What we owe to family:

The impact of special obligations on moral judgment

Ryan M. McManus

Liane Young

Department of Psychology,

Boston College, Chestnut Hill, MA 02467

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 recommit 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.

	Stranger	Kin
No Choice	A new tenant is moving into an apartment down the hall from John. She is <u>a stranger.</u> John helps his <u>new neighbor</u> move their furniture in.	A new tenant is moving into an apartment down the hall from John. She is <u>his cousin whom he has not seen or spoken to in years.</u> John helps his <u>cousin</u> move their furniture in.
Choice	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 <u>his cousin</u>, John helps <u>his other new neighbor</u> move their furniture in.	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 <u>the stranger</u>, John helps <u>his cousin</u> move their furniture in.
Failed to Help	A new tenant is moving into an apartment down the hall from John. She is <u>a stranger.</u> Rather than helping <u>the stranger</u> move their furniture in, John stays inside his apartment and plays video games.	A new tenant is moving into an apartment down the hall from John. She is <u>his cousin whom he has not seen or spoken to in years.</u> Rather than helping <u>his cousin</u> move their furniture in, John stays inside his apartment and plays video games.

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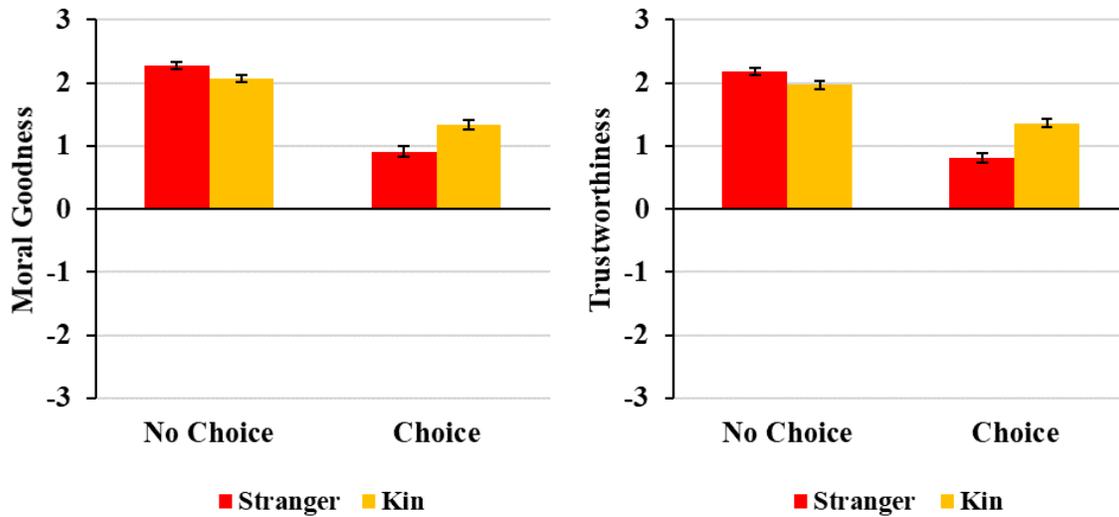
Figure 1. Judgments of target agent in Study 1. Error bars represent +/- 1 SEM.

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, $M_{Age} = 38.30$).

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").[2]

Results

Descriptive statistics for moral goodness and trustworthiness are shown in Figure 2. Data were analyzed with identical mixed effects models as Study 1.
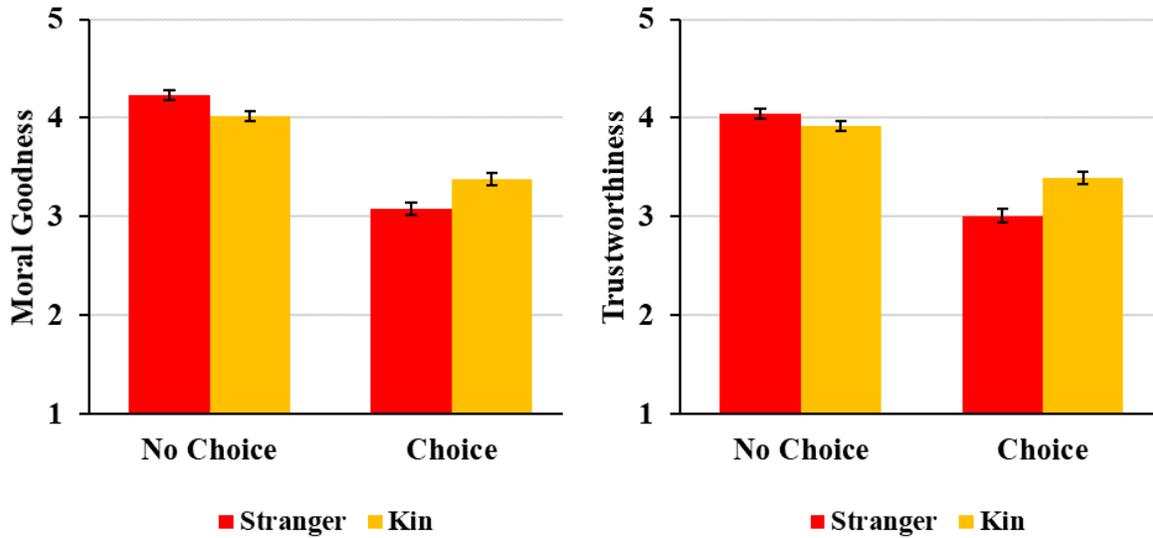
Figure 2. Judgments of target agent in Study 2. Error bars represent +/- 1 SEM.

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, $\chi^2(1) = 49.02$, $p < .001$.

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

		No Choice				Choice			
		Kin	Stranger	*t*	*p*	Kin	Stranger	*t*	*p*
Study 1	*Moral Goodness*	6.07 (5.86, 6.29)	6.27 (6.05, 6.49)	3.00	.003	5.35 (5.13, 5.57)	4.91 (4.69, 5.12)	6.66	< .001
	Trustworthiness	5.97 (5.77, 6.18)	6.19 (5.98, 6.40)	3.20	.001	5.37 (5.16, 5.57)	4.81 (4.60, 5.02)	8.35	< .001
Study 2	*Moral Goodness*	4.02 (3.86, 4.18)	4.24 (4.08, 4.40)	4.27	< .001	3.38 (3.22, 3.54)	3.07 (2.91, 3.23)	6.02	< .001
	Trustworthiness	3.92 (3.78, 4.06)	4.05 (3.91, 4.19)	2.49	.013	3.39 (3.25, 3.52)	3.00 (2.86, 3.13)	7.50	< .001

Note. In Study 1, moral goodness and trustworthiness were rated on 1-7 scales (1 = "extremely bad/untrustworthy" to 4 = "neither bad nor good/untrustworthy nor trustworthy" to 7 = "extremely good/trustworthy). Study 1 total $N = 209$; total observations per variable = 1672. In Study 2, moral goodness and trustworthiness were rated on 1-5 scales (1 = "not at all good/trustworthy" to "extremely good/trustworthy"). Study 2 total $N = 193$; total observations per variable = 1544. *p*-values are unadjusted for multiple comparisons as all comparisons were pre-registered.

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.[3] 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").[4]

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-

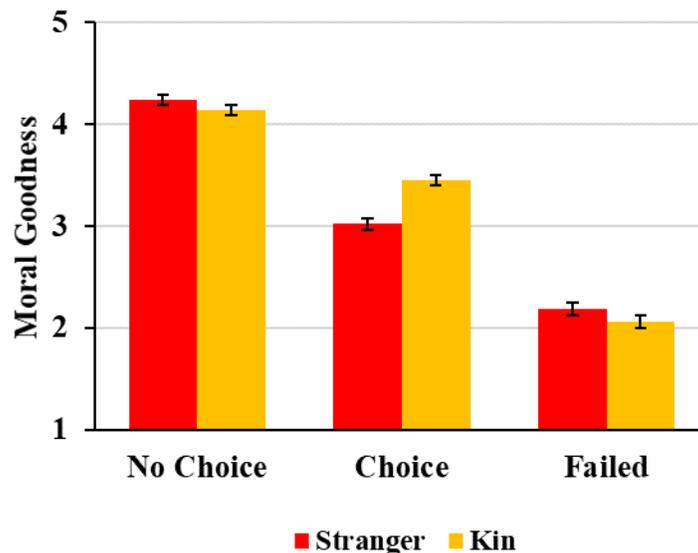4 at the end of Study 4 Methods/Results.

Figure 3. 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, $M_{Age} = 36.33$).

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"). [5]

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.
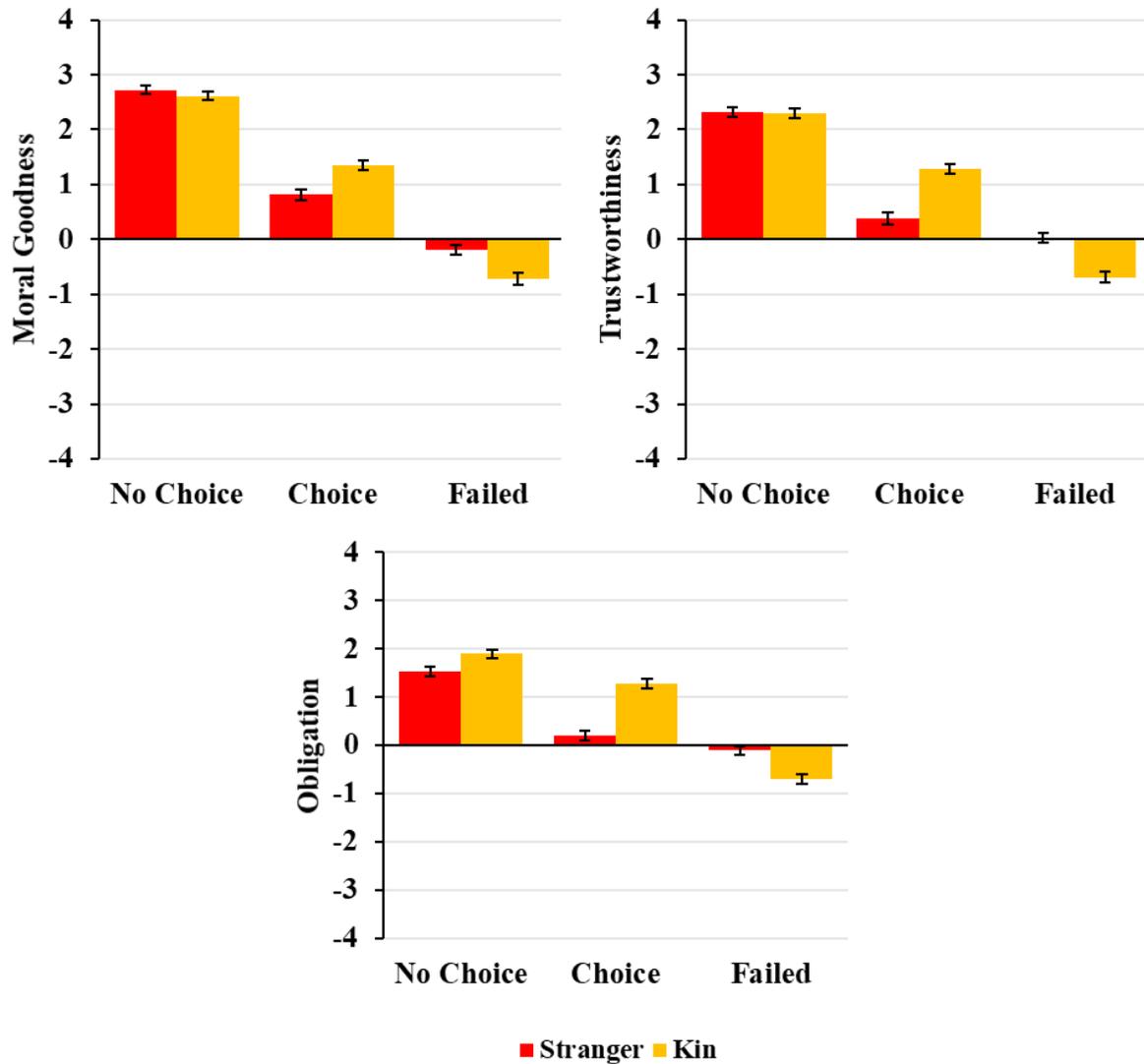
Figure 4. Judgments of target agent in Study 4. Error bars represent +/- 1 SEM.

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.

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

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.

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

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, rs(301) ≥ .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) ≥ .40, ps < .001.

