



“But he’s my brother”: The impact of family obligation on moral judgments and decisions

Junho Lee¹ · Keith J. Holyoak¹

Published online: 23 July 2019
© The Psychonomic Society, Inc. 2019

Abstract

We created practical moral dilemmas for which participants imagined witnessing a transgression by a target person. The identity of the transgressor was manipulated to be either a stranger or the participant's brother. In Experiment 1, whether the target person committed a violation was left ambiguous. Participants made factual (how strongly they believe the target person actually committed a transgression) and unethicity judgments regarding the incident, and rated their willingness to report the transgressor to the police. Given ambiguity (Experiment 1), participants interpreted the facts in favor of their brother, but not in favor of a stranger. This interpretation led to moral judgments and willingness to report that favored family over strangers, while creating overall coherence in reasoning. In Experiment 2, we eliminated the ambiguity of the factual situation so that the possibility of achieving coherence between unethicity of an act and leniency toward a family member was blocked. Nonetheless, participants were less willing to report their brother to the police. Experiment 3 replicated the findings of the first two experiments within an integrated study design. Results from path analyses indicated that the factual judgment depended on factual understanding of an event, but willingness to report depended on identity of the target (i.e., brother vs. stranger), even at the cost of reduced coherence in reasoning. Moral decisions are thus strongly influenced by agent-relative obligations, such as duty to protect a family member.

Keywords Morality · Judgment · Decision-making · Family obligation · Favoritism · Motivated reasoning

Moral reasoning is often viewed as being based on an individual's assessment of his or her responsibilities toward strangers or near strangers within some broadly defined group (e.g., fellow citizens). Particularly when moral judgments are considered within the ethical framework of utilitarianism (e.g., Singer, 1979), the value of each affected person is held independent of the unique perspective of the individual decision maker. Under this view, moral judgments are *agent neutral* (Nagel, 1986). The majority of studies in moral psychology (notably, those focusing on sacrificial dilemmas based on the trolley problem; Foot, 1967) involve scenarios in which a hypothetical decision is made concerning the fates of anonymous strangers (e.g., Cikara, Farnsworth, Harris, & Fiske, 2010; Greene, Sommerville, Nystrom, Darley, & Cohen, 2001). Any hint of a personal relationship between the reasoner and those affected by the moral judgment is avoided in the

interests of supposed objectivity, consistent with experimental controls employed in psychological research.

This focus in moral psychology on agent-neutral judgments has contributed to the neglect of moral issues that arise in everyday life. Few of us will ever have to choose whether to redirect a runaway trolley so as to kill one stranger in order to save five others; many of us will have to choose whether to support our own children or donate all our income to charity. Of course, for most people the latter decision is not likely to be a difficult one—our own children need to be fed first. One case in which a parent exhibited extreme utilitarianism—giving his money to strangers rather than family members—was the result of brain damage (Ferreira-Garcia, Fontenelle, Moll, & de Oliveira-Souza, 2014; see also Kahane et al., 2012; Kahane et al., 2018).

Introducing a personal relationship between the moral agent and the affected parties emphasizes *agent-relative* responsibilities (Nagel, 1986). These include family obligations, patriotism, or in-group loyalty, concerns of the sort that Baron and Spranca (1997) characterized as “protected values” that resist trade-offs with other values. Agent-relative decisions inevitably evoke the deontological concepts of rights and duties (Holyoak & Powell, 2016) that cannot be universal

✉ Junho Lee
leejunho@ucla.edu

¹ Department of Psychology, University of California, Los Angeles, Los Angeles, CA 90095, USA

given the potential conflicts between agents upholding the interests of different parties. For example, the duties of soldiers in two opposing nations at war generally cannot be reconciled, since their individual duties to be loyal to their respective countries conflict with each other.

As Nagel (1979) pointed out, agent-relative moral decisions are inevitably personal and situated. One's moral obligation to provide care and protection for a child is not equal between the case where that child is one's own daughter and the case where that child is anonymous and unrelated. Family obligations are essentially agent-relative, in that the ethicality of a decision depends on who is making the decision, and on the relationship between the moral agent and the affected parties.

Similarly, Bloom (2011) observed that those about whom we make moral judgments and decisions are often friends, family, or in-group members who share some personal attributes with us (e.g., students from the same school or people of the same ethnicity or nation; Cikara et al., 2010; Haidt, 2007; Haidt & Graham, 2007). Consequently, our moral judgments may often be partial and biased, especially when the safety, well-being, or interest of a close group conflicts with those of a more distant group. People's general tendency to favor in-group members, and more specifically family and kin, has been documented in experimental studies (Burnstein, Crandall, & Kitayama, 1994; Cikara et al., 2010; Haidt, 2007) and evolutionary analyses (Hamilton, 1964; Shackelford & Hansen, 2016). For example, Petrinovich, O'Neill, and Jorgensen (1993) showed that people's moral decisions in trolley problems and lifeboat problems (selecting who to throw out of a sinking ship to save other passengers) are significantly influenced by the relatedness between the participant and the victim. These investigators argued that people's moral decisions are strongly oriented toward favoring themselves and their relatives in order to maximize inclusive fitness. Bonnefon, Shariff, and Rahwan (2016) obtained a similar pattern in surveys about the decisions of autonomous vehicles. Participants were asked whether it is moral to create an algorithm for such vehicles that sacrifices passengers in order to save more pedestrians. When one of the passengers was the participant's family member, people judged such an algorithm to be less moral. Some researchers have argued that this sort of tribalism or parochialism was central to the evolution of human morality because it served the function of promoting cooperation and group loyalty, thereby increasing group-level fitness in the primal world (Greene, 2014; Haidt, 2012; Petrinovich et al., 1993).

In this paper, we examine how perceived family obligations—a ubiquitous source of agent-relative responsibilities—affect moral decisions in realistic scenarios that evoke moral judgments. Our central aim is to understand the mechanisms by which family obligation may affect such judgments. One general possibility is that family relationships

may enter into an interconnected network of beliefs and attitudes that collectively shift so as to maximize coherence. For example, people faced with a legal decision involving contradictory and ambiguous factors will shift their views on all the relevant factors to maximize coherence with their eventual decision (Holyoak & Simon, 1999). Coherence shifts, which can be modeled in terms of constraint satisfaction within a belief network, have been shown to affect decisions ranging from consumer choice (Russo, Carlson, Meloy, & Yong, 2008; Simon, Krawczyk & Holyoak, 2004) to judgments of legal and moral culpability (Simon, Stenstrom, & Read, 2015; Simon, Snow, & Read, 2004).

