typicality</td>
<td>-.06 (.04)</td>
<td>-.11 -.10</td>
<td>-1.67 .096</td>
<td>-.09 (.04)</td>
<td>-.15 -.13</td>
<td>-2.52 .012</td>
</tr>
<tr>
<td>Obligation</td>
<td>- - - - -</td>
<td>- - - -</td>
<td>- - - -</td>
<td>- - - -</td>
<td>- - - -</td>
<td>- - - -</td>
</tr>
<tr>
<td>F(2, 302) =</td>
<td>2.99</td>
<td>R² = .02</td>
<td>.052</td>
<td>F(3, 301) =</td>
<td>22.71</td>
<td>R² = .19</td>
</tr>
<tr>
<td>KNC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predictability</td>
<td>.18 (.04)</td>
<td>.26 .24</td>
<td>4.38 &lt; .001</td>
<td>.11 (.04)</td>
<td>.15 .14</td>
<td>2.78 .006</td>
</tr>
<tr>
<td>Typicality</td>
<td>.00 (.04)</td>
<td>.01 .01</td>
<td>0.10 .917</td>
<td>-.04 (.03)</td>
<td>-.06 -.06</td>
<td>-1.12 .262</td>
</tr>
<tr>
<td>Obligation</td>
<td>- - - - -</td>
<td>- - - -</td>
<td>- - - -</td>
<td>- - .25</td>
<td>.47 .45</td>
<td>9.02 &lt; .001</td>
</tr>
<tr>
<td>F(2, 302) =</td>
<td>11.53</td>
<td>R² = .07</td>
<td>&lt; .001</td>
<td>F(3, 301) =</td>
<td>36.84</td>
<td>R² = .27</td>
</tr>
<tr>
<td>SOK</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predictability</td>
<td>.11 (.05)</td>
<td>.13 .11</td>
<td>2.11 .035</td>
<td>.01 (.04)</td>
<td>.01 .01</td>
<td>0.11 .912</td>
</tr>
<tr>
<td>Typicality</td>
<td>.28 (.05)</td>
<td>.35 .29</td>
<td>5.70 &lt; .001</td>
<td>.08 (.05)</td>
<td>.10 .08</td>
<td>1.76 .080</td>
</tr>
<tr>
<td>Obligation</td>
<td>- - - - -</td>
<td>- - - -</td>
<td>- - - -</td>
<td>- - - -</td>
<td>- - - -</td>
<td>- - - -</td>
</tr>
<tr>
<td>F(2, 302) =</td>
<td>36.30</td>
<td>R² = .19</td>
<td>&lt; .001</td>
<td>F(3, 301) =</td>
<td>70.55</td>
<td>R² = .41</td>
</tr>
<tr>
<td>KOS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predictability</td>
<td>.39 (.06)</td>
<td>.43 .32</td>
<td>6.60 &lt; .001</td>
<td>.16 (.06)</td>
<td>.17 .12</td>
<td>2.82 .005</td>
</tr>
<tr>
<td>Typicality</td>
<td>.11 (.05)</td>
<td>.13 .10</td>
<td>1.99 .048</td>
<td>.05 (.05)</td>
<td>.06 .04</td>
<td>1.05 .293</td>
</tr>
<tr>
<td>Obligation</td>
<td>- - - - -</td>
<td>- - - -</td>
<td>- - - -</td>
<td>- - - -</td>
<td>- - - -</td>
<td>- - - -</td>
</tr>
<tr>
<td>F(2, 302) =</td>
<td>57.79</td>
<td>R² = .28</td>
<td>&lt; .001</td>
<td>F(3, 301) =</td>
<td>88.72</td>
<td>R² = .47</td>
</tr>
<tr>
<td>SFH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predictability</td>
<td>.09 (.05)</td>
<td>.10 .09</td>
<td>1.76 .079</td>
<td>.04 (.04)</td>
<td>.05 .04</td>
<td>1.01 .312</td>
</tr>
<tr>
<td>Typicality</td>
<td>.28 (.05)</td>
<td>.37 .33</td>
<td>6.28 &lt; .001</td>
<td>.13 (.04)</td>
<td>.17 .15</td>
<td>3.58 &lt; .001</td>
</tr>
<tr>
<td>Obligation</td>
<td>- - - - -</td>
<td>- - - -</td>
<td>- - - -</td>
<td>- - - -</td>
<td>- - - -</td>
<td>- - - -</td>
</tr>
<tr>
<td>F(2, 302) =</td>
<td>32.88</td>
<td>R² = .18</td>
<td>&lt; .001</td>
<td>F(3, 301) =</td>
<td>103.38</td>
<td>R² = .51</td>
</tr>
<tr>
<td>KFH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predictability</td>
<td>.11 (.05)</td>
<td>.12 .11</td>
<td>2.24 .026</td>
<td>.05 (.04)</td>
<td>.05 .05</td>
<td>1.37 .171</td>
</tr>
<tr>
<td>Typicality</td>
<td>.43 (.04)</td>
<td>.52 .45</td>
<td>9.66 &lt; .001</td>
<td>.16 (.04)</td>
<td>.20 .16</td>
<td>4.65 &lt; .001</td>
</tr>
<tr>
<td>Obligation</td>
<td>- - - - -</td>
<td>- - - -</td>
<td>- - - -</td>
<td>- - - -</td>
<td>- - - -</td>
<td>- - - -</td>
</tr>
<tr>
<td>F(2, 302) =</td>
<td>77.06</td>
<td>R² = .34</td>
<td>&lt; .001</td>
<td>F(3, 301) =</td>
<td>194.07</td>
<td>R² = .66</td>
</tr>
</tbody>
</table>