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

		Model 1					Model 2				
	Predictor	b (SE)	β	sp-r	t	p	b (SE)	β	sp-r	t	p
SNC – KNC	*Predictability*	.10 (.04)	.16	.15	2.58	.010	.06 (.04)	.11	.10	1.76	.080
Moral Goodness	*Typicality*	-.07 (.03)	-.15	-.13	-2.30	.022	-.09 (.03)	-.18	-.16	-2.97	.003
	Obligation	-	-	-	-	-	.18 (.04)	.26	.25	4.48	< .001
		F(2, 302) =	**4.15**		R^2 = **.03**	.017	F(3, 301) =	**9.62**		R^2 = **.09**	< .001
										ΔR^2 = **.06**	< .001
KOS – SOK	*Predictability*	.29 (.05)	.35	.28	5.63	< .001	.11 (.05)	.13	.10	2.37	.018
Moral Goodness	*Typicality*	.17 (.05)	.22	.18	3.54	< .001	.03 (.04)	.03	.03	0.64	.521
	Obligation	-	-	-	-	-	.56 (.05)	.61	.50	12.25	< .001
		F(2, 302) =	**51.09**		R^2 = **.25**	< .001	F(3, 301) =	**100.85**		R^2 = **.50**	< .001
										ΔR^2 = **.25**	< .001
SFH - KFH	*Predictability*	.18 (.05)	.20	.19	3.93	< .001	.10 (.04)	.12	.11	2.75	.006
Moral Goodness	*Typicality*	.33 (.04)	.47	.45	9.34	< .001	.19 (.03)	.27	.24	6.09	< .001
	Obligation	-	-	-	-	-	.47 (.04)	.52	.46	11.63	< .001
		F(2, 302) =	**68.71**		R^2 = **.31**	< .001	F(3, 301) =	**111.25**		R^2 = **.53**	< .001
										ΔR^2 = **.22**	< .001
	Predictor	b (SE)	β	sp-r	t	p	b (SE)	β	sp-r	t	p
SNC – KNC	*Predictability*	.12 (.04)	.18	.16	2.78	.006	.10 (.04)	.14	.12	2.19	.030
Trustworthiness	*Typicality*	-.01 (.04)	-.02	-.02	-0.37	.714	-.03 (.04)	-.05	-.05	-0.80	.427
	Obligation	-	-	-	-	-	.15 (.05)	.18	.17	3.05	.003
		F(2, 302) =	**4.34**		R^2 = **.03**	.014	F(3, 301) =	**6.06**		R^2 = **.06**	.001
										ΔR^2 = **.03**	.003
KOS - SOK	*Predictability*	.34 (.05)	.37	.30	6.28	< .001	.15 (.05)	.16	.13	3.19	.002
Trustworthiness	*Typicality*	.24 (.05)	.28	.22	4.74	< .001	.09 (.04)	.11	.08	2.11	.036
	Obligation	-	-	-	-	-	.57 (.05)	.57	.46	11.79	< .001
		F(2, 302) =	**73.25**		R^2 = **.33**	< .001	F(3, 301) =	**117.52**		R^2 = **.54**	< .001
										ΔR^2 = **.21**	< .001
SFH - KFH	*Predictability*	.13 (.05)	.14	.13	2.80	.005	.08 (.04)	.08	.07	1.71	.089
Trustworthiness	*Typicality*	.38 (.04)	.50	.48	10.02	< .001	.27 (.04)	.35	.31	7.22	< .001
	Obligation	-	-	-	-	-	.38 (.05)	.40	.35	8.19	< .001
		F(2, 302) =	**68.94**		R^2 = **.31**	< .001	F(3, 301) =	**78.37**		R^2 = **.44**	< .001
										ΔR^2 = **.13**	< .001

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

References

Bartels, D.M., Bauman, C.W., Cushman, F.A., Pizarro, D.A., and McGraw, A.P. (2016), "Moral

    judgment and decision making," In G. Keren & G. Wu (Eds.) *The Wiley Blackwell*

    Handbook of Judgment and Decision Making. Chichester, UK: Wiley.

Bearne, S. (2017, Dec. 14). The millennials donating 10% of their pay to save the world.

    Retrieved from https://www.theguardian.com/money/2017/dec/04/millennials-donating-

    10-of-their-pay-to-save-the-world

Bloom, P. (2011). Family, community, trolley problems, and the crisis in moral psychology. *The*

    Yale Review, 99(2), 26-43.

Brambilla, M., Carraro, L., Castelli, L., & Sacchi, S. (2019). Changing impressions: Moral

    character dominates impression updating. *Journal of Experimental Social Psychology,*

    82, 64-73.

Burnstein, E., Crandall, C., & Kitayama, S. (1994). Some neo-Darwinian decision rules for

    altruism: Weighing cues for inclusive fitness as a function of the biological importance of

    the decision. *Journal of Personality and Social Psychology, 67*(5), 773-789.

Chalik, L., & Dunham, Y. (2018). Beliefs about moral obligation structure children's social

    category-based expectations. *Child Development*.

Cottrell, C. A., Neuberg, S. L., & Li, N. P. (2007). What do people desire in others? A

    sociofunctional perspective on the importance of different valued characteristics. *Journal*

    of Personality and Social Psychology, 92, 208–231.

Curry, O.S., Chesters, M.J., & Van Lissa, C.J. (2019). Mapping morality with a compass:

    Testing the theory of 'morality-as-cooperation' with a new questionnaire. *Journal of*

    Research in Personality, 78, 106-124.

Curry, O.S., Mullins, D.A., & Whitehouse, H. (in press). Is it good to cooperate? Testing the

      theory of morality-as-cooperation in 60 societies. *Current Anthropology.*

DeScioli, P., & Kurzban, R. (2009). Mysteries of morality. *Cognition, 112,* 281-299.

DeScioli, P., & Kurzban, R. (2013). A solution to the mysteries of morality. *Psychological*

      Bulletin, 139(2)*,* 477-496.

Duntley, J., & Buss, D. (2016). The evolution of good and evil. In A.G. Miller (Ed.), *The social*

      psychology of good and evil (2nd edition, pp. 17-40). New York, New York: The Guilford

      Press.

Everett, J.A.C., Faber, N.S., Savulescu, J., & Crockett, M.J. (2018). The costs of being

      consequentialist: Social inference from instrumental harm and impartial beneficence.

      Journal of Experimental Social Psychology, 79, 200-216.

Goodwin, G.P. (2015). Moral character in person perception. *Psychological Science, 24,* 38-44.

Goodwin, G.P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person

      perception and evaluation. *Journal of Personality and Social Psychology, 106*(1), 148-

      168.

Graham, J., Nosek, B.A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. (2011). Mapping the moral

      domain. *Journal of Personality and Social Psychology, 101,* 366-385.

Greitemeyer, T., Rudolph, U., & Weiner, B. (2003). Whom would you rather help: An

      acquaintance not responsible for her plight or a responsible sibling? *The Journal of Social*

      Psychology, 143(3), 331-340.

Haidt, J. & Baron, J. (1996). Social roles and the moral judgment of acts and omissions.

      European Journal of Social Psychology, 26, 201-218.

Henrich, J., Heine, S.J., & Norenzayan, A. (2010a). The weirdest people in the world?

    Behavioral and Brain Science, 33, 61-83.

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.

Jeske, D. (1998). Family, friends, and special obligations. *Canadian Journal of Philosophy,*

    28(4), 527-556.

Judd, C.M., Westfall, J., & Kenny, D.A. (2012). Treating stimuli as a random factor in social

    psychology: A new and comprehensive solution to a pervasive but largely ignored

    problem. *Journal of Personality and Social Psychology, 103*(1), 54-69.

Kurzban, R., DeScioli, P., & Fein, D. (2012). Hamilton vs. Kant: Pitting adaptations for altruism

    against adaptations for moral judgment. *Evolution and Human Behavior, 33,* 323-333.

Lee, J., & Holyoak, K.J. (2018). "But he's my brother": How family obligation impacts moral

    judgments. In C. Kalish, M. Rau, J. Zhu, & T.T. Rogers (Eds.), *Proceedings of the 40th*

    Annual Conference of the Cognitive Science Society. Austin, TX: Cognitive Science

    Society.

Linke, L.H. (2012). Social closeness and decision making: Moral, attributive, and emotional

    reactions to third party transgressions. *Current Psychology, 31*(3), 291-312.

MacFarquhar, L. (2015). *Strangers drowning: Grappling with impossible idealism, drastic*

    choices, and the overpowering urge to help. New York, New York: Penguin Press.

Madsen, E.A., Tunney, R.J., Fieldman, G., Plotkin, H.C., Dunbar, R.I.M., Richardson, J., &

    McFarland, D. (2007). Kinship and altruism: A cross-cultural experimental study. *British*

    Journal of Psychology, 98, 339-359.

Marshall, J. & Bloom, P. (in prep). Children's and adult's reasoning about prosocial obligations

    in relational contexts.

Mende-Siedlecki, P., Baron, S. G., & Todorov, A. (2013). Diagnostic value underlies

    asymmetric updating of impressions in the morality and ability domains. *The Journal of*

    Neuroscience, 33, 19406–19415.

Morgan, G.S., Mullen, E., & Skitka, L.J. (2010). When values and attributions collide: Liberals'

    and conservatives' values motivate attributions for alleged misdeeds. *Personality and*

    Social Psychology Bulletin, 36(9), 1241-1254.

Niemi, L., Wasserman, E., & Young, L. (2018). The behavioral and neural signatures of distinct

    conceptions of fairness. *Social Neuroscience, 13*(4), 399-415.

Olson, K. R., & Spelke, E.S. (2008). Foundations of cooperation in young children. *Cognition,*

    108, 222-231.

Rai, T.S., & Fiske, A.P. (2011). Moral psychology is relationship regulation: Moral motives for

    unity, hierarchy, equality, and proportionality. *Psychological Review, 188,* 57-75.

Shaw, A., & Olson, K. (2014). Fairness as partiality aversion: The development of procedural

    justice. *Journal of Experimental Child Psychology, 119,* 40-53.

Schein, C. & Gray, K. (2018). The theory of dyadic morality: Reinventing moral judgment by

    redefining harm. *Personality and Social Psychology Review, 22*(1), 32-70.

Schulz, J., Bahrami-Rad, D., Beauchamp, J., & Henrich, J. (2018). The origins of WEIRD

    psychology. *Social Science Research Network.*

Sommers, C. (1986). Filial morality. *The Journal of Philosophy, 83*(8), 439-456.

Strombach, T*.,* Weber, B., Hangebrauk, Z., Kenning, P., Karipidis, I.I., Tobler, P.N., &

    Kalenscher, T. (2015). Social discounting involves modulation of neural value signals by

temporoparietal junction. *Proceedings of the National Academy of Sciences of the United States of America, 112*(5), 1619-1624.

Tepe, B., & Aydinli-Karakulak, A. (in press). Beyond harmfulness and impurity: Moral wrongness as a violation of relational motivations. *Journal of Personality and Social Psychology*.

Waytz, A., Dungan, J., & Young, L. (2013). The whistleblower's dilemma and the fairness-loyalty tradeoff. *Journal of Experimental Social Psychology, 49,* 1027-1033.

Supplemental Online Material for:

What we owe to family: The impact of special obligations on moral judgment

Ryan M. McManus & Liane Young

Study 1

1A. 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, $\chi^2(1) = 63.14$, $p < .001$. In No Choice conditions, helping a stranger was judged as significantly righter than helping kin, whereas in Choice conditions, helping kin instead of a stranger was judged as significantly righter than helping a stranger instead of kin (see SOM Table 1 for linear mixed effects means, 95% CIs, t-statistics, and p-values).


Wrongness
(1 = extremely right; 4 = neither; 7 = extremely wrong)
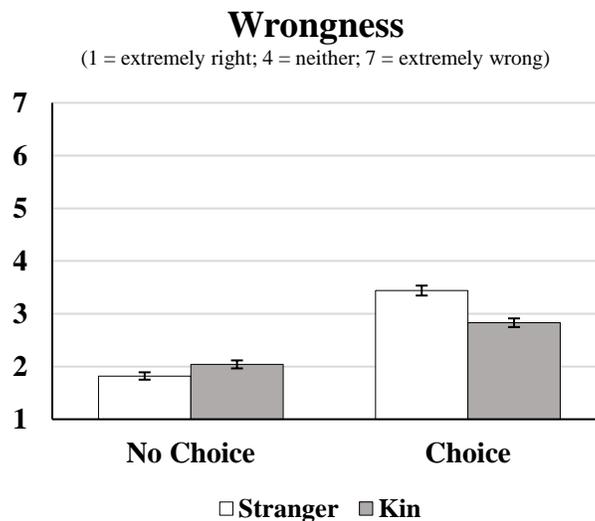
SOM Figure 1.1. Wrongness judgments in Study 1. Error bars represent +/- 1 SEM.

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, $\chi^2(1) = 27.63$, $p < .001$. In No Choice conditions, helping a stranger was judged as significantly more diagnostic of future behavior than helping kin, whereas in

Choice conditions, helping kin instead of a stranger was judged as significantly more diagnostic of future behavior than helping a stranger instead of kin (see SOM Table 1 for linear mixed effects means, 95% CIs, *t*-statistics, and *p*-values).

Diagnosticity
(1 = strongly disagree; 4 = neither; 7 = strongly agree)
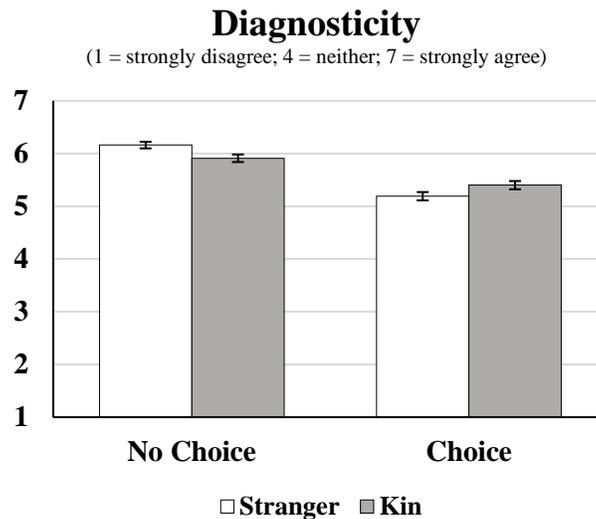
SOM Figure 1.2. Diagnosticity judgments in Study 1. Error bars represent +/- 1 SEM.