In the case of an apparent transgression, family favoritism may alter assessments of uncertain aspects of the situation so as to make a family member appear less blameworthy than a stranger would be (an instance of motivated reasoning; Ditto, Pizarro, & Tannenbaum, 2009; Kunda, 1990). Such coherence shifts would be consistent with what Holyoak and Powell (2016) termed deontological coherence, whereby a network of beliefs and values is altered so as to minimize conflict between rival moral values. A number of studies have investigated this tendency to achieve coherence in reasoning. Uhlmann, Pizarro, Tannenbaum, and Ditto (2009) showed that people selectively adopt and endorse moral principles (consequentialist or deontological) so that the selected principle supports their desired moral conclusion. Liu and Ditto (2013) showed that inducing moral beliefs for or against capital punishment by making participants read essays led to significant change in factual beliefs about the consequences of the punishment, suggesting that the mutual impact of factual and moral beliefs can be bidirectional. Similarly, Kahan, Hoffman, Braman, Evans, and Rachlinski (2012) showed that people's worldview (e.g., egalitarianism, communitarianism) affected their perception of protestors' behaviors in a video. Depending on the perceived purpose of the protest, which was manipulated by the researchers (e.g., supporting abortion rights), participants interpreted the same behaviors of the police and the protestors differently, in ways that created coherence with their views.

Coherence-driven decision-making takes advantage of factual and moral/legal ambiguities, which make it possible to shift beliefs without blatantly contradicting points of certainty. Sometimes, however, key facts are incontrovertible—the culprit may be caught red-handed. In such situations, the moral agent may be placed in a true dilemma, being forced to choose whether to act in accord with the dictates of the law and society, or to honor a perceived obligation to protect a family member. We first conducted exploratory studies to examine the influence of family favoritism in situations where factual ambiguity was present (Experiment 1) or absent (Experiment 2). These findings were replicated in Experiment 3 using a factorial design in which ambiguity was one of the independent variables.

Experiment 1

Participants read a scenario in which they witnessed a hypothetical transgression. They were asked a set of questions including one that assessed their willingness to report the transgressor to the police. We manipulated the identity of the purported transgressor as either a stranger or the participant's brother. When the transgressor is one's brother, two plausible moral goals are pitted against one another: civic duty (to report a transgression) versus family obligation (to protect a family member, even from a social punishment). The scenarios were constructed to be factually ambiguous as to whether or not the purported transgressor actually committed a transgression. Participants could therefore potentially interpret the facts in a way that tended to be exculpatory for their brother, thus maintaining coherence while making a decision based in part on family favoritism (Holyoak & Simon, 1999; Simon, Snow, & Read, 2004; Simon et al., 2015).

Method

Participants

Amazon Mechanical Turk was used to recruit 169 participants (92 females, $M_{\text{age}} = 35.9$ years, $SD_{\text{age}} = 11.6$ years) residing in the United States. In this exploratory experiment, we considered the identity of transgressor and severity of transgression as the two main factors of interest and aimed to collect at least 40 participants for each of the four main conditions. We used TurkGate (Goldin & Darlow, 2013) for all of the three experiments reported in this paper to prevent the same participants from participating in more than one experiment. Ethnicity was self-reported: 77% were European/European American, 12% were Asian/Asian American, 7% were African American/Black, and 5% were Hispanic or of Spanish origin. Participants received \$0.70 as compensation for completing the experiment, which took a median of 3.2 minutes.

Design, materials, and procedure

Each participant read and made judgments about a single scenario in which a target person is implicated in a purported crime. The scenarios used one of two basic cover stories (traffic violation or street battery). The rest of the design was a 2 (identity of transgressor: stranger/brother) $\times$ 2 (severity of violation: misdemeanor/felony) factorial. About 20 participants were assigned to each of the eight smallest cells (including the variation in cover story). Collapsing across the scenarios, there were about 40 participants in each condition (stranger*misdemeanor = 42, brother*misdemeanor = 43, stranger*felony = 44, and brother*felony = 40). All factors were manipulated between subjects.

Participants were instructed to imagine themselves as the person in the given scenario and to answer the questions after careful consideration. The participants imagined being a witness to a possible transgression. (All scenarios are provided as supplemental online materials; see Author's Note.) For example, one scenario, involving a street battery incident (misdemeanor), was as follows:

One evening, you are walking home after a long work-day. A block ahead, you see a man wearing a red baseball cap, who seems to be arguing with another man. Soon, the two men disappear into an alley. As you walk up to where they were, you see the other man lying on the ground in an alley, covering his face and groaning, though his injury doesn't seem to be serious. The injured man is drunk, so he may not be reliable or truthful about what happened.

You consider the possibility that the man with the red cap may have attacked the drunken man and then ran away, and should be reported to the police. However, you also consider the possibility that the drunken man may have been trying to pick a fight with the man in the red cap. The man with the red cap may have tried to defend himself, or perhaps hit the drunken man accidentally while trying to run away to avoid a fight.

The purported transgressor (man in the red cap in the above scenario; car driver in the traffic violation scenario) was the target person for factual and moral judgments. The purported transgression was either a misdemeanor, as above (punched a drunken man in the face and then ran away; drove under the influence), or a felony (stabbed a drunken man and then ran away; hit-and-run). To leave the factual details unclear, we provided two possible explanations in each of the scenarios (e.g., the man in the red cap may have attacked the drunken man, or he may have tried to defend himself from the drunken man). Following the description of the scenario, the target person was described as either a stranger or the participant's brother, depending on the condition. If participants did not have a brother, they were told to imagine they had a brother about their own age.