Note. SNC = Stranger, No Choice; KNC = Kin, No Choice; SOK = Stranger Over Kin; KOS = Kin Over Stranger; SFH = Failed to Help Stranger; and KFH = Failed to Help Kin. Total N = 305.
SOM Table 5
Hierarchical regressions predicting trustworthiness within individual conditions for Study 4.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b (SE)</td>
<td>β</td>
</tr>
<tr>
<td>SNC</td>
<td>.17 (.05)</td>
<td>.23</td>
</tr>
<tr>
<td>Typicality</td>
<td>-.03 (.04)</td>
<td>-.04</td>
</tr>
<tr>
<td>Obligation</td>
<td>-.41 (.05)</td>
<td>-.41</td>
</tr>
<tr>
<td>F(2, 302)</td>
<td>7.51</td>
<td>.05</td>
</tr>
<tr>
<td>KNC</td>
<td>.28 (.05)</td>
<td>.35</td>
</tr>
<tr>
<td>Typicality</td>
<td>.05 (.04)</td>
<td>.07</td>
</tr>
<tr>
<td>Obligation</td>
<td>-.40 (.04)</td>
<td>-.46</td>
</tr>
<tr>
<td>F(2, 302)</td>
<td>25.31</td>
<td>-.14</td>
</tr>
<tr>
<td>SOK</td>
<td>.24 (.05)</td>
<td>.29</td>
</tr>
<tr>
<td>Typicality</td>
<td>.31 (.05)</td>
<td>.38</td>
</tr>
<tr>
<td>Obligation</td>
<td>-.53 (.05)</td>
<td>-.55</td>
</tr>
<tr>
<td>F(2, 302)</td>
<td>79.52</td>
<td>-.35</td>
</tr>
<tr>
<td>KOS</td>
<td>.41 (.06)</td>
<td>.44</td>
</tr>
<tr>
<td>Typicality</td>
<td>.09 (.06)</td>
<td>.10</td>
</tr>
<tr>
<td>Obligation</td>
<td>-.47 (.05)</td>
<td>-.48</td>
</tr>
<tr>
<td>F(2, 302)</td>
<td>54.40</td>
<td>-.27</td>
</tr>
<tr>
<td>SFH</td>
<td>.12 (.05)</td>
<td>.13</td>
</tr>
<tr>
<td>Typicality</td>
<td>.28 (.04)</td>
<td>.36</td>
</tr>
<tr>
<td>Obligation</td>
<td>-.45 (.05)</td>
<td>-.46</td>
</tr>
<tr>
<td>F(2, 302)</td>
<td>36.11</td>
<td>-.19</td>
</tr>
<tr>
<td>KFH</td>
<td>.07 (.05)</td>
<td>.07</td>
</tr>
<tr>
<td>Typicality</td>
<td>.42 (.05)</td>
<td>.50</td>
</tr>
<tr>
<td>Obligation</td>
<td>-.66 (.04)</td>
<td>-.68</td>
</tr>
<tr>
<td>F(2, 302)</td>
<td>61.05</td>
<td>-.29</td>
</tr>
</tbody>
</table>