1B. An Explanation for Reported Effect Sizes

To increase the interpretability of effect sizes across studies, and more generally, to increase transparency, for all traditional repeated-measures analyses, we have reported partial eta-squared and generalized eta-squared for ANOVAs, and Cohen's d_z, Hedge's g_{av}, and the common language effect size (CLES) for all one-to-one comparisons. We have provided all relevant information so that other researchers can re-calculate these statistics as well.

The reporting of and distinguishing of effect sizes is important for at least four reasons: (1) to identify that the appropriate effect size was computed given the research design, (2) to allow researchers to decide whether a specific effect is meaningful given the research context, (3) to facilitate a cumulative science in which all published studies have enough statistical

information in them for other researchers to infer power for the reported study and plan power

for future studies, and (4) to easily allow for the published data to be used in future meta-

analyses (Lakens, 2013).

Partial eta-squared and generalized eta-squared are reported alongside the repeated-

measures ANOVAs. Partial eta-squared is reported so that other researchers can conduct power

analyses for similar within-subjects designs, whereas generalized eta-squared is reported so that

other researchers can use these results in future meta-analyses that collapse across both within-

and between-subjects designs. Additionally, our OSF page for this project (https://osf.io/bnwdv/)

provides all relevant information needed to calculate partial Omega-squared and generalized

Omega-squared.

Cohen's d_z and Hedge's g_{av} are reported as the standardized measures of effect size for

one-to-one within-subjects comparisons. Cohen's d_z (which takes into account the correlation

between measures) is reported because it is typically used in power analyses for within-subjects

comparisons and therefore can be used by other researchers to appropriately power similarly

designed within-subjects comparisons (as well as to compute achieved power for a reported

effect). Hedge's g_{av} and its confidence intervals are reported because it is typically less biased

than Cohen's d_{av}, which can both be used to make effect size comparisons across within- and

between-subjects designs. Hedge's g_{av} is recommended to be reported so it can be used in future

meta-analyses (though, importantly, as sample size increases, the difference between Hedge's g_{av}

and Cohen's d_{av} becomes smaller; see Lakens, 2013). Hedge's g_{av} is reported here instead of

Hedge's g_{rm} so that these data can be used in future meta-analyses that investigate effect sizes

collapsed across within- and between-subjects designs. However, we have provided all

information necessary to calculate Cohen's d_{av}, Cohen's d_{rm}, and Hedge's g_{rm} as well.

Finally, to enhance interpretability (and potentially, the meaningfulness) of our effects, we have reported a *common language effect size* (CLES) for all comparisons of interest (see McGraw and Wong, 1992, for a more detailed description of the CLES). The CLES translates a standardized effect size into an easy-to-understand percentage, such that a researcher can say, for example, "After controlling for individual differences, and taking into account the correlation between Condition A and Condition B, the likelihood that a participant will rate Condition A higher than Condition B is XX%." We report this effect size to more accurately describe, and make clearer the magnitude of, our effects for a more general audience (see Lakens, 2013, for an explanation of how a between-subjects effect can be translated into a CLES, and see https://osf.io/6zxd2/ to download his spreadsheet that will compute this effect size for both design types).

1C. Traditional Repeated-Measures Analyses

Moral Goodness. A 2 (Relationship) x 2 (Choice) repeated-measures ANOVA was conducted and yielded a main effect of Relationship, $F(1, 208) = 5.08$, $p = .025$, $\eta_p^2 = .02$, $\eta_G^2 = .003$, such that agents who helped kin were judged as more morally good than agents who helped a stranger; a main effect of Choice, $F(1, 208) = 296.52$, $p < .001$, $\eta_p^2 = .59$, $\eta_G^2 = .220$, 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, $F(1, 208) = 34.72$, $p < .001$, $\eta_p^2 = .14$, $\eta_G^2 = .025$. No conditions crossed below the mid-point of "neither bad nor good," and so all results are presented in terms of "moral goodness," with higher scores meaning more morally good. 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 ($M = 6.28$, $SD = 0.81$) than agents who helped kin ($M = 6.07$,

$SD = 0.91$), $t(208) = 3.55$, $p < .001$, Cohen's $d_z = 0.25$, Hedge's $g_{av} = 0.23$, 95% CI [0.11, 0.37].

The common language effect size (CLES) indicates that after controlling for individual

differences, and taking into account the correlation between conditions, $r(208) = 0.549$, the

likelihood that a participant judges agents who helped a stranger as more morally good than

agents who helped kin is 60%. In the Choice conditions, however, agents who helped kin instead

of a stranger were judged as more morally good ($M = 5.34$, $SD = 1.01$) than agents who helped a

stranger instead of kin ($M = 4.91$, $SD = 1.18$), $t(208) = 4.88$, $p < .001$, Cohen's $d_z = 0.34$,

Hedge's $g_{av} = 0.39$ [0.23, 0.56]. The CLES indicates that after controlling for individual

differences, and taking into account the correlation between conditions, $r(208) = 0.324$, the

likelihood that a participant judges agents who helped kin instead of a stranger as more morally

good than agents who helped a stranger instead of kin is 63%.

    Trustworthiness. A 2 (Relationship) x 2 (Choice) repeated-measures ANOVA showed a

main effect of Relationship, $F(1, 208) = 10.63$, $p = .001$, $\eta_p^2 = .05$, $\eta_G^2 = .007$, such that agents

were judged as more trustworthy when they helped kin than when they helped a stranger; a main

effect of Choice, $F(1, 208) = 243.63$, $p < .001$, $\eta_p^2 = .54$, $\eta_G^2 = .194$, 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, $F(1,$

$208) = 47.81$, $p < .001$, $\eta_p^2 = .19$, $\eta_G^2 = .034$. No conditions crossed the below the mid-point of

"neither untrustworthy nor trustworthy," and so all results are presented in terms of

"trustworthiness," with higher scores meaning more trustworthy. In the No Choice conditions,

agents who helped a stranger were judged as significantly more trustworthy ($M = 6.19$, $SD =$

0.89) than agents who helped kin ($M = 5.97$, $SD = 0.89$), $t(208) = 4.22$, $p < .001$, Cohen's $d_z =$

0.29, Hedge's g_{av} = 0.24 [0.13, 0.36]. The CLES indicates that after controlling for individual

differences, and taking into account the correlation between conditions, $r(208)$ = 0.658, the

likelihood that a participant judges agents who helped a stranger as more trustworthy than agents

who helped kin is 61%. In Choice conditions, however, agents who helped kin instead of a

stranger were judged as significantly more trustworthy (M = 5.36, SD = 1.06) than agents who

helped a stranger instead of kin (M = 4.81, SD = 1.19), $t(208)$ = 5.88, p < .001, Cohen's d_z =

0.41, Hedge's g_{av} = 0.48 [0.32, 0.65]. The CLES indicates that after controlling for individual

differences, and taking into account the correlation between conditions, $r(208)$ = 0.286, the

likelihood that a participant judges agents who helped kin instead of a stranger as more

trustworthy than agents who helped a stranger instead of kin is 66%.

    Wrongness. A 2 (Relationship) x 2 (Choice) repeated-measures ANOVA yielded a main

effect of Relationship, $F(1, 208)$ = 10.84, p = .001, η_p^2 = .05, η_G^2 = .007, such that it was judged

as righter when agents helped kin than when they helped a stranger; a main effect of Choice, $F(1,$

$208)$ = 275.66, p < .001, η_p^2 = .57, η_G^2 = .210, such that help was judged as righter 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, η_p^2 = .20,

η_G^2 = .031. No conditions crossed above the mid-point of "neither right nor wrong," and so all

means are presented in terms of "rightness," with lower scores meaning righter. In the No Choice

conditions, helping a stranger was judged as significantly righter (M = 1.82, SD = 1.01) than

helping kin (M = 2.04, SD = 1.10), $t(208)$ = 3.53, p = .001, Cohen's d_z = 0.24, Hedge's g_{av} = 0.21

[0.09, 0.33]. The CLES indicates that after controlling for individual differences, and taking into

account the correlation between conditions, $r(208)$ = 0.638, the likelihood that a participant

judges helping a stranger as righter than helping kin is 60%. In the Choice conditions, however,

helping kin instead of a stranger was judged as significantly righter ($M = 2.83$, $SD = 1.87$) than

helping a stranger instead of kin ($M = 3.44$, $SD = 1.36$), $t(208) = 6.16$, $p = < .001$, Cohen's $d_z =$

0.43, Hedge's $g_{av} = 0.48$ [0.32, 0.64]. The CLES indicates that after controlling for individual

differences, and taking into account the correlation between conditions, $r(208) = 0.364$, the

likelihood that a participant judges helping kin instead of a stranger as righter than helping a

stranger instead of kin is 67%.

Diagnosticity. A 2 (Relationship) x 2 (Choice) repeated-measures ANOVA revealed no

main effect of Relationship, $F(1, 208) = 0.27$, $p = .604$, $\eta_p^2 = .00$, $\eta_G^2 = .000$. There was,

however, a main effect of Choice, $F(1, 208) = 158.95$, $p < .001$, $\eta_p^2 = .43$, $\eta_G^2 = .109$, such that

agents' actions were judged to be more diagnostic of future behavior when no alternative

beneficiary was available to help than when there were two potential beneficiaries to help; and a

significant Relationship x Choice interaction, $F(1, 208) = 25.65$, $p < .001$, $\eta_p^2 = .11$, $\eta_G^2 = .011$.

In the No Choice conditions, helping a stranger was judged as being significantly more

diagnostic of 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 $d_z = 0.32$, Hedge's $g_{av} = 0.25$ [0.15, 0.36]. The CLES

indicates that after controlling for individual differences, and taking into account the correlation

between conditions, $r(208) = 0.736$, the likelihood that a participant judges helping a stranger as

being more diagnostic of future behavior than helping kin is 64%. In the Choice conditions,

however, helping kin instead of a stranger was judged as being 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 $d_z = 0.20$, Hedge's $g_{av} = 0.18$ [0.06, 0.30]. The

CLES indicates that after controlling for individual differences, and taking into account the

correlation between conditions, $r(208) = 0.594$, the likelihood that a participant judges helping

kin instead of a stranger as being more diagnostic of future behavior than helping a stranger

instead of kin is 58%.

1D. 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

2A. 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 a stranger was judged as significantly less likely than helping kin, and similarly, in Choice conditions, helping a stranger instead of kin was also judged as significantly less likely than helping kin instead of a stranger (see SOM Table 1 for linear mixed effects means, 95% CIs, *t*-statistics, and *p*-values).

Typicality

(1 = not at all likely; 5 = extremely likely)
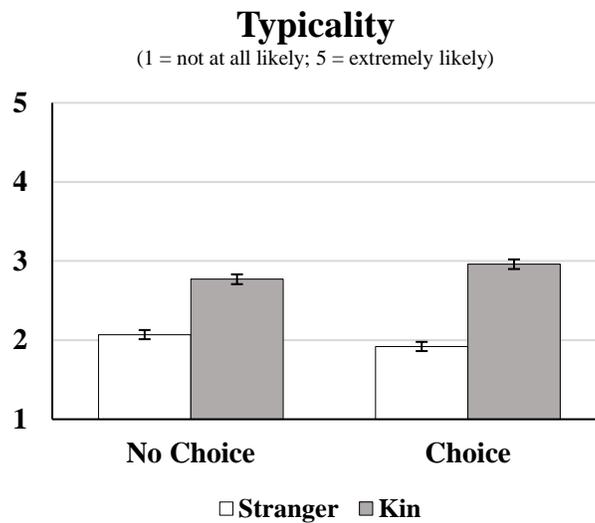
SOM Figure 2.1. Typicality judgments in Study 2. Error bars represent +/- 1 SEM.

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, for agents who helped a stranger, their future behavior was judged as significantly more unpredictable than

agents who helped kin, and similarly, in Choice conditions, for agents who helped a stranger instead of kin, their future behavior was judged as significantly more unpredictable than agents who helped kin instead of a stranger (see SOM Table 1 for linear mixed effects means, 95% CIs, *t*-statistics, and *p*-values).

Unpredictability
(1 = not at all unpredictable; 5 = extremely unpredictable)

SOM Figure 2.2. Unpredictability judgments in Study 2. Error bars represent +/- 1 SEM.

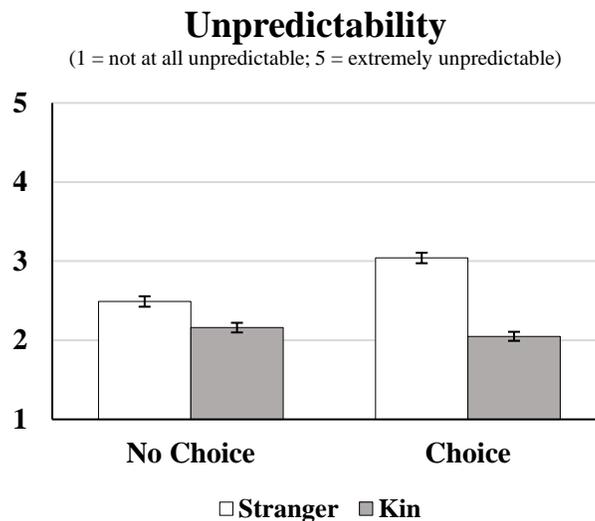*Harm.* There was a significant interaction between Relationship and Choice, $B = -0.175$, $SE_B = 0.059$, $t = -2.94$, $p = .003$; this interaction model provided a better fit to the data than a main effects-only model, $\chi^2(1) = 8.61$, $p = .003$. In No Choice conditions, participants judged scenarios involving helping a stranger as containing marginally less harm than scenarios involving helping kin, whereas in Choice conditions, participants judged scenarios involving helping a stranger instead of kin as containing significantly more harm than scenarios involving helping kin instead of a stranger (see SOM Table 1 for linear mixed effects means, 95% CIs, *t*-statistics, and *p*-values).
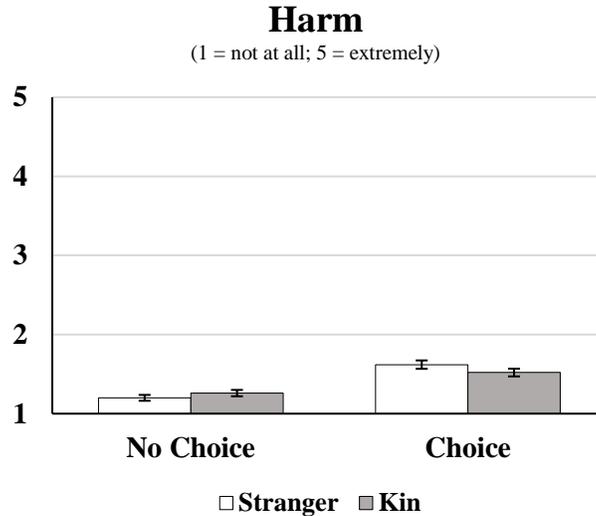
SOM Figure 2.3. Harm judgments in Study 2. Error bars represent +/- 1 SEM.