The participants were told that they saw a policeman a few blocks back, and were asked questions in the following fixed order: (1) "Do you think your brother/stranger actually [committed a transgression?]" (different transgression inserted for each condition), using a 6-point scale (1 = *certainly not*, 2 = *probably not*, 3 = *guess not*, 4 = *guess so*, 5 = *probably*, 6 = *certainly*). (2) "Given what you believe happened, how unethical was the behavior of your brother/stranger?", using a 5-point scale (1 = *not problematic*, 2 = *slightly unethical*, 3 = *moderately unethical*, 4 = *seriously unethical*, 5 = *extremely unethical*). Because the target person's behavior could only be interpreted as neutral at best, the lowest rating for this question

was labeled as “not problematic” rather than “ethical.” (3) “Given what you believe happened, would you report your brother/stranger to the policeman?”, using a 6-point scale, same as (1). (4) “How difficult was it to make the decision to report (or not report) your brother to the policeman?”, using a 5-point scale (1 = *not difficult*, 2 = *slightly difficult*, 3 = *moderately difficult*, 4 = *seriously difficult*, 5 = *extremely difficult*). For brevity, the questions and the corresponding mean scores will be termed *factual*, *unethicality*, *report*, and *difficulty*, respectively.

After the main questions, participants were asked to briefly explain the reason why they reported (or did not report) the target person to the police (answers to this open-ended question were not formally analyzed). Finally, basic demographic questions were asked, including whether the participant had a sibling, and if they did, how close they were with their sibling, on a 5-point scale (1 = *very distant*, 2 = *distant*, 3 = *neither close nor distant*, 4 = *close*, 5 = *very close*). Although the hypothetical family member in our scenarios was always a brother, participants were not specifically asked if their siblings were male or female. We were interested in the potential effect of having a sibling per se, rather than the effect of specifically having a brother (vs. sister). It seemed possible that having an actual sibling might make the scenarios in the brother conditions more personally relatable. For participants who do not have siblings, it may be harder to imagine having a brother with whom they have a close bond, or to feel a responsibility to protect the hypothetical brother when he is vulnerable.

Results and discussion

We first conducted preliminary analyses to test the potential effects of scenarios, having a sibling, and closeness with siblings. Using 2 (identity of the transgressor: brother/stranger) $\times$ 2 (severity of transgression: misdemeanor/felony) $\times$ 2 (scenario type: traffic violation/street battery) between-subjects ANOVAs for each dependent measure, we first found that the mean scores for *factual*, *unethicality*, and *report* were all significantly higher in the traffic violation scenario than in the street battery scenario (significant main effects of scenario, $F_s > 4.42$, $ps < .038$, $\eta_p^2 < .034$). However, the effect of identity was essentially the same across the two scenarios. Accordingly, data were collapsed across scenarios.

About 87% of the participants reported having one or more siblings. Given the relatively small number of participants without a sibling, having a sibling could not be used as a factor in ANOVAs for dependent measures. However, performing the key ANOVAs for the brother condition ($n = 83$) after excluding participants without siblings yielded essentially the same pattern of results. For participants in the brother condition who had siblings ($n = 75$), we

checked whether closeness with siblings was correlated with the dependent measures. Only the *factual* score had a significant correlation with closeness ($r = -.26$, $p = .022$); *unethicality*, *report*, and *difficulty* did not ($|r|s < .08$, $ps > .50$). Hence, participants who reported being relatively closer to their actual siblings showed a slightly lower tendency to believe that a hypothetical brother committed a transgression in the scenarios. To maintain equal samples between cells in the full-factorial ANOVAs, we included all participants in the key analyses.

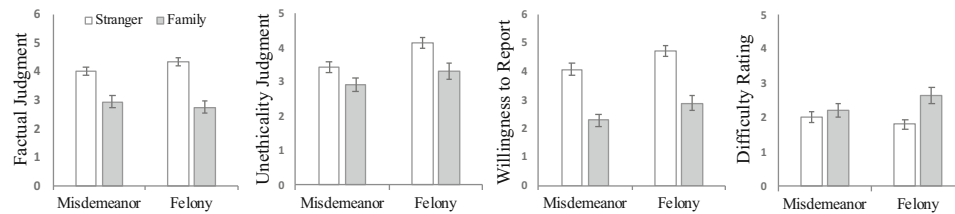
For the primary analyses, we conducted a 2 (identity) $\times$ 2 (severity) ANOVA for each of the dependent variables. The major findings are plotted in Fig. 1a. The identity of the target person yielded a significant main effect for *factual* score, $F(1, 165) = 55.17$, $p < .001$, $\eta_p^2 = .251$. Participants displayed a stronger belief that the target person committed a transgression when he was a stranger ($M = 4.19$, $SD = 0.98$) than when he was a brother ($M = 2.83$, $SD = 1.37$). Neither the main effect of severity nor the interactions between identity and severity were significant ($ps > .15$).

The ANOVA for *unethicality* indicated that both identity, $F(1, 165) = 12.18$, $p < .001$, $\eta_p^2 = .069$, and severity, $F(1, 165) = 8.69$, $p = .004$, $\eta_p^2 = .050$, yielded significant main effects. Participants judged the behavior as more unethical when the target person was a stranger ($M = 3.79$, $SD = 1.12$) than when he was their brother ($M = 3.11$, $SD = 1.41$), and when the transgression was a felony ($M = 3.75$, $SD = 1.36$) than when it was a misdemeanor ($M = 3.16$, $SD = 1.19$). The interaction of these variables was not significant ($p = .45$).

Similarly, for *report*, identity, $F(1, 165) = 64.49$, $p < .001$, $\eta_p^2 = .281$, and severity, $F(1, 165) = 7.61$, $p = .006$, $\eta_p^2 = .044$, yielded significant main effects. Participants were more willing to report the incident to the police when the target person was a stranger ($M = 4.41$, $SD = 1.38$) than when he was their brother ($M = 2.57$, $SD = 1.61$), and when the transgression was a felony ($M = 3.85$, $SD = 1.73$) than when it was a misdemeanor ($M = 3.16$, $SD = 1.72$). Again, the interaction was not significant ($p = .90$).

Finally, the ANOVA for *difficulty* yielded a significant main effect of identity, $F(1, 165) = 7.42$, $p = .007$, $\eta_p^2 = .043$. Difficulty in answering the *report* question was higher when the target person was their brother ($M = 2.42$, $SD = 1.48$) rather than a stranger ($M = 1.90$, $SD = 1.03$). Neither severity nor the interaction between identity and severity yielded significant differences ($ps > .10$). However, the trends suggest that the *report* question was slightly more difficult to answer in the felony condition than in the misdemeanor condition when the target person was a brother. When the target person was a stranger, difficulty tended to be slightly lower for felony than misdemeanor, presumably because a severe offense led the participants to report with less hesitation.