Note. SNC = Stranger, No Choice; KNC = Kin, No Choice; SOK = Stranger Over Kin; KOS = Kin Over Stranger; SFH = Failed to Help Stranger; and KFH = Failed to Help Kin. Total N = 305.

Exploratory Analyses of Obligation Principles
After reading and making judgments of all scenarios, participants were then asked to rate their agreement with four general principles related to obligation: (1) If people have obligations to family members, (2) if people do not have obligations to strangers, (3) if people do not deserve praise for helping family members, and (4) if people deserve praise for helping strangers (1 = “strongly disagree” to 9 = “strongly agree”). See SOM Table 6 for descriptive statistics.
SOM Table 6
Means, SDs, and correlations of principle endorsements in Study 4.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Family Obligation</td>
<td>6.24</td>
<td>2.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. No Stranger Obligation</td>
<td>5.79</td>
<td>2.24</td>
<td>*-.13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. No Family Praise</td>
<td>4.10</td>
<td>2.17</td>
<td>-.01</td>
<td>.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Stranger Praise</td>
<td>6.93</td>
<td>1.82</td>
<td>**.18</td>
<td></td>
<td>.00</td>
<td>***-.40</td>
</tr>
</tbody>
</table>

Note. *p < .05; **p < .01; ***p < .001.

Dependent Variable Wording and Scales

*Moral Goodness.* “How morally bad or good is [agent’s name] as a person?”
(1 = “extremely morally bad”; 5 = “neither bad nor good”; 9 = “extremely morally good”)

*Trustworthiness.* “How untrustworthy or trustworthy is [agent’s name]?”
(1 = “extremely untrustworthy”; 5 = “neither untrustworthy not trustworthy”; 9 = “extremely trustworthy”)

*Obligation.* “Did [agent’s name] violate or fulfill an obligation s/he had?”
(1 = “completely violated”; 5 = “neither violated nor fulfilled”; 9 = “completely fulfilled”)

*Rightness.* “How wrong or right were [agent’s name’s] actions?”
(1 = “extremely wrong”; 5 = “neither wrong nor right”; 9 = “extremely right”)

*Praise.* “How morally blameworthy or praiseworthy were [agent’s name’s] actions?”
(1 = “extremely blameworthy”; 5 = “neither blameworthy nor praiseworthy”; 9 = “extremely praiseworthy”)

*Fairness.* “Was the outcome of [agent’s name’s] behavior unfair or fair?”
(1 = “completely fair”; 5 = “neither unfair nor fair”; 9 = “completely fair”)

*Predictability.* “How unpredictable or predictable is [agent’s name’s] future behavior?”
(1 = “extremely unpredictable”; 5 = “neither unpredictable nor predictable”; 9 = “extremely predictable”)

*Typicality.* “How likely is it that the average person would do what [agent’s name] did?”
(1 = “extremely unlikely”; 5 = “neither unlikely nor likely”; 9 = “extremely likely”)
Obligation Principle Wording and Scales

Family Obligation. “People have obligations to help their family members.”

(1 = “strongly disagree”; 5 = “neither disagree nor agree”; 9 = “strongly agree”)

No Stranger Obligation. “People do not have obligations to help people they do not know.”

(1 = “strongly disagree”; 5 = “neither disagree nor agree”; 9 = “strongly agree”)

No Family Praise. “People do not deserve praise for helping their family members.”

(1 = “strongly disagree”; 5 = “neither disagree nor agree”; 9 = “strongly agree”)

Stranger Praise. “People deserve praise for helping people they do not know.”

(1 = “strongly disagree”; 5 = “neither disagree nor agree”; 9 = “strongly agree”)