2B. Mixed Effects Table for Studies 1-2

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

		No Choice				Choice			
		Kin	Stranger	*t*	*p*	Kin	Stranger	*t*	*p*
Study 1	*Wrongness*	2.04 (1.77, 2.30)	1.82 (1.55, 2.08)	2.93	.003	2.82 (2.56, 3.09)	3.45 (3.18, 3.71)	8.42	< .001
	Diagnosticity	5.91 (5.74, 6.08)	6.16 (5.99, 6.33)	4.08	< .001	5.40 (5.23, 5.57)	5.19 (5.02, 5.36)	3.38	< .001
Study 2	*Typicality*	2.74 (2.45, 3.04)	2.10 (1.81, 2.40)	10.74	< .001	2.95 (2.65, 3.24)	1.92 (1.63, 2.21)	17.28	< .001
	Unpredictability	2.18 (1.97, 2.39)	2.46 (2.25, 2.67)	4.44	< .001	2.05 (1.84, 2.26)	3.04 (2.83, 3.25)	15.74	< .001
	Harm	1.27 (1.13, 1.40)	1.19 (1.06, 1.32)	1.81	.070	1.52 (1.39, 1.66)	1.62 (1.49, 1.76)	2.35	.019

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.

2C. Traditional Repeated-Measures Analyses

Moral Goodness. A 2 (Relationship) x 2 (Choice) repeated-measures ANOVA revealed

no main effect of Relationship, $F(1, 192) = 1.03$, $p = .311$, $\eta_p^2 = .01$, $\eta_G^2 = .000$; a main effect of

Choice, $F(1, 192) = 341.04$, $p < .001$, $\eta_p^2 = .64$, $\eta_G^2 = .249$, 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, $F(1, 192) = 42.63$, $p < .001$, $\eta_p^2 = .18$, $\eta_G^2 = .026$. 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 ($M = 4.23$, $SD = 0.70$) than agents who helped kin ($M = 4.02$, $SD = 0.74$), $t(192) = 4.97$, $p < .001$, Cohen's $d_z = 0.36$, Hedge's $g_{av} = 0.30$, 95% CI [0.18, 0.42]. The common language effect size (CLES) indicates that after controlling for individual differences, and taking into account the correlation between conditions, $r(192) = 0.656$, the likelihood that a participant judges agents who helped a stranger as more moral than agents who helped kin is 64%. In the Choice conditions, however, agents who helped kin instead of a stranger were judged as significantly more morally good ($M = 3.38$, $SD = 0.83$) than agents who helped a stranger instead of kin ($M = 3.08$, $SD = 0.83$), $t(192) = 4.42$, $p = < .001$, Cohen's $d_z = 0.32$, Hedge's $g_{av} = 0.36$ [0.19, 0.52]. The CLES indicates that after controlling for individual differences, and taking into account the correlation between conditions, $r(192) = 0.367$, the likelihood that a participant judges agents who helped kin instead of a stranger as more moral than agents who helped a stranger instead of kin is 62%.

Trustworthiness. A 2 (Relationship) x 2 (Choice) repeated-measures ANOVA showed a main effect of Relationship, $F(1, 192) = 9.64$, $p = .002$, $\eta_p^2 = .05$, $\eta_G^2 = .007$, such that agents were judged as more trustworthy when they helped kin than when they helped a stranger; a main effect of Choice, $F(1, 192) = 273.49$, $p < .001$, $\eta_p^2 = .59$, $\eta_G^2 = .199$, 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, $p < .001$, $\eta_p^2 = .16$, $\eta_G^2 = .025$. In the No Choice conditions, agents who helped a stranger were judged as significantly more trustworthy ($M = 4.04$, $SD = 0.73$) than agents who helped kin ($M = 3.92$, $SD = 0.73$), $t(192) = 2.94$, $p = .004$, Cohen's $d_z = 0.21$, Hedge's $g_{av} = 0.17$ [0.05, 0.28]. The CLES indicates that after controlling for individual differences, and taking into account the correlation between conditions, $r(192) = 0.690$, the likelihood that a participant judges agents who helped a stranger as more trustworthy than agents who helped kin is 58%. In the Choice conditions, however, agents who helped kin instead of a stranger were judged as significantly more trustworthy ($M = 3.39$, $SD = 0.81$) than agents who helped a stranger instead of kin ($M = 3.01$, $SD = 0.89$), $t(192) = 5.30$, $p < .001$, Cohen's $d_z = 0.38$, Hedge's $g_{av} = 0.45$ [0.28, 0.62]. The CLES indicates that after controlling for individual differences, and taking into account the correlation between conditions, $r(192) = 0.313$, the likelihood that a participant judges agents who helped kin instead of a stranger as more trustworthy than agents who helped a stranger instead of kin is 65%.

Typicality. A 2 (Relationship) x 2 (Choice) repeated-measures ANOVA yielded a main effect of Relationship, $F(1, 192) = 245.85$, $p < .001$, $\eta_p^2 = .56$, $\eta_G^2 = .217$, such that helping a stranger was judged to be less likely than helping kin; no main effect of Choice, $F(1, 192) = 0.18$, $p = .671$, $\eta_p^2 = .00$, $\eta_G^2 = .000$; and a significant Relationship x Choice interaction, $F(1, 192) = 15.29$, $p < .001$, $\eta_p^2 = .07$, $\eta_G^2 = .010$. 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 $d_z = 0.79$, Hedge's $g_{av} = 0.84$ [0.67, 1.02]. The CLES indicates that after controlling for individual differences, and taking into account the correlation between conditions, $r(192) = 0.425$, the likelihood that a participant judges helping a stranger as less likely than helping kin is 78%. Similarly, in Choice conditions, helping a stranger instead of

kin was judged as significantly less likely ($M = 1.92$, $SD = 0.81$) than helping kin instead of a

stranger ($M = 2.96$, $SD = 0.85$), $t(192) = 13.60$, $p < .001$, Cohen's $d_z = 0.98$, Hedge's $g_{av} = 1.25$

[1.03, 1.47]. The CLES indicates that after controlling for individual differences, and taking into

account the correlation between conditions, $r(192) = 0.181$, the likelihood that a participant

judges helping a stranger instead of kin as less likely than helping kin instead of a stranger is

84%.

Unpredictability. A 2 (Relationship) x 2 (Choice) repeated-measures ANOVA resulted in

a main effect of Relationship, $F(1, 192) = 142.29$, $p < .001$, $\eta_p^2 = .43$, $\eta_G^2 = .126$, such that when

agents helped a stranger, their future behavior was judged to be more unpredictable than agents

who helped kin; a main effect of Choice $F(1, 192) = 18.44$, $p < .001$, $\eta_p^2 = .09$, $\eta_G^2 = .015$, 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, $F(1, 192) = 56.90$, $p < .001$, $\eta_p^2 = .23$, $\eta_G^2 = .036$. In the No

Choice conditions, when agents helped a stranger, their future behavior was judged as being

significantly more unpredictable ($M = 2.49$, $SD = 0.92$) than the future behavior of agents who

helped kin ($M = 2.16$, $SD = 0.84$), $t(192) = 5.43$, $p < .001$, Cohen's $d_z = 0.39$, Hedge's $g_{av} = 0.36$

[0.23, 0.50]. The CLES indicates that after controlling for individual differences, and taking into

account the correlation between conditions, $r(192) = 0.563$, the likelihood that a participant

judges the future behavior of agents who helped a stranger as more unpredictable than agents

who helped kin is 65%. In Choice conditions, again, for agents who helped a stranger instead of

kin, their future behavior was judged as significantly more unpredictable ($M = 3.04$, $SD = 0.91$)

than the future behavior of agents who helped kin instead of a stranger ($M = 2.05$, $SD = 0.79$),

$t(192) = 12.38$, $p < .001$, Cohen's $d_z = 0.89$, Hedge's $g_{av} = 1.16$ [0.94, 1.38]. The CLES indicates

that after controlling for individual differences, and taking into account the correlation between conditions, $r(192) = 0.153$, the likelihood that a participant judges the future behavior of agents who helped a stranger instead of kin as more unpredictable than agents who helped kin instead of a stranger is 81%.

   Harm. A 2 (Relationship) x 2 (Choice) repeated-measures ANOVA revealed no main effect of Relationship, $F(1, 192) = 0.42$, $p = .520$, $\eta_p^2 = .00$, $\eta_G^2 = .000$; a main effect of Choice, $F(1, 192) = 84.89$, $p < .001$, $\eta_p^2 = .31$, $\eta_G^2 = .069$, such that participants 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, $F(1, 192) = 6.15$, $p = .014$, $\eta_p^2 = .03$, $\eta_G^2 = .004$. In the No Choice conditions, participants perceived marginally less harm in scenarios that involved agents helping a stranger ($M = 1.20$, $SD = 0.52$) than in scenarios that involved agents helping kin ($M = 1.26$, $SD = 0.55$), $t(192) = 1.89$, $p = .061$, Cohen's $d_z = 0.17$, Hedge's $g_{av} = 0.14$ [0.03, 0.24]. The CLES indicates that after controlling for individual differences, and taking into account the correlation between conditions, $r(192) = 0.695$, the likelihood that a participant judges scenarios that involved helping a stranger as containing less harm than scenarios that involved helping kin is 55%. In the Choice conditions, however, participants perceived marginally more harm in scenarios that involved agents helping a stranger instead of kin ($M = 1.62$, $SD = 0.73$) than in scenarios that involved agents helping kin instead of a stranger ($M = 1.52$, $SD = 0.68$), $t(192) = 1.81$, $p = .072$, Cohen's $d_z = 0.13$, Hedge's $g_{av} = 0.14$ [-0.01, 0.28]. The CLES indicates that after controlling for individual differences, and taking into account the correlation between conditions, $r(192) = 0.458$, the likelihood that a participant judges scenarios that involved helping kin instead of a stranger as containing more harm than scenarios that involved helping a stranger instead of kin is 55%.

2D. 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

3A. 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, for agents who helped a stranger, their future behavior was judged to be significantly more unpredictable than agents who helped kin, and similarly, in Choice conditions, for agents who helped a stranger instead of kin, their future behavior was judged to be significantly more unpredictable than agents who helped kin instead of a stranger (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 a stranger, their future behavior was judged to be significantly less unpredictable than agents who failed to help kin. For agents who failed to help kin, their future behavior was also judged to be significantly less unpredictable than agents who helped a stranger instead of kin (see SOM Table 3 for linear mixed effects means, 95% CIs, t-statistics, and p-values).

Unpredictability

(1 = not at all unpredictable; 5 = extremely unpredictable)
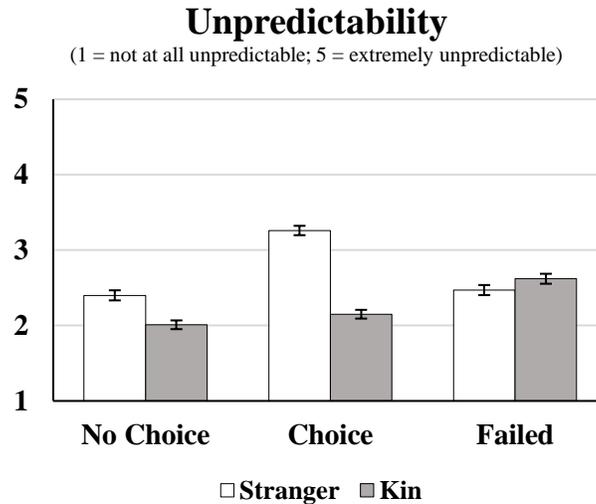
SOM Figure 3. Unpredictability judgments in Study 3. Error bars represent +/- 1 SEM.