(a) Experiment 1



(b) Experiment 2

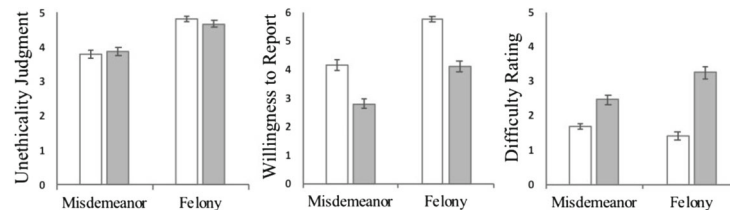


Fig. 1 Means of *factual*, *unethicality*, *report*, and *difficulty* scores for (a) Experiment 1, in which the scenarios provided factual ambiguity, and (b) Experiment 2, in which the scenarios were factually unambiguous. (A factual judgment was not elicited in the unambiguous case, Experiment

2.) In Experiment 1, family favoritism was observed for *factual*, *unethicality*, and *report* scores. In Experiment 2, family favoritism was not observed for *unethicality*, but nonetheless was observed for *report*. Error bars indicate standard error of the mean

Given the factual ambiguity in the scenarios, a coherence shift apparently took place: The identity of the transgressor significantly affected all of our dependent variables. The pattern of average scores in *factual*, *unethicality*, and *report* were coherent within each of the brother and stranger conditions, with all three scores significantly lower in the brother condition than in the stranger condition. Even though the observed behavior was identical, participants made more favorable judgments and decisions for their brother than for a stranger.

We then conducted correlation analyses separately for the brother and stranger conditions to further examine the relationships among the four dependent variables. To control for the main effect of severity, *unethicality* and *report* scores were standardized within each of the misdemeanor ($n = 85$) and felony ($n = 84$) conditions prior to these correlation analyses. As summarized in Table 1 (top), the correlations between

factual, *unethicality*, and *report* scores were all significant for both the brother and stranger conditions, and did not differ significantly between the two identity conditions (comparison after Fisher z transformation, $ps > .80$). Regardless of identity, when the target person was strongly believed to have committed a transgression, he was likely to be judged as unethical, and participants were more willing to report him to the police.

On the other hand, the correlations between *difficulty* and other dependent variables were significantly different between the brother and stranger conditions (all $ps < .018$). Strikingly, all correlations were positive for the brother condition (i.e., the more the brother was judged to have done wrong, the more difficult the decision), whereas all were negative for the stranger condition (i.e., the more the stranger was judged to have done wrong, the *easier* the decision). This pattern could be due to the participants experiencing greater difficulty when

Table 1 Correlations between dependent variables in Experiments 1 and 2

Experiment 1 ($n = 169$)						
Brother condition				Stranger condition		
	Unethicality	Report	Difficulty	Unethicality	Report	Difficulty
Factual	.55***	.49***	.20	.61***	.46***	-.17
Unethicality		.53***	.38***		.44***	-.14
Report			.36***			-.30**
Experiment 2 ($n = 327$)						
Brother condition				Stranger condition		
		Report	Difficulty		Report	Difficulty
Unethicality		.18*	.21**		.56***	-.43***
Report			-.02			-.44***

The variables significantly affected by severity of transgression (*unethicality* and *report* in Experiment 1 and all three variables in Experiment 2) were standardized within each of the misdemeanor and felony conditions and merged, thus controlling for the overall impact of type of transgression. * $p < .05$. ** $p < .01$. *** $p < .001$

their reluctance to report a family member conflicted with the judgment that he had committed a serious transgression. Conversely, in the stranger condition, fulfilling a stronger civic duty to blame and report the transgressor led to reduced difficulty.

Experiment 2

Experiment 2 used the same design and basic materials as Experiment 1, except that the scenarios precluded factual ambiguity about the transgression. When the facts are clear, it should be harder to shift unethicity judgments so as to maintain coherence with a motivation to protect one's brother. If family favoritism is nonetheless observed with respect to deciding whether to report the target to the police, this finding would imply that people are willing to protect a relative even at the cost of violating their apparent civic duty.

Method

Participants

We recruited 327 participants (158 females, $M_{\text{age}} = 36.9$ years, $SD_{\text{age}} = 12.6$ years) residing in the United States through Amazon Mechanical Turk. We expected to find main effects and interactions of three factors (identity, severity, and question order; see the Design section), and recruited about 40 participants for each of the eight conditions (stranger*misdemeanor*unethicity-first = 42, stranger*misdemeanor*report-first = 41, stranger*felony*unethicity-first = 39, stranger*felony*report-first = 40, brother*misdemeanor*unethicity-first = 42, brother*misdemeanor*report-first = 42, brother*felony*unethicity-first = 41, brother*felony*report-first = 40). Among the participants, 69% were European/European American, 13% were African American/Black, 9% were Asian/Asian American, 7% were Hispanic or Spanish origin, and 2% classified themselves as others. Participants received \$0.50 as compensation for the experiment, which took a median of 2.25 minutes.

Design, materials, and procedure

As in Experiment 1, the scenarios involved either a street battery or a traffic violation. The rest of the design of Experiment 2 was a 2 (identity of transgressor) $\times$ 2 (severity) $\times$ 2 (question order: unethicity-first/report-first) factorial. All factors were manipulated between participants, with about 20 participants in each of the smallest cells. The materials and general procedures were largely the same as

those from Experiment 1. However, factual ambiguity in the scenarios was eliminated, as participants were explicitly told that the target person indeed committed a transgression (see supplemental online materials). Thus, scenarios did not offer an alternative explanation of what might have caused the incident. Since factual ambiguity was eliminated, the *factual* question was omitted in Experiment 2.

One additional variable was manipulated in Experiment 2: the order of *unethicity* and *report* questions. In the *unethicity*-first condition, the *unethicity* question was asked before the *report* question (as in Experiment 1). In the *report*-first condition, the *report* question was asked first before the *unethicity* question. The identity of the target person was always revealed immediately before the participants answered the *report* question. This variation in question order was intended to provide an opportunity for family favoritism to impact *unethicity* judgments even when factual ambiguity was not present. Because the identity of the transgressor was not yet known when the *unethicity* judgment was administered first, it was impossible for family favoritism to operate. But when the *unethicity* question followed the *report* question, the identity of the transgressor (as well as the more lenient report judgment that may have been made in the brother condition) might guide participants to interpret the observed transgression as less unethical. This component of the design is analogous to classical studies of cognitive dissonance (e.g., Festinger, 1957) in which people's responses to later questions were dependent on their responses to earlier questions. If participants alter their subsequent *unethicity* judgment to cohere with their *report* decision, then the *unethicity* score might be reduced in the *report*-first condition when the target person was the brother. However, if factual ambiguity is a precondition for coherence shifts, then identity would not impact the *unethicity* score regardless of the order of questions.