3B. Traditional Repeated-Measures Analyses

Moral Goodness (No Choice vs. Choice). A 2 (Relationship) x 2 (Choice) repeated-measures ANOVA was conducted and revealed a main effect of Relationship, $F(1, 303) = 14.58$, $p < .001$, $\eta_p^2 = .05$, $\eta_G^2 = .009$, such that agents were judged as more morally good when they helped kin than when they helped a stranger; a main effect of Choice, $F(1, 303) = 357.70$, $p < .001$, $\eta_p^2 = .54$, $\eta_G^2 = .221$, 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, $F(1, 303) = 20.27$, $p < .001$, $\eta_p^2 = .12$, $\eta_G^2 = .020$. 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 $d_z = 0.13$, Hedge's $g_{av} = 0.11$, 95% CI [0.01, 0.22]. The common language effect size (CLES) indicates that after controlling for individual differences, and taking into account the correlation

between conditions, $r(303) = 0.585$, the likelihood that a participant judges agents who helped a stranger as more moral than agents who helped kin is 55%. In the Choice conditions, however, agents who helped kin instead of a stranger were judged as significantly more morally good ($M = 3.45$, $SD = 0.93$) than agents who helped a stranger instead of kin ($M = 3.02$, $SD = 1.04$), $t(303) = 5.85$, $p = < .001$, Cohen's $d_z = 0.34$, Hedge's $g_{av} = 0.43$ [0.28, 0.58]. The CLES indicates that after controlling for individual differences, and taking into account the correlation between conditions, $r(303) = 0.177$, the likelihood that a participant judges agents who helped kin instead of a stranger as more moral than agents who helped a stranger instead of kin is 63%.

Moral Goodness (Choice vs. Failed). A 2 (Relationship) x 2 (Choice) repeated-measures ANOVA was conducted while adding the Failed to Help conditions and leaving out the No Choice conditions. This revealed a main effect of Relationship, $F(1, 303) = 8.28$, $p = .004$, $\eta_p^2 = .03$, $\eta_G^2 = .005$, such that agents were judged as more morally good when they interacted with kin than when they interacted with a stranger; a main effect of Choice, $F(1, 303) = 297.95$, $p < .001$, $\eta_p^2 = .50$, $\eta_G^2 = .224$, such that agents were judged as more morally good when they helped rather than failed to help; and a significant Relationship x Choice interaction, $F(1, 303) = 31.90$, $p < .001$, $\eta_p^2 = .10$, $\eta_G^2 = .018$. In the Failed to Help conditions, agents who failed to help a stranger were judged as marginally 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 $d_z = 0.11$, Hedge's $g_{av} = 0.12$ [0.00, 0.25]. The CLES indicates that after controlling for individual differences, and taking into account the correlation between conditions, $r(303) = 0.385$, the likelihood that a participant judges an agent who failed to help a stranger as more morally good than an agent who failed to help kin is 54%. Agents who failed to help kin were also judged as less morally good ($M = 2.06$, $SD = 1.10$) than agents who helped a stranger instead of kin ($M = 3.02$, $SD = 1.04$),

$t(303) = 11.26$, $p = < .001$, Cohen's $d_z = 0.65$, Hedge's $g_{av} = 0.90$ [0.73, 1.07]. The CLES

indicates that after controlling for individual differences, and taking into account the correlation

between conditions, $r(303) = 0.028$, the likelihood that a participant judges agents failing to help

kin as less moral than agents who helped a stranger instead of kin is 74%.

   Unpredictability (No Choice vs. Choice). A 2 (Relationship) x 2 (Choice) repeated-

measures ANOVA yielded a main effect of Relationship, $F(1, 303) = 196.62$, $p < .001$, $\eta_p^2 = .39$,

$\eta_G^2 = .111$, such that for agents who helped strangers, their future behavior was judged as being

more unpredictable than agents who helped kin; a main effect of Choice, $F(1, 303) = 78.95$, $p <$

$.001$, $\eta_p^2 = .21$, $\eta_G^2 = .052$, such that when no alternative beneficiary was available to help, an

agent's future behavior was judged as being less unpredictable than when there were two

potential beneficiaries; and a significant Relationship x Choice interaction, $F(1, 303) = 52.73$, p

$< .001$, $\eta_p^2 = .15$, $\eta_G^2 = .027$. 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 the

future behavior of agents helped kin ($M = 2.01$, $SD = 1.00$), $t(303) = 5.81$, $p < .001$, Cohen's $d_z = $

0.24, Hedge's $g_{av} = 0.36$ [0.24, 0.49]. The CLES indicates that after controlling for individual

differences, and taking into account the correlation between conditions, $r(303) = 0.401$, the

likelihood that a participant judges agents who helped a stranger as more unpredictable than

agents who helped kin is 63%. In the Choice conditions, again, for agents who helped a stranger

instead of kin, their future behavior was judged as significantly more unpredictable ($M = 3.26$,

$SD = 1.08$) 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 $d_z = 0.83$, Hedge's $g_{av} = 1.06$ [0.89, 1.23]. The CLES

indicates that after controlling for individual differences, and taking into account the correlation

between conditions, $r(303) = 0.179$, the likelihood that a participant judges agents who helped a

stranger instead of kin as more unpredictable than agents who helped kin instead of a stranger is

80%.

Unpredictability (Choice vs. Failed). A 2 (Relationship) x 2 (Choice) repeated-measures

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, $F(1, 303) = 100.27$,

$p < .001$, $\eta_p^2 = .25$, $\eta_G^2 = .045$, such that for agents who interacted with strangers, their future

behavior was judged as being more unpredictable than agents who interacted with kin; a main

effect of Choice, $F(1, 303) = 6.28$, $p = .013$, $\eta_p^2 = .02$, $\eta_G^2 = .005$, such that an agent's future

behavior was judged as being less unpredictable when they helped rather than when they failed

to help; and a significant Relationship x Choice interaction, $F(1, 303) = 127.70$, $p < .001$, $\eta_p^2 = $

$.30$, $\eta_G^2 = .076$. 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 the

future behavior of agents who failed to help kin ($M = 2.62$, $SD = 1.15$), $t(303) = 2.16$, $p = .032$,

Cohen's $d_z = 0.12$, Hedge's $g_{av} = 0.13$ [0.01, 0.25]. The CLES indicates that after controlling for

individual differences, and taking into account the correlation between conditions, $r(303) = $

0.444, the likelihood that a participant judges agents who failed to help a stranger as less

unpredictable than agents who failed to help kin is 55%. For agents who failed to help kin, their

future behavior was also judged as significantly less unpredictable ($M = 2.62$, $SD = 1.15$) than

the future behavior of agents who helped a stranger instead of kin ($M = 3.26$, $SD = 1.08$), $t(303)$

$= 7.74$, $p < .001$, Cohen's $d_z = 0.44$, Hedge's $g_{av} = 0.57$ [0.42, 0.72]. The CLES indicates that

after controlling for individual differences, and taking into account the correlation between

conditions, $r(303) = 0.172$, the likelihood that a participant judges agents who failed to help a

stranger as less unpredictable than agents who helped a stranger instead of kin is 67%.

3C. 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?"
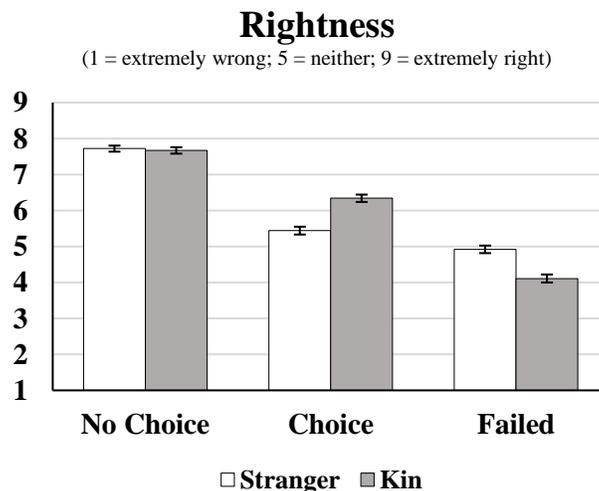
(1 = "not at all unpredictable"; 5 = "extremely unpredictable")

Study 4

4A. 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, $\chi^2(1) = 29.78$, $p < .001$. In No

Choice conditions, helping a stranger was judged as no righter than helping kin, whereas in

Choice conditions, helping kin instead of a stranger was judged as significantly righter than

helping a stranger instead of kin (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, $\chi^2(1) = 76.20$, $p < .001$. In Failed to Help

conditions, failing to help a stranger was judged as significantly less wrong (or significantly

righter) than failing to help kin. 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).

Rightness

(1 = extremely wrong; 5 = neither; 9 = extremely right)

SOM Figure 4.1. Rightness judgments in Study 4. Error bars represent +/- 1 SEM.

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, $\chi^2(1) = 16.29$, $p < .001$. In No Choice conditions, agents who helped a stranger were praised no more than agents who helped kin, whereas in Choice conditions, agents who helped kin instead of a stranger were praised significantly more than agents who helped a stranger instead of kin (see SOM Table 2 for linear mixed effects means, 95% CIs, t-statistics, and p-values).
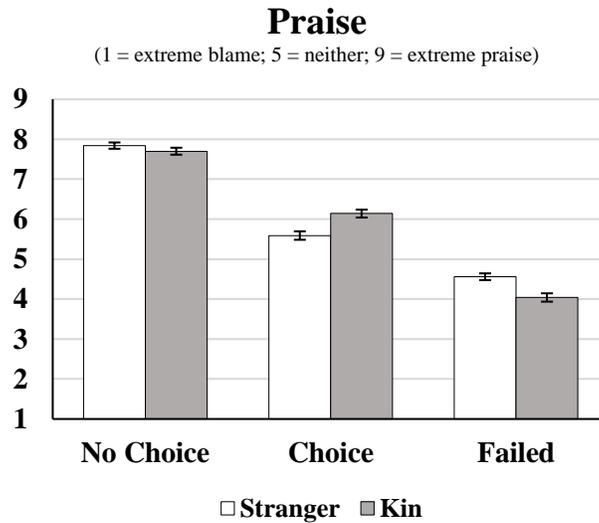
Praise (Choice vs. Failed). There was a significant interaction between Relationship and Choice, $B = -1.056$, $SE_B = 0.177$, $t = -5.97$, $p < .001$; this interaction model provided a better fit to the data than a main effects-only model, $\chi^2(1) = 35.14$, $p < .001$. In Failed to Help conditions, agents who failed to help strangers were blamed significantly less than agents who failed to help kin. Agents who failed to help kin were also blamed significantly more than agents who helped a stranger instead of kin (see SOM Table 3 for linear mixed effects means, 95% CIs, t-statistics, and p-values).

Praise

(1 = extreme blame; 5 = neither; 9 = extreme praise)

□ **Stranger**   ■ **Kin**

SOM Figure 4.2. 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, $\chi^2(1) = 33.50$, $p < .001$. In No Choice conditions, when strangers were helped, it was judged as no fairer than when kin were helped, whereas in Choice conditions, when kin were being helped instead of a stranger, this was judged as significantly fairer than when a stranger was being helped instead of kin (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, $\chi^2(1) = 74.83$, $p < .001$. Failing to help strangers was judged as significantly fairer (or less unfair) than failing to help kin. Failing to help kin was also judged as significantly less fair (or more unfair) than helping a stranger being instead of kin (see SOM Table 3 for linear mixed effects means, 95% CIs, t-statistics, and p-values).

Fairness

(1 = completely unfair; 5 = neither; 9 = completely fair)
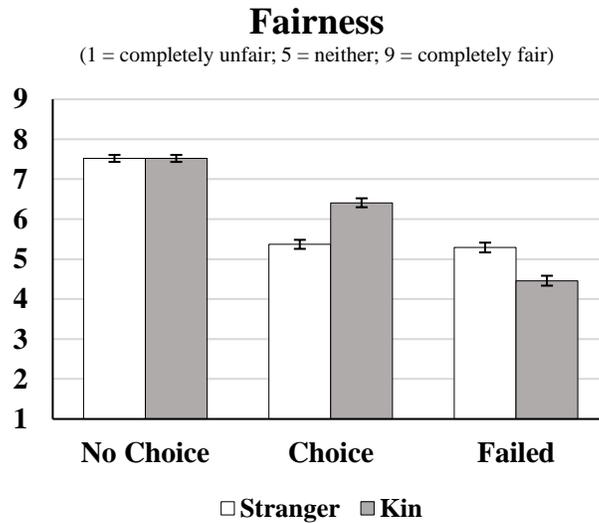
SOM Figure 4.3. Fairness judgments in Study 4. Error bars represent +/- 1 SEM.

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, when agents helped a stranger, their future behavior was judged as significantly less predictable than the future behavior of agents who helped kin. Similarly, in Choice conditions, when agents helped a stranger instead of kin, their future behavior was also judged as significantly less predictable than the future behavior of agents who helped kin instead of a stranger (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 a stranger, their future behavior was judged to be significantly more

predictable than the future behavior of agents who failed to help kin. For agents who failed to help kin, their future behavior was also judged to be significantly more predictable (or less unpredictable) than the future behavior of agents who helped a stranger instead of kin (see SOM Table 3 for linear mixed effects means, 95% CIs, *t*-statistics, and *p*-values).