After the main questions, participants were asked to provide the reason for their decision to report or not, as well as demographic questions including whether the participants had siblings.¹

Results and discussion

Preliminary analyses revealed that neither specific scenario nor having a sibling (89% of participants reported that they did) critically altered the way in which target identity affected scores on other questions. Accordingly, we collapsed across these two variables in all further analyses.

The pattern of scores for all questions is displayed in Fig. 1b. A three-way ANOVA including identity, severity, and the

¹ We did not ask about closeness with siblings in Experiment 2. However, results from Experiments 1 and 3 suggest that closeness ratings did not significantly predict the pattern of results.

order of the *report* and *unethicality* questions as independent variables was conducted for *unethicality* scores. Results indicated that only the main effect of severity was significant, $F(1, 319) = 78.85, p < .001, \eta_p^2 = .198$, with *unethicality* scores being higher for the felony ($M = 4.74, SD = 0.76$) than for the misdemeanor ($M = 3.83, SD = 1.06$) condition. No other main effects or interactions achieved significance ($ps > .135$). The identity of the target person did not influence *unethicality* score in Experiment 2 ($p = .78$). Moreover, there was no significant interaction between identity and question order ($p = .36$). Thus, even when the *report* question had been answered first, participants did not judge the target behavior to be less unethical for their brother than for a stranger.

For *report*, a three-way ANOVA indicated that the main effects of identity, $F(1, 319) = 73.29, p < .001, \eta_p^2 = .187$, and severity, $F(1, 319) = 78.05, p < .001, \eta_p^2 = .197$, were significant, while the other terms were not ($ps > .38$). Participants were more willing to report when the target person was a stranger ($M = 4.93, SD = 1.55$) than when he was their brother ($M = 3.44, SD = 1.81$), and when the transgression was a felony ($M = 4.91, SD = 1.60$) than when it was a misdemeanor ($M = 3.47, SD = 1.78$).

Finally, a three-way ANOVA for *difficulty* revealed a marginally significant main effect of severity, $F(1, 319) = 3.85, p = .051, \eta_p^2 = .012$, a significant main effect of identity, $F(1, 319) = 103.53, p < .001, \eta_p^2 = .245$, and a significant interaction between severity and identity, $F(1, 319) = 16.33, p < .001, \eta_p^2 = .049$. Test of simple effects showed that when the target person was a brother, participants experienced greater difficulty for a felony ($M = 3.23, SD = 1.48$) than a misdemeanor ($M = 2.46, SD = 1.27$), $t(163) = 4.25, p < .001, d = 0.66$. When the target person was a stranger, lower difficulty was reported for the felony scenario ($M = 1.41, SD = 0.94$) than misdemeanor scenario ($M = 1.67, SD = 0.84$), $t(160) = 1.47, p = .142$. This finding is consistent with the trend toward a comparable interaction observed in Experiment 1.

As in Experiment 1, correlations between scores for the major questions were calculated after standardizing all three variables within each of the misdemeanor and felony conditions to control for the main effect of severity (see Table 1, bottom). In contrast to the pattern observed in Experiment 1, the correlation between *unethicality* and *report* scores was significantly different for the brother versus stranger conditions ($r = .18$ vs. $.56, z = 4.01, p < .001$). Relative to the corresponding correlation obtained in Experiment 1, the correlation observed for the brother condition in Experiment 2 was significantly lower ($r = .55$ vs. $.18, z = 3.00, p = .003$). Thus, eliminating factual ambiguity (Experiment 2) weakened the association between *unethicality* and *report* scores when the target person was a brother. In Experiment 2, *report* scores were lower for a brother than a stranger, even though identity did not influence *unethicality* scores. This pattern suggests

that the effect of judged unethicality on willingness to report was overridden by family obligation constraint. This explanation was formally tested using path analyses, as we will discuss after reporting the results of Experiment 3.

As in Experiment 1, the correlation between *difficulty* rating and both *unethicality* and *report* scores was significantly lower (negative) in the stranger than the brother condition ($ps < .001$). In the stranger condition only, participants reported greater difficulty when they did not judge the behavior to be unethical or were not willing to report to the police.

In summary, when factual ambiguity was eliminated from the scenarios in Experiment 2, family favoritism was not observed for *unethicality* scores. When the transgression was clearly witnessed, participants conceded that the brother's behavior was as unethical as that of a stranger. Nonetheless, favoritism persisted in *report* scores: participants continued to protect their brother when faced with the more consequential decision of whether to report the transgressor to the police. The pattern of difficulty ratings indicated that participants experienced greater difficulty when making moral judgments about their brother than about a stranger, with the moral dilemma reducing willingness to report.

Experiment 3

The basic aim of Experiment 3 was to replicate the overall pattern observed in Experiments 1 and 2, while manipulating factual ambiguity of scenarios within a unified experimental design.

Method

Participants

Amazon Mechanical Turk was used to recruit 341 participants (189 females, $M_{\text{age}} = 35.8$ years, $SD_{\text{age}} = 12.0$ years) residing in the United States. Distribution of ethnicity was as follows: 75% European/European American, 11% African American/Black, 7% Hispanic or Spanish origin, 6% Asian/Asian American, and 1% classifying themselves as others. Participants received \$1.40 as compensation for completing the experiment, which took a median of 6.5 minutes.

After collecting data, we conducted post-hoc power analyses using G*Power software. We targeted the smallest significant effect found in the previous experiments when predicting the three key dependent variables: *factual*, *unethicality*, and *report* scores. The effect of severity on *report* observed in Experiment 1 was the smallest in size: Cohen's f of 0.21 ($\eta_p^2 = .044$). Given the sample size of 341, numerator degree of freedom of 1, and eight groups, the power was .97. The smallest significant effect of identity—the central independent

variable—was observed when predicting the *unethicality* score in Experiment 1. Cohen's f of 0.27 ($\eta_p^2 = .069$). The corresponding power was higher than .99.

Design, materials, and procedure

The basic materials were very similar to those used in Experiments 1–2. The design was a 2 (identity) $\times$ 2 (severity) $\times$ 2 (ambiguity: ambiguous/unambiguous) $\times$ 2 (scenario) factorial, with all factors manipulated between-subjects. About 20 participants were assigned to each of the 16 groups. Merging across the scenario types, there were at least 40 participants in each of the analyzed groups (stranger*misdemeanor*unambiguous = 44, stranger*misdemeanor*ambiguous = 41, stranger*felony*unambiguous = 40, stranger*felony*ambiguous = 43, brother*misdemeanor*unambiguous = 43, brother*misdemeanor*ambiguous = 42, brother*felony*unambiguous = 43, brother*felony*ambiguous = 45). The central question in this experiment was whether the interaction of identity and ambiguity would significantly predict the *unethicality* and *report* scores.

The general procedure of the main task was very similar to the procedures of Experiments 1–2. As a minor change in all felony scenarios, we stated that “someone else already called for an ambulance, but you’re the only one who knows about (the target person).” This addition was introduced to rule out the possibility that participants might decide to report the target person to the police only in order to ensure that the victim would receive medical treatment. In the ambiguous condition only, scenarios provided a possible alternative explanation of what could have caused the incident, and the *factual* question was asked as in Experiment 1. The *unethicality* and *report* questions were asked in all conditions, in that order. All participants were told about the identity of the transgressor before answering any questions.

Prior to the main experiment, participants were asked two pairs of questions (provided in the supplemental online materials; see Author's Note) to assess their attitudes to the general moral values of loyalty to family and upholding the law. These questions were used to provide covariates for analyses of responses to the main experimental questions. Similar questions were asked after the main experiment, but stated in the specific terms used in the scenarios (i.e., loyalty to one's brother). Because participant feedback indicated that the latter

questions made little sense for those in the stranger condition (in which a brother had never been mentioned), responses were not analyzed. Participants were also asked why they decided to report (or not report) the target person, demographic information, and whether they had siblings and if so how close they were with their siblings.

Results and discussion

As in the previous experiments, neither the specific scenario nor having versus not having a sibling (89% reported that they did) influenced the basic impact of identity on the major questions. For participants whose target person was their brother and the situation was ambiguous, we identified those who had siblings ($n = 77$; one participant was dropped for not providing a closeness rating, leaving 76) and checked if their closeness rating correlated with any of the four dependent measurements (*factual*, *unethicality*, *report*, and *difficulty*). None of the four correlations were significant ($|r|s < .12$, $ps > .31$). Thus, participants who were close to their siblings did not produce notably different responses from those who were not close to their siblings. As in the previous experiments, we collapsed the scores across scenario and sibling variables for further analyses.

Responses to the main questions (*factual*, *unethicality*, *report*, and *difficulty*) were analyzed in a series of ANOVAs.² The *factual* question was only asked in the ambiguous condition. For this condition, a 2 (identity) $\times$ 2 (severity) ANOVA was performed ($n = 171$). The main effect of identity of the target person was significant, $F(1, 167) = 23.65$, $p < .001$, $\eta_p^2 = .124$, while the main effect of severity ($p = .10$) and its interaction with identity was not ($p = .698$). As shown in Fig. 2a (left), participants had a stronger belief that the target person committed a transgression when the target person was a stranger ($M = 4.07$, $SD = 0.97$) than when he was a brother ($M = 3.17$, $SD = 1.42$).

For *unethicality* and *report* scores, ANOVAs were performed using data for all conditions, with ambiguity of scenario as an additional factor. For *unethicality*, this analysis yielded significant effects of identity, $F(1, 333) = 10.32$, $p = .001$, $\eta_p^2 = .030$, severity, $F(1, 333) = 43.64$, $p < .001$, $\eta_p^2 = .116$, and ambiguity, $F(1, 333) = 71.75$, $p < .001$, $\eta_p^2 = .177$. The interaction between identity and ambiguity approached significance, $F(1, 333) = 3.19$, $p = .075$; no other interactions were significant ($ps > .34$). In the ambiguous condition, *unethicality* was considerably higher for the stranger condition ($M = 3.61$, $SD = 1.11$) than for the brother condition ($M = 3.02$, $SD = 1.41$), whereas in the unambiguous condition these means were much more similar in stranger ($M = 4.37$, $SD = 1.00$) and brother conditions ($M = 4.22$, $SD = 1.04$).

Because the comparison of results from Experiments 1 and 2 indicated that identity only influenced *unethicality* when scenarios were unambiguous (in accord with earlier work;

² Parallel ANCOVAs were performed, which included as covariates scores on the pretest assessing prior attitudes about family loyalty and upholding the law. These analyses (reported in supplementary online materials) confirmed all the main statistical results from the ANOVA analyses.

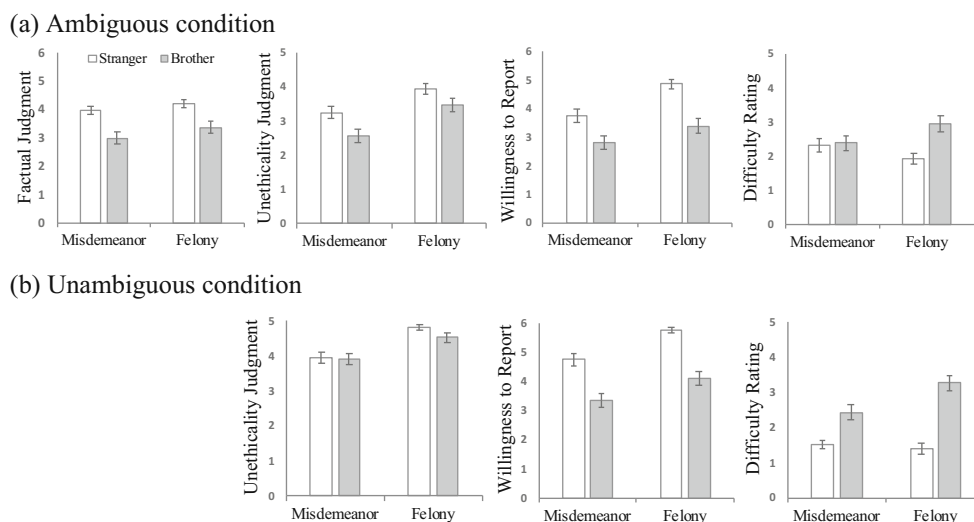


Fig. 2 Means of *factual*, *unethicality*, *report*, and *difficulty* scores for (a) ambiguous and (b) unambiguous conditions in Experiment 3. Error bars indicate standard error of the mean