Predictability

(1 = extremely unpredictable; 5 = neither; 9 = extremely predictable)
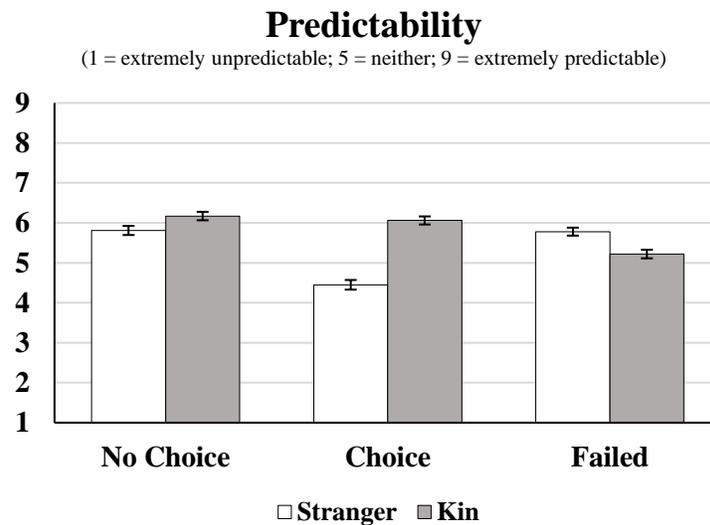

□ **Stranger**   ■ **Kin**

SOM Figure 4.4. 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 a stranger was judged as significantly less likely than helping kin, and similarly, in Choice conditions, helping a stranger instead of kin was also judged as significantly less likely than helping kin instead of a stranger (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, $\chi^2(1) = 250.54$, $p < .001$. Failing to help a

stranger was judged as significantly more likely than failing to help kin. Failing to help kin was

also judged as more likely (or less unlikely) than helping a stranger instead of helping kin (see

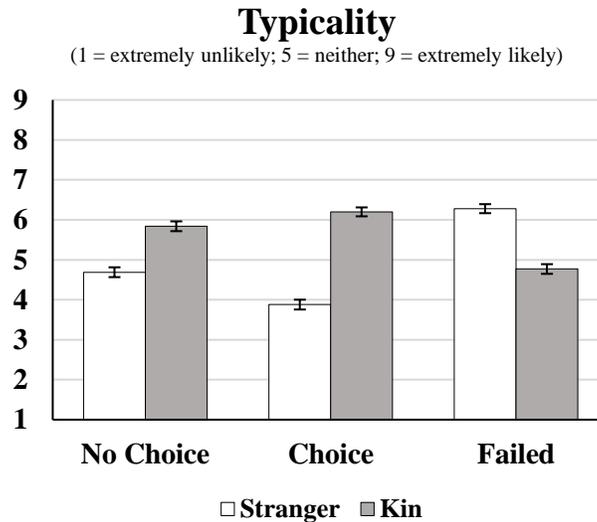SOM Table 3 for linear mixed effects means, 95% CIs, t-statistics, and p-values).

SOM Figure 4.5. Typicality judgments in Study 4. Error bars represent +/- 1 SEM.

4B. Mixed Effects Tables for Studies 3-4

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

		No Choice				Choice			
		Kin	Stranger	t	p	Kin	Stranger	t	p
Study 3	*Unpredictability*	2.03 (1.82, 2.23)	2.38 (2.18, 2.59)	4.88	< .001	2.14 (1.93, 2.34)	3.24 (3.04, 3.45)	15.12	< .001
Study 4	*Rightness*	7.66 (7.41, 7.91)	7.72 (7.47, 7.97)	0.55	.582	6.34 (6.09, 6.59)	5.44 (5.19, 5.69)	7.22	< .001
	Praise	7.70 (7.50, 7.89)	7.84 (7.65, 8.03)	1.19	.234	6.14 (5.95, 6.34)	5.60 (5.40, 5.79)	4.55	< .001
	Fairness	7.51 (7.26, 7.75)	7.53 (7.29, 7.77)	0.14	.893	6.41 (6.17, 6.65)	5.37 (5.13, 5.61)	8.12	< .001
	Predictability	6.13 (5.73, 6.54)	5.80 (5.40, 6.20)	2.54	.011	6.05 (5.64, 6.45)	4.45 (4.05, 4.85)	12.11	< .001
	Typicality	5.77 (5.18, 6.37)	4.68 (4.08, 5.27)	7.53	< .001	6.19 (5.59, 6.78)	3.89 (3.30, 4.49)	15.78	< .001

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.

		Failed to Help				Neglected Kin			
		Kin	Stranger	*t*	*p*	Failed	Stranger Instead	*t*	*p*
Study 3	*Unpredictability*	2.63 (2.45, 2.82)	2.46 (2.28, 2.64)	2.23	.026	2.63 (2.45, 2.82)	3.26 (3.08, 3.45)	8.05	< .001
Study 4	*Rightness*	4.13 (3.80, 4.45)	4.91 (4.58, 5.23)	5.85	< .001	4.13 (3.80, 4.45)	5.43 (5.10, 5.76)	9.76	< .001
	Praise	4.04 (3.76, 4.32)	4.55 (4.27, 4.83)	4.03	< .001	4.04 (3.76, 4.32)	5.59 (5.31, 5.87)	12.35	< .001
	Fairness	4.47 (4.08, 4.86)	5.27 (4.87, 5.66)	5.44	< .001	4.47 (4.08, 4.86)	5.36 (4.97, 5.76)	6.10	< .001
	Predictability	5.24 (4.94, 5.54)	5.78 (5.48, 6.07)	3.99	< .001	5.24 (4.94, 5.54)	4.45 (4.15, 4.75)	5.86	< .001
	Typicality	4.80 (4.43, 5.16)	6.26 (5.89, 6.63)	9.32	< .001	4.80 (4.43, 5.16)	3.89 (3.52, 4.26)	5.75	< .001

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.

4C. Traditional Repeated-Measures Analyses

Moral Goodness (No Choice vs. Choice). A 2 (Relationship) x 2 (Choice) repeated-measures ANOVA revealed a main effect of Relationship, $F(1, 304) = 8.56$, $p = .027$, $\eta_p^2 = .03$, $\eta_G^2 = .006$, such that agents who helped kin were judged as more morally good than agents who helped a stranger; a main effect of Choice, $F(1, 304) = 370.23$, $p < .001$, $\eta_p^2 = .55$, $\eta_G^2 = .226$, 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, $F(1, 304) = 21.41$, $p < .001$, $\eta_p^2 = .07$, $\eta_G^2 = .012$. No conditions in this 2x2 crossed below the mid-point, and so all analyses are presented in terms of "goodness," with higher scores meaning more morally 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 $d_z = 0.09$, Hedge's $g_{av} = 0.08$, 95% CI [-0.02, 0.18]. The

common language effect size (CLES) indicates that after controlling for individual differences, and taking into account the correlation between conditions, $r(304) = 0.609$, the likelihood that a participant judges an agent who helped a stranger as more moral than an agent who helped kin is 54%. In the Choice conditions, however, agents who helped kin instead of a stranger were judged as significantly more morally good ($M = 6.35$, $SD = 1.60$) than agents who helped a stranger instead of kin ($M = 5.81$, $SD = 1.69$), $t(304) = 4.18$, $p = < .001$, Cohen's $d_z = 0.24$, Hedge's $g_{av} = 0.33$ [0.17, 0.48]. The CLES indicates that after controlling for individual differences, and taking into account the correlation between conditions, $r(304) = 0.059$, the likelihood that a participant judges agents who helped kin instead of a stranger as more moral than agents who helped a stranger instead of kin is 59%.

Moral Goodness (Choice vs. Failed). A 2 (Relationship) x 2 (Choice) repeated-measures ANOVA revealed no main effect of Relationship, $F(1, 304) = 0.01$, $p = .938$, $\eta_p^2 = .00$, $\eta_G^2 = .000$; a main effect of Choice, $F(1, 304) = 287.29$, $p < .001$, $\eta_p^2 = .49$, $\eta_G^2 = .179$, such that agents were judged as more morally good when they helped rather than failed to help; and a significant Relationship x Choice interaction, $F(1, 304) = 42.79$, $p < .001$, $\eta_p^2 = .12$, $\eta_G^2 = .026$. Failed to Help judgments were below the mid-point indicating that agents who failed to help were judged as somewhat morally bad, with lower scores meaning more morally bad. 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 agents who failed to help kin ($M = 4.28$, $SD = 1.75$), $t(304) = 5.12$, $p < .001$, Cohen's $d_z = 0.29$, Hedge's $g_{av} = 0.32$ [0.20, 0.45]. The CLES indicates that after controlling for individual differences, and taking into account the correlation between conditions, $r(304) = 0.404$, the likelihood that a participant judges agents who failed to help a stranger as less morally bad than agents who failed to help kin is 62%. Agents who failed to help

kin were also judged as more morally bad (or less morally good; $M = 4.28$, $SD = 1.75$) than

agents who helped a stranger instead of kin ($M = 5.81$, $SD = 1.69$), $t(304) = 11.68$, $p = < .001$,

Cohen's $d_z = 0.67$, Hedge's $g_{av} = 0.89$ [0.72, 1.05]. The CLES indicates that after controlling for

individual differences, and taking into account the correlation between conditions, $r(304) =$

0.118, the likelihood that a participant judges agents who failed to help kin as more morally bad

than agents who helped a stranger instead of kin is 75%.

      Trustworthiness (No Choice vs. Choice). A 2 (Relationship) x 2 (Choice) repeated-

measures ANOVA showed a main effect of Relationship, $F(1, 304) = 29.66$, $p < .001$, $\eta_p^2 = .09$,

$\eta_G^2 = .019$, such that agents were judged as more trustworthy when they helped kin than when

they helped a stranger; a main effect of Choice, $F(1, 304) = 294.95$, $p < .001$, $\eta_p^2 = .49$, $\eta_G^2 =$

.175, 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, $F(1, 304) = 32.44$, $p < .001$, $\eta_p^2 = .10$, $\eta_G^2 = .020$. No conditions in this 2x2

crossed below the mid-point into untrustworthiness, and so all results are presented in terms of

"trustworthiness," with higher scores meaning more trustworthy. In the No Choice conditions,

agents who helped a stranger were judged as no more trustworthy ($M = 7.31$, $SD = 1.48$) than

agents who helped kin ($M = 7.30$, $SD = 1.51$), $t(304) = 0.25$, $p = .800$, Cohen's $d_z = 0.02$,

Hedge's $g_{av} = 0.01$ [-0.09, 0.12]. The CLES indicates that after controlling for individual

differences, and taking into account the correlation between conditions, $r(304) = 0.587$, the

likelihood that a participant judges agents who helped a stranger as more trustworthy than agents

who helped kin is 51%. In the Choice conditions, however, agents who helped kin instead of a

stranger were judged as significantly more trustworthy ($M = 6.28$, $SD = 1.65$) than agents who

helped a stranger instead of kin ($M = 5.38$, $SD = 1.76$), $t(304) = 6.35$, $p < .001$, Cohen's $d_z =$

0.36, Hedge's g_{av} = 0.53 [0.36, 0.70]. The CLES indicates that after controlling for individual differences, and taking into account the correlation between conditions, $r(304)$ = -0.054, the likelihood that a participant judges agents who helped kin instead of a stranger as more trustworthy than agents who helped a stranger instead of kin is 64%.

Trustworthiness (Choice vs. Failed). A 2 (Relationship) x 2 (Choice) repeated-measures ANOVA yielded no main effect of Relationship, $F(1, 304)$ = 1.34, p = .249, η_p^2 = .00, η_G^2 = .001; a main effect of Choice, $F(1, 304)$ = 168.65, p < .001, η_p^2 = .36, η_G^2 = .107, 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, η_p^2 = .19, η_G^2 = .054. Some Failed to Help judgments in this 2x2 crossed below the mid-point into "untrustworthiness," with lower numbers meaning more untrustworthy. 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 agents who failed to help kin (M = 4.31, SD = 1.78), $t(304)$ = 6.38, p < .001, Cohen's d_z = 0.37, Hedge's g_{av} = 0.43 [0.29, 0.56]. The CLES indicates that after controlling for individual differences, and taking into account the correlation between conditions, $r(304)$ = 0.323, the likelihood that a participant judges agents who failed to help a stranger as more trustworthy than agents who failed to help kin is 64%. Agents who failed to help kin were also judged as significantly less trustworthy (or more untrustworthy; M = 4.31, SD = 1.78) than agents who helped a stranger instead of kin (M = 5.38, SD = 1.76), $t(304)$ = 8.25, p = < .001, Cohen's d_z = 0.47, Hedge's g_{av} = 0.60 [0.45, 0.75]. The CLES indicates that after controlling for individual differences, and taking into account the correlation between conditions, $r(304)$ = 0.191, the likelihood that a participant judges agents who failed to help kin as less trustworthy than agents who helped a stranger instead of kin is 68%.

Obligation (No Choice vs. Choice). A 2 (Relationship) x 2 (Choice) repeated-measures ANOVA revealed a main effect of Relationship, $F(1, 304) = 66.21$, $p < .001$, $\eta_p^2 = .18$, $\eta_G^2 = .041$, such that agents were judged as fulfilling an obligation more for helping kin than for helping a stranger; a main effect of Choice, $F(1, 304) = 106.34$, $p < .001$, $\eta_p^2 = .26$, $\eta_G^2 = .073$, 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, $F(1, 304) = 18.65$, $p < .001$, $\eta_p^2 = .06$, $\eta_G^2 = .010$. No conditions in this 2x2 crossed below the mid-point into violating an obligation, and so all results are presented in terms of "fulfilling obligations," with higher scores meaning judgments closer to completely fulfilling an obligation the agent had. In the No Choice conditions, agents who helped a stranger were judged as fulfilling an obligation less ($M = 6.51$, $SD = 1.67$) than agents who helped kin ($M = 6.89$, $SD = 1.73$), $t(304) = 3.96$, $p < .001$, Cohen's $d_z = 0.23$, Hedge's $g_{av} = 0.22$ [0.11, 0.33]. The CLES indicates that after controlling for individual differences, and taking into account the correlation between conditions, $r(304) = 0.538$, the likelihood that a participant judges agents who helped a stranger as fulfilling an obligation less than agents who helped kin is 59%. In the Choice conditions, again, helping a stranger instead of helping kin was judged as fulfilling less of an obligation ($M = 5.20$, $SD = 1.82$) than helping kin instead of helping a stranger ($M = 6.27$, $SD = 1.71$), $t(304) = 7.57$, $p < .001$, Cohen's $d_z = 0.43$, Hedge's $g_{av} = 0.60$ [0.44, 0.77]. The CLES indicates that after controlling for individual differences, and taking into account the correlation between conditions, $r(304) = 0.023$, the likelihood that a participant judges agents who helped a stranger instead of kin as fulfilling an obligation less than agents who helped kin instead of a stranger is 67%.

Obligation (Choice vs. Failed). A 2 (Relationship) x 2 (Choice) repeated-measures

ANOVA demonstrated a main effect of Relationship, $F(1, 304) = 7.26$, $p = .007$, $\eta_p^2 = .02$, $\eta_G^2 =$

.005, such that judgments of fulfilling obligation were higher for interactions involving kin than

for interactions involving a stranger; a main effect of Choice, $F(1, 304) = 139.66$, $p < .001$, $\eta_p^2 =$

.32, $\eta_G^2 = .098$, such that helping was judged as fulfilling an obligation more than failing to help;

and a significant Relationship x Choice interaction, $F(1, 304) = 80.67$, $p < .001$, $\eta_p^2 = .21$, $\eta_G^2 =$

.054. Failure to help judgments were below the mid-point indicating that failing to help was

judged as *violating* an obligation, with lower scores meaning more of a violation of an

obligation. In the Failed to Help conditions, agents who failed to help a stranger were judged as

violating an obligation significantly less ($M = 4.89$, $SD = 1.54$) than agents who failed to help kin

($M = 4.30$, $SD = 1.82$), $t(304) = 5.13$, $p < .001$, Cohen's $d_z = 0.29$, Hedge's $g_{av} = 0.34$ [0.21,

0.48]. The CLES indicates that after controlling for individual differences, and taking into

account the correlation between conditions, $r(304) = 0.307$, the likelihood that a participant

judges agents who failed to help a stranger as violating an obligation less than agents who failed

to help kin is 62%. Agents who failed to help kin were also judged as violating an obligation

significantly more (or fulfilling an obligation significantly less; $M = 4.30$, $SD = 1.82$) than agents

who helped a stranger instead of kin ($M = 5.20$, $SD = 1.82$), $t(304) = 6.78$, $p = < .001$, Cohen's d_z

$= 0.39$, Hedge's $g_{av} = 0.49$ [0.35, 0.64]. The CLES indicates that after controlling for individual

differences, and taking into account the correlation between conditions, $r(304) = 0.193$, the

likelihood that a participant judges agents who failed to help a stranger as violating an obligation

more than agents who helped a stranger instead of kin is 65%.

Rightness (No Choice vs. Choice). A 2 (Relationship) x 2 (Choice) repeated-measures

ANOVA yielded a main effect of Relationship, $F(1, 304) = 23.89$, $p < .001$, $\eta_p^2 = .07$, $\eta_G^2 = .016$,

such that participants judged helping kin as righter than helping a stranger; a main effect of

Choice, $F(1, 304) = 342.21$, $p < .001$, $\eta_p^2 = .53$, $\eta_G^2 = .225$, such that help was judged as righter

when no other potential beneficiary was available to help than when there were two potential

beneficiaries; and a significant Relationship x Choice interaction, $F(1, 304) = 35.67$, $p < .001$, η_p^2

$= .11$, $\eta_G^2 = .020$. No conditions in this 2x2 crossed below the mid-point into wrongness, and so

all results are presented in terms of "rightness," with higher scores meaning righter. In the No

Choice conditions, helping a stranger was judged as no righter ($M = 7.72$, $SD = 1.46$) than

helping kin ($M = 7.67$, $SD = 1.55$), $t(304) = 0.61$, $p = .542$, Cohen's $d_z = 0.04$, Hedge's $g_{av} = 0.03$

[-0.07, 0.15]. The CLES indicates that after controlling for individual differences, and taking into

account the correlation between conditions, $r(304) = 0.505$, the likelihood that a participant

judges helping a stranger as righter than helping kin is 51%. In the Choice conditions, however,

helping kin instead of a stranger ($M = 6.34$, $SD = 1.79$) was judged as significantly righter than

helping a stranger instead of kin ($M = 5.44$, $SD = 1.90$), $t(304) = 6.29$, $p = < .001$, Cohen's $d_z =$

0.36, Hedge's $g_{av} = 0.49$ [0.33, 0.65]. The CLES indicates that after controlling for individual

differences, and taking into account the correlation between conditions, $r(304) = 0.071$, the

likelihood that a participant judges helping kin instead of a stranger as righter than helping a

stranger instead of kin is 64%.

   Rightness (Choice vs. Failed). A 2 (Relationship) x 2 (Choice) repeated-measures

ANOVA demonstrated no main effect of Relationship, $F(1, 304) = 0.31$, $p = .578$, $\eta_p^2 = .00$, η_G^2

$= .000$; a main effect of Choice, $F(1, 304) = 177.60$, $p < .001$, $\eta_p^2 = .37$, $\eta_G^2 = .119$, such that

helping was judged as righter than failing to help; and a significant Relationship x Choice

interaction, $F(1, 304) = 78.89$, $p < .001$, $\eta_p^2 = .21$, $\eta_G^2 = .050$. Failed to Help judgments were

below the mid-point indicating that failing to help was perceived as somewhat wrong, with lower

scores meaning more wrong. 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_z = 0.40, Hedge's g_{av} = 0.43 [0.30, 0.55]. The CLES indicates that after controlling for individual differences, and taking into account the correlation between conditions, $r(304)$ = 0.426, the likelihood that a participant judges failing to help a stranger as less wrong than failing to help kin is 65%. Failing to help kin was also judged as more wrong (or less right; 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_z = 0.54, Hedge's g_{av} = 0.69 [0.53, 0.84]. The CLES indicates that after controlling for individual differences, and taking into account the correlation between conditions, $r(304)$ = 0.202, the likelihood that a participant judges failing to help kin as more wrong than helping a stranger instead of kin is 71%.

Praise (No Choice vs. Choice). A 2 (Relationship) x 2 (Choice) repeated-measures ANOVA yielded a main effect of Relationship, $F(1, 304)$ = 6.31, $p = .013$, $\eta_p^2 = .02$, $\eta_G^2 = .004$, such that agents who helped kin were praised more than agents who helped a stranger; a main effect of Choice, $F(1, 304)$ = 419.90, $p < .001$, $\eta_p^2 = .58$, $\eta_G^2 = .253$, such that agents were praised more when no other potential beneficiary was available to help than when there were two potential beneficiaries; and a significant Relationship x Choice interaction, $F(1, 304)$ = 18.26, $p < .001$, $\eta_p^2 = .06$, $\eta_G^2 = .011$. No conditions in this 2x2 crossed the below the mid-point into blame, and so all results are presented in terms of "praise," with higher scores meaning more praise. In the No Choice conditions, agents who helped a stranger were praised marginally more (M = 7.84, SD = 1.38) than agents who helped kin (M = 7.70, SD = 1.52), $t(304)$ = 1.67, $p = .095$, Cohen's d_z = 0.10, Hedge's g_{av} = 0.09 [-0.02, 0.21]. The CLES indicates that after controlling for individual differences, and taking into account the correlation between conditions, $r(304)$ =

0.513, the likelihood that a participant praises an agent who helped a stranger more than an agent

who helped kin is 54%. In the Choice conditions, however, agents who helped kin instead of a

stranger were praised significantly more (M = 6.14, SD = 1.74) than agents who helped a

stranger instead of kin (M = 5.59, SD = 1.85), $t(304)$ = 3.93, $p < .001$, Cohen's d_z = 0.23,

Hedge's g_{av} = 0.30 [0.15, 0.46]. The CLES indicates that after controlling for individual

differences, and taking into account the correlation between conditions, $r(304)$ = 0.088, the

likelihood that a participant praises an agent who helped kin instead of a stranger more than an

agent who helped a stranger instead of kin is 59%.

Praise (Choice vs. Failed). A 2 (Relationship) x 2 (Choice) repeated-measures ANOVA

revealed no main effect of Relationship, $F(1, 304)$ = 0.02, p = .878, η_p^2 = .00, η_G^2 = .000; a main

effect of Choice, $F(1, 304)$ = 278.43, $p < .001$, η_p^2 = .48, η_G^2 = .170, 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, $p < .001$, η_p^2 = .11, η_G^2 = .023. Failed to Help judgments were

below the mid-point indicating that agents who failed to help were somewhat blamed, with lower

scores meaning more blame. Agents who failed to help a stranger were blamed significantly less

(M = 4.56, SD = 1.47) than agents who failed to help kin (M = 4.04, SD = 1.94), $t(304)$ = 4.99, p

< .001, Cohen's d_z = 0.29, Hedge's g_{av} = 0.31 [0.19, 0.44]. The CLES indicates that after

controlling for individual differences, and taking into account the correlation between conditions,

$r(304)$ = 0.398, the likelihood that a participant blames an agent who failed to help a stranger less

than an agent who failed to help kin is 61%. Agents who failed to help kin were also blamed

more (M = 4.04, SD = 1.94) than agents who helped a stranger instead of kin (M = 5.59, SD =

1.85), $t(304)$ = 11.45, $p = < .001$, Cohen's d_z = 0.66, Hedge's g_{av} = 0.84 [0.68, 1.00]. The CLES

indicates that after controlling for individual differences, and taking into account the correlation

between conditions, $r(304) = 0.169$, the likelihood that a participant blames an agent who failed

to help kin more than an agent who helped a stranger instead of kin is 74%.

Fairness (No Choice vs. Choice). A 2 (Relationship) x 2 (Choice) repeated-measures

ANOVA yielded a main effect of Relationship, $F(1, 304) = 38.03$, $p < .001$, $\eta_p^2 = .11$, $\eta_G^2 = .021$,

such that when agents helped kin, it was judged as fairer than when agents helped a stranger; a

main effect of Choice, $F(1, 304) = 249.19$, $p < .001$, $\eta_p^2 = .45$, $\eta_G^2 = .178$, such that help 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, $p < .001$, $\eta_p^2 = .12$, $\eta_G^2 = .022$. No conditions in this 2x2 crossed below the mid-point into

unfair, and so all results are presented in terms of "fairness," with higher scores meaning fairer.

In the No Choice 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 $d_z = 0.00$, Hedge's $g_{av} = 0.00$ [-0.10, 0.10]. The CLES indicates that after

controlling for individual differences, and taking into account the correlation between conditions,

$r(304) = 0.601$, the likelihood that a participant judges helping a stranger as fairer than helping

kin is 50%. 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 $d_z = 0.40$, Hedge's $g_{av} = 0.53$ [0.38, 0.68]. The CLES

indicates that after controlling for individual differences, and taking into account the correlation

between conditions, $r(304) = 0.133$, the likelihood that a participant judges helping kin instead of

a stranger as fairer than helping a stranger instead of kin is 66%.

Fairness (Choice vs. Failed). A 2 (Relationship) x 2 (Choice) repeated-measures

ANOVA resulted in no main effect of Relationship, $F(1, 304) = 1.12$, $p = .292$, $\eta_p^2 = .00$, $\eta_G^2 =$

.001; a main effect of Choice, $F(1, 304) = 77.91$, $p < .001$, $\eta_p^2 = .20$, $\eta_G^2 = .058$, such that it was judged as fairer when agents helped rather than failed to help; and a significant Relationship x Choice interaction, $F(1, 304) = 83.85$, $p < .001$, $\eta_p^2 = .22$, $\eta_G^2 = .050$. Some Failed to Help judgments were below the mid-point indicating that, depending on relationship, failing to help was judged as somewhat unfair, with lower scores meaning more unfair. 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 $d_z = 0.35$, Hedge's $g_{av} = 0.39$ [0.26, 0.52]. The CLES indicates that after controlling for individual differences, and taking into account the correlation between conditions, $r(304) = 0.379$, the likelihood that a participant judges failing to help a stranger as fairer than failing to help kin is 64%. Failing to help kin was also judged as significantly more unfair (or significantly less fair; $M = 4.46$, $SD = 2.17$) than helping a stranger instead of kin ($M = 5.37$, $SD = 1.98$), $t(304) = 5.85$, $p = < .001$, Cohen's $d_z = 0.34$, Hedge's $g_{av} = 0.44$ [0.29, 0.59]. The CLES indicates that after controlling for individual differences, and taking into account the correlation between conditions, $r(304) = 0.141$, the likelihood that a participant judges failing to help kin as more unfair than helping a stranger instead of kin is 63%.

Predictability (No Choice vs. Choice). A 2 (Relationship) x 2 (Choice) repeated-measures ANOVA demonstrated a main effect of Relationship, $F(1, 304) = 96.06$, $p < .001$, $\eta_p^2 = .24$, $\eta_G^2 = .062$, such that, for agents who helped a stranger, their future behavior was judged as less predictable than the future behavior of agents who helped kin; a main effect of Choice, $F(1, 304) = 54.19$, $p < .001$, $\eta_p^2 = .15$, $\eta_G^2 = .035$, 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, $F(1, 304) = 51.80$, $p < .001$, $\eta_p^2 = .15$, $\eta_G^2 = .023$. Some judgments in this 2x2 fell below the mid-point of "neither

unpredictable or predictable," with lower scores (below the mid-point) meaning more

unpredictable. In the No Choice 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

($M = 6.17$, $SD = 1.85$), $t(304) = 3.26$, $p = .001$, Cohen's $d_z = 0.19$, Hedge's $g_{av} = 0.19$ [0.07,

0.30]. The CLES indicates that after controlling for individual differences, and taking into

account the correlation between conditions, $r(304) = 0.490$, the likelihood that a participant

judges agents who helped a stranger as less predictable than agents who helped kin is 57%. In the

Choice conditions, again, when agents helped a stranger instead of kin, their future behavior was

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, $p < .001$, Cohen's $d_z = 0.61$, Hedge's $g_{av} = 0.83$ [0.67, 1.00]. The CLES indicates that

after controlling for individual differences, and taking into account the correlation between

conditions, $r(304) = 0.068$, the likelihood that a participant judges agents who helped a stranger

instead of kin as less predictable than agents who helped kin instead of a stranger is 73%.