Holyoak & Simon, 1999; Simon et al., 2015), a follow-up analysis was performed to examine the influence of identity separately for the ambiguous and unambiguous conditions. For the ambiguous conditions, both identity, $F(1,167) = 10.18, p = .002, \eta_p^2 = .057$, and severity, $F(1,167) = 19.29, p < .001, \eta_p^2 = .10$, yielded significant main effects, while the interaction between the two factors was not significant ($p = .57$). For the unambiguous conditions, only severity had a significant main effect, $F(1,166) = 26.02, p < .001, \eta_p^2 = .135$; neither the main effect nor the interaction involving identity was significant ($ps > .25$). Thus, participants in Experiment 3 only judged *unethicality* to be lower for a brother than a stranger when factual ambiguity was present.

For *report* scores, a three-way ANOVA revealed significant main effects of identity, $F(1,333) = 80.1, p < .001, \eta_p^2 = .194$, severity, $F(1,333) = 31.88, p < .001, \eta_p^2 = .087$, and ambiguity, $F(1,333) = 25.43, p < .001, \eta_p^2 = .071$. None of the interactions were significant ($ps > .197$). As shown in Fig. 2, participants were more likely to report the transgression when the target person was a stranger ($M = 4.78, SD = 1.40$) rather than a brother ($M = 3.42, SD = 1.63$), whether or not the facts of the situation were ambiguous. Thus, the central findings of Experiments 1 and 2 were replicated. When the scenario was factually ambiguous, participants' *factual*, *unethicality*, and *report* scores all shifted to favor a brother over a stranger; when the scenario was unambiguous, *unethicality* score did not significantly predict *report* score, but participants were nonetheless reluctant to report their brother to the police.

For *difficulty* ratings, a three-way ANOVA revealed a significant main effect of identity, $F(1, 333) = 53.92, p < .001, \eta_p^2 = .139$, an interaction between identity and severity, $F(1, 333) = 13.34, p < .001, \eta_p^2 = .039$, and an interaction between

identity and ambiguity, $F(1, 333) = 10.59, p = .001, \eta_p^2 = .031$. Difficulty rating was generally higher for the brother condition ($M = 2.77, SD = 1.44$) than the stranger condition ($M = 1.79, SD = 1.07$). Severity of transgression significantly increased difficulty in the brother condition ($M_{mis} = 2.41$ vs. $M_{felony} = 3.11, SD_{mis} = 1.32, SD_{felony} = 1.47, t(171) = 3.75, p < .001, d = 0.57$, but slightly decreased it in the stranger condition ($M_{mis} = 1.91$ vs. $M_{felony} = 1.66, SD_{mis} = 1.05, SD_{felony} = 1.09, t(166) = 1.28, p = .202$, consistent with the results from the two earlier experiments. Elimination of ambiguity significantly decreased difficulty in the stranger condition ($M_{amb} = 2.12$ vs. $M_{unamb} = 1.45, SD_{amb} = 1.14, SD_{unamb} = 0.88, t(166) = 3.52, p < .001, d = 0.54$, but not in the brother condition ($M_{amb} = 2.68$ vs. $M_{unamb} = 2.86, SD_{amb} = 1.43, SD_{unamb} = 1.45, t(171) = 0.83, p = .33$).

Table 2 shows the correlations between the dependent scores, computed separately for ambiguous and unambiguous conditions. We controlled for the main effect of severity on *unethicality*, *report*, and *difficulty* scores by standardizing them within each of the misdemeanor ($n = 170$) and felony ($n = 171$) conditions prior to the correlational analyses. In the ambiguous condition, the association between *factual* and *unethicality* scores was marginally different for brother ($r = .57$) versus stranger conditions ($r = .33, z_{\text{difference}} = 1.96, p = .051$). The association between *factual* and *report* scores was not significantly different between brother ($r = .57$) and stranger conditions ($r = .47, z_{\text{difference}} = 0.88, p = .38$); nor was the correlation between *unethicality* and *report* scores significantly different for brother ($r = .57$) versus stranger conditions ($r = .45, z_{\text{difference}} = 1.03, p = .30$). In contrast, when the factual situation was unambiguous, the correlation between *unethicality* and *report* scores was higher for the stranger condition ($r = .67$) than that for the brother condition ($r = .41, z_{\text{difference}} = 2.35, p = .019$).

Table 2 Correlations between dependent variables in Experiment 3

	Ambiguous scenarios (<i>n</i> = 171)			Stranger condition		
	Brother condition					
	Unethicality	Report	Difficulty	Unethicality	Report	Difficulty
Factual	.57***	.57***	.26*	.33**	.47***	-.07
Unethicality		.57***	.43***		.45***	.00
Report			.36***			-.27*
	Unambiguous scenarios (<i>n</i> = 170)			Stranger condition		
	Brother condition					
		Report	Difficulty		Report	Difficulty
Unethicality		.41***	.32**		.67***	-.33**
Report			-.03			-.54***

The variables significantly affected by severity (*unethicality*, *report*, and *difficulty*) were standardized within each of the misdemeanor and felony conditions and merged, thus controlling for the overall impact of type of transgression. * $p < .05$. ** $p < .01$. *** $p < .001$

This pattern of correlations largely replicates the relationships observed in Experiments 1 and 2. In the ambiguous condition, the same conduct is interpreted as factually less problematic when the target is a brother rather than a stranger. Judged *unethicality* is altered so as to cohere with the factual judgment; after this tacit “correction,” the relationship between *unethicality* and *report* scores is approximately the same regardless of target identity. In contrast, the unambiguous condition does not support a coherence shift, so participants judged an action as about equally unethical regardless of identity. Faced, therefore, with the unresolved dilemma of duty to report versus duty to protect a family member, participants often favor the family member despite the perceived unethicality of his action, yielding a weaker correlation between *unethicality* and *report* for the brother condition.

The associations between *difficulty* rating and other dependent variables were similar to the pattern observed in the previous studies. The difference in coefficients between brother and stranger conditions was tested for each pair. For all pairwise comparisons, the correlation was lower (more negative) in the stranger than in the brother condition, both in the ambiguous conditions ($ps < .032$) and in the unambiguous conditions ($ps < .001$).

Path analyses

As a formal test of the hypothesized links connecting identity to the various factual and moral queries, we conducted path analyses separately for the ambiguous and unambiguous conditions, combining the data from all three experiments. Specifically, we combined the data from ambiguous conditions in Experiments 1 and 3 ($n = 340$) and from unambiguous conditions in Experiments 2 and 3 ($n = 497$). Separate path analyses were conducted using the R package lavaan (Version 0.5-23; Rosseel, 2012). The main advantage of using lavaan is convenient access to Lagrange multiplier tests and advanced

model fit measures such as maximum likelihood estimation with robust standard errors and Satorra–Bentler scaled test statistics (Satorra & Bentler, 2010). Our focus was mainly on whether the identity and severity variables had direct or indirect effect on each of the dependent variables. We chose the best-fitting models in which dependent variables (*factual*, *unethicality*, and *report*) were either predicted or not predicted by experimentally manipulated variables (identity and severity). Also, we constrained the direction of effects between the dependent variables to be consistent with the order of presentation of corresponding questions from Experiments 1 and 3. Hence, *report* could not predict *factual* or *unethicality*, and *unethicality* could not predict *factual*. The identity and severity variables were dummy coded (1: brother, 0: stranger; 1: felony, 0: misdemeanor, respectively).

For the ambiguous conditions, the model with the best fit, $\chi^2(2) = 1.91$, $p = .385$, RMSEA = .000, 90% CI of RMSEA [.000, .114], CFI = 1.000, TLI = 1.000, had *factual* score predicted by identity, *unethicality* predicted by *factual* and severity, and *report* predicted by all other variables: *unethicality*, *factual*, identity, and severity. All paths were significant ($ps < .007$). This model (see Fig. 3a) implies that identity had a direct influence on *factual* and *report*, but not on *unethicality*. We also performed confirmatory mediation analyses for indirect effects using bootstrap confidence intervals (Preacher & Hayes, 2008), which confirmed the above patterns. These results are provided as supplemental online materials (see Author’s Note).

For the unambiguous conditions, the best fitting model, $\chi^2(1) = 0.66$, $p = .416$, RMSEA = .000, 90% CI of RMSEA [.000, .121], CFI = 1.000, TLI = 1.000, had *unethicality* predicted by severity and *report* predicted by identity, *unethicality*, and severity (see Fig. 3b). All paths were significant ($ps < .001$). Notably, identity had a significant direct effect on *report* but not on *unethicality*.

Overall, the path analyses favored similar models for the ambiguous and unambiguous conditions, with the salient

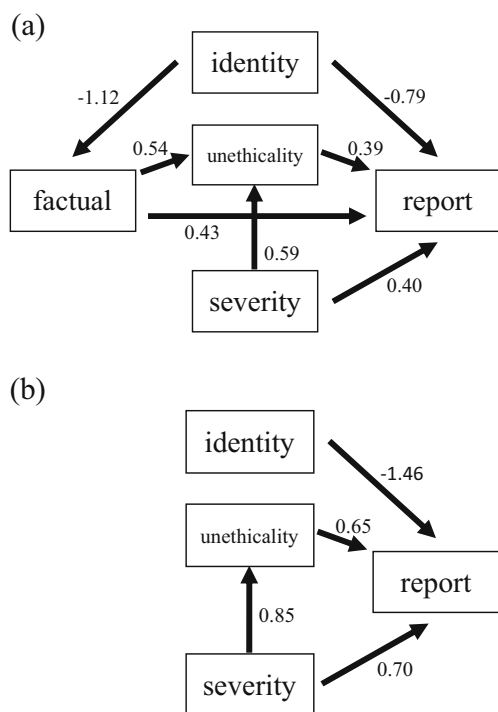


Fig. 3 Path models that best fit the data from (a) ambiguous (Experiments 1 and 3) and (b) unambiguous conditions (Experiments 2 and 3). Presented coefficients were all significant

difference that in the ambiguous condition, identity directly influenced the *factual* score (which in turn influenced both *unethicity* and *report*). In contrast, in the unambiguous condition the influence of identity on *report* was solely a direct one. These analyses provide more formal support for the hypotheses proposed earlier. Participants' judgments of unethicity were dependent on their factual understanding of the situation. When factual ambiguity was absent, participants arrived at an impartial unethicity judgment, but family favoritism was still able to override the influence of the unethicity judgment on willingness to report.

General discussion

The present study considered people's reasoning about a moral event in terms of three components: factual understanding of the event, moral judgment toward the act, and practical decision about how to respond to the event (whether or not to report the person who apparently committed a transgression to the civil authorities). Across three experiments, we found that both factual and moral judgments are affected by the personal relationship between the moral reasoner and the target person who appears to commit a transgression. Given factual ambiguity, participants judged the same behavior to be less likely to constitute an actual transgression (e.g., the target person may have only tried to defend himself) when the target was identified as their brother rather than a stranger.

Both standard ANOVAs and path analyses showed that given factual ambiguity participants rated an action as less unethical, and were less willing to report it to police, when the target person was their brother rather than a stranger. The path analysis indicated both a direct influence of target identity on willingness to report and also an indirect influence via the impact on the assessment of the situational facts. This pattern is consistent with a coherence shift (Holyoak & Powell, 2016; Holyoak & Simon, 1999; Simon et al., 2015): Ambiguity is exploited to interpret the facts in a way that favors a family member, thereby reducing judged unethicity and decreasing willingness to report.

When the situation was rendered unambiguous, such that the target person incontrovertibly committed the transgression, it was no longer possible to generate a coherent set of beliefs that would excuse the family member. The action was then judged equally unethical regardless of whether it was done by a brother or a stranger. However, when faced with a clear conflict between the agent-neutral duty to report crimes and the agent-relative duty to protect a family member, participants often elected to fulfill the latter duty at the expense of the former. Consequential moral decisions such as reporting a family member to the police are not made easily because people typically have a strong sense of obligation to protect closely related family members, even at a high cost (Burnstein et al., 1994).

We have used the term "coherence shift" as a general description for a process by which one's beliefs, motivations, and decisions dynamically align with one another during reasoning so as to achieve higher internal consistency (Holyoak & Simon, 1999; Holyoak