Predictability (Choice vs. Failed). A 2 (Relationship) x 2 (Choice) repeated-measures

ANOVA yielded a main effect of Relationship, $F(1, 304) = 31.02$, $p < .001$, $\eta_p^2 = .09$, $\eta_G^2 = .019$,

such that when agents interacted with kin, their future behavior was judged as more predictable

than when agents interacted with a stranger; a main effect of Choice, $F(1, 304) = 6.32$, $p = .012$,

$\eta_p^2 = .02$, $\eta_G^2 = .004$, such that agents who helped were judged to be more predictable than

agents who failed to help; and a significant Relationship x Choice interaction, $F(1, 304) =$

127.50, $p < .001$, $\eta_p^2 = .30$, $\eta_G^2 = .078$. For agents who failed to help a stranger, their future

behavior was judged as significantly more predictable ($M = 5.78$, $SD = 1.75$) than the future

behavior of agents who failed to help kin ($M = 5.22$, $SD = 1.88$), $t(304) = 4.92$, $p < .001$, Cohen's

$d_z = 0.28$, Hedge's $g_{av} = 0.31$ [0.19, 0.44]. The CLES indicates that after controlling for

individual differences, and taking into account the correlation between conditions, $r(304) =$

0.387, the likelihood that a participant judges agents who failed to help a stranger as more

predictable than agents who failed to help kin is 61%. Agents who failed to help kin were also

judged as significantly more predictable ($M = 5.22$, $SD = 1.88$) than agents who helped a stranger

instead of kin ($M = 4.45$, $SD = 2.09$), $t(304) = 5.47$, $p = < .001$, Cohen's $d_z = 0.31$, Hedge's $g_{av} =$

0.39 [0.24, 0.53]. The CLES indicates that after controlling for individual differences, and taking

into account the correlation between conditions, $r(304) = 0.237$, the likelihood that a participant

judges agents who failed to help kin as more predictable than agents who helped a stranger

instead of kin is 62%.

     Typicality (No Choice vs. Choice). A 2 (Relationship) x 2 (Choice) repeated-measures

ANOVA resulted in a main effect of Relationship, $F(1, 304) = 246.70$, $p < .001$, $\eta_p^2 = .49$, $\eta_G^2 =$

.146, such that helping kin was judged to be more likely than helping a stranger; a marginal main

effect of Choice, $F(1, 304) = 3.68$, $p = .056$, $\eta_p^2 = .01$, $\eta_G^2 = .003$, such that helping someone

when no other potential beneficiary was available to help was judged as marginally more likely

than helping one agent instead of another; and a significant Relationship x Choice interaction,

$F(1, 304) = 32.15$, $p < .001$, $\eta_p^2 = .10$, $\eta_G^2 = .019$. Some judgments in this 2x2 fell below the

mid-point, indicating that participants believed some behaviors were somewhat unlikely, with

lower scores (below the mid-point) meaning more unlikely. In the No Choice 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 $d_z = 0.48$, Hedge's $g_{av} = 0.53$

[0.40, 0.67]. The CLES indicates that after controlling for individual differences, and taking into

account the correlation between conditions, $r(304) = 0.367$, the likelihood that a participant

judges helping a stranger as less likely than helping kin is 68%. 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 $d_z = 0.82$, Hedge's $g_{av} = 1.13$ [0.95, 1.31]. The CLES indicates that after controlling for individual differences, and taking into account the correlation between conditions, $r(304) = 0.037$, the likelihood that a participant judges helping a stranger instead of kin as less likely than helping kin instead of a stranger is 79%.

Typicality (Choice vs. Failed). A 2 (Relationship) x 2 (Choice) repeated-measures ANOVA showed a main effect of Relationship, $F(1, 304) = 14.12$, $p < .001$, $\eta_p^2 = .04$, $\eta_G^2 =$ .010, such judgments of likelihood were higher when agents interacted with kin than when agents interacted with a stranger; a main effect of Choice, $F(1, 304) = 16.77$, $p < .001$, $\eta_p^2 = .05$, $\eta_G^2 = .014$, such that helping was judged as *less* likely than failing to help; and a significant Relationship x Choice interaction, $F(1, 304) = 295.33$, $p < .001$, $\eta_p^2 = .49$, $\eta_G^2 = .179$. Some judgments of failing to help were again below the mid-point. Failing to help a stranger was judged as significantly 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 $d_z = 0.59$, Hedge's $g_{av} = 0.73$ [0.58, 0.89]. The CLES indicates that after controlling for individual differences, and taking into account the correlation between conditions, $r(304) = 0.214$, the likelihood that a participant judges failing to help a stranger as more likely than failing to help kin is 72%. Failing to help kin was also judged as significantly 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 $d_z = 0.33$, Hedge's $g_{av} = 0.42$ [0.27, 0.56]. The CLES indicates that after controlling for individual differences, and taking into

account the correlation between conditions, $r(304) = 0.190$, the likelihood that a participant

judges failing to help kin as more likely than helping a stranger instead of kin is 63%.

4D. Within-Condition Hierarchical Regressions

As stated in the main text, to rule out the possibility that moral judgments were solely the

result of perceived behavior frequency and future predictability (and to provide evidence that

moral judgments were in fact due to perceptions of obligations being violated versus fulfilled),

we conducted hierarchical regressions within each individual condition. In the first models,

participants' typicality and future predictability judgments were entered as predictors, and in the

second models, obligation judgments were entered as the predictor of interest. This allowed us to

test, within individual conditions, if obligation judgments accounted for unique variance in moral

judgments. Within all conditions, obligation judgments were always significant (and always the

strongest) predictors 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.

	Predictor	Model 1					Model 2				
		b (SE)	β	sp-r	t	p	b (SE)	β	sp-r	t	p
SNC	Predictability	.10 (.04)	.15	.13	2.35	.019	.02 (.04)	.03	.03	0.56	.574
	Typicality	-.06 (.04)	-.11	-.10	-.1.67	.096	-.09 (.04)	-.15	-.13	-2.52	.012
	Obligation	-	-	-	-	-	.33 (.04)	.43	.41	7.81	< .001
		F(2, 302) =	**2.99**		**R² = .02**	.052	**F(3, 301) =**	**22.71**		**R² = .19**	< .001
										ΔR² = .17	< .001
KNC	Predictability	.18 (.04)	.26	.24	4.38	< .001	.11 (.04)	.15	.14	2.78	.006
	Typicality	.00 (.04)	.01	.01	0.10	.917	-.04 (.03)	-.06	-.06	-1.12	.262
	Obligation	-	-	-	-	-	.35 (.04)	.47	.45	9.02	< .001
		F(2, 302) =	**11.53**		**R² = .07**	< .001	**F(3, 301) =**	**36.84**		**R² = .27**	< .001
										ΔR² = .20	< .001
SOK	Predictability	.11 (.05)	.13	.11	2.11	.035	.01 (.04)	.01	.01	0.11	.912
	Typicality	.28 (.05)	.35	.29	5.70	< .001	.08 (.05)	.10	.08	1.76	.080
	Obligation	-	-	-	-	-	.54 (.05)	.58	.47	10.60	< .001
		F(2, 302) =	**36.30**		**R² = .19**	< .001	**F(3, 301) =**	**70.55**		**R² = .41**	< .001
										ΔR² = .22	< .001
KOS	Predictability	.39 (.06)	.43	.32	6.60	< .001	.16 (.06)	.17	.12	2.82	.005
	Typicality	.11 (.05)	.13	.10	1.99	.048	.05 (.05)	.06	.04	1.05	.293
	Obligation	-	-	-	-	-	.50 (.05)	.54	.44	10.45	< .001
		F(2, 302) =	**57.79**		**R² = .28**	< .001	**F(3, 301) =**	**88.72**		**R² = .47**	< .001
										ΔR² = .19	< .001
SFH	Predictability	.09 (.05)	.10	.09	1.76	.079	.04 (.04)	.05	.04	1.01	.312
	Typicality	.28 (.05)	.37	.33	6.28	< .001	.13 (.04)	.17	.15	3.58	< .001
	Obligation	-	-	-	-	-	.61 (.04)	.62	.57	14.17	< .001
		F(2, 302) =	**32.88**		**R² = .18**	< .001	**F(3, 301) =**	**103.38**		**R² = .51**	< .001
										ΔR² = .33	< .001
KFH	Predictability	.11 (.05)	.12	.11	2.24	.026	.05 (.04)	.05	.05	1.37	.171
	Typicality	.43 (.04)	.52	.45	9.66	< .001	.16 (.04)	.20	.16	4.65	< .001
	Obligation	-	-	-	-	-	.64 (.04)	.67	.57	16.85	< .001
		F(2, 302) =	**77.06**		**R² = .34**	< .001	**F(3, 301) =**	**194.07**		**R² = .66**	< .001
										ΔR² = .32	< .001

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.

	Predictor	Model 1					Model 2				
		b (SE)	β	sp-r	t	p	b (SE)	β	sp-r	t	p
SNC	Predictability	.17 (.05)	.23	.21	3.68	<.001	.08 (.04)	.11	.09	1.82	.069
	Typicality	-.03 (.04)	-.04	-.03	-0.57	.569	-.06 (.04)	-.08	-.07	-1.42	.158
	Obligation	-	-	-	-	-	.41 (.05)	.46	.44	8.74	<.001
		F(2, 302) =	7.51		R² = .05	.001	F(3, 301) =	31.74		R² = .24	<.001
										ΔR² = .19	<.001
KNC	Predictability	.28 (.05)	.35	.32	6.02	<.001	.20 (.04)	.24	.22	4.57	<.001
	Typicality	.05 (.04)	.07	.06	1.13	.261	.00 (.04)	.00	.00	-0.01	.991
	Obligation	-	-	-	-	-	.40 (.04)	.46	.44	9.26	<.001
		F(2, 302) =	25.31		R² = .14	<.001	F(3, 301) =	50.21		R² = .33	<.001
										ΔR² = .19	<.001
SOK	Predictability	.24 (.05)	.29	.24	5.16	<.001	.14 (.04)	.17	.14	3.55	<.001
	Typicality	.31 (.05)	.38	.31	6.72	<.001	.11 (.04)	.14	.11	2.68	.008
	Obligation	-	-	-	-	-	.53 (.05)	.55	.44	11.34	<.001
		F(2, 302) =	79.52		R² = .35	<.001	F(3, 301) =	118.24		R² = .54	<.001
										ΔR² = .20	<.001
KOS	Predictability	.41 (.06)	.44	.33	6.69	<.001	.19 (.06)	.21	.14	3.25	.001
	Typicality	.09 (.06)	.10	.08	1.57	.118	.03 (.05)	.04	.03	0.68	.497
	Obligation	-	-	-	-	-	.47 (.05)	.48	.40	9.01	<.001
		F(2, 302) =	54.40		R² = .27	<.001	F(3, 301) =	72.97		R² = .42	<.001
										ΔR² = .16	<.001
SFH	Predictability	.12 (.05)	.13	.12	2.33	.021	.08 (.04)	.09	.08	1.80	.073
	Typicality	.28 (.04)	.36	.32	6.26	<.001	.17 (.04)	.22	.18	4.05	<.001
	Obligation	-	-	-	-	-	.45 (.05)	.46	.43	9.34	<.001
		F(2, 302) =	36.11		R² = .19	<.001	F(3, 301) =	60.00		R² = .37	<.001
										ΔR² = .18	<.001
KFH	Predictability	.07 (.05)	.07	.06	1.24	.217	.00 (.04)	.00	.00	0.03	.980
	Typicality	.42 (.05)	.50	.44	9.06	<.001	.15 (.04)	.18	.14	3.98	<.001
	Obligation	-	-	-	-	-	.66 (.04)	.68	.58	16.16	<.001
		F(2, 302) =	61.05		R² = .29	<.001	F(3, 301) =	162.76		R² = .62	<.001
										ΔR² = .33	<.001

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.

4E. Exploratory Analyses for Obligation Principles

After participants had completed the main experimental phase in Study 4, they were then asked to read and rate their agreement with four general principles related to obligations to strangers and family: (1) If people have obligations to help family members, (2) if people do not have obligations to help 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.

Variable	Mean	SD	1	2	3	4
1. Family Obligation	6.24	2.06	-			
2. No Stranger Obligation	5.79	2.24	*-.13	-		
3. No Family Praise	4.10	2.17	-.01	.11	-	
4. Stranger Praise	6.93	1.82	**.18	.00	***-.40	-

Note. *p < .05; **p < .01; ***p < .001. *N* = 305.

4F. Dependent Variable and Obligation Principles 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")

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")

Additional References

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A

practical primer for *t*-tests and ANOVAs. *Frontiers in Psychology, 4:863.*

McGraw, K.O., & Wong, S.P. (1992). A common language effect size statistic. *Psychological*

Bulletin, 111(2), 361-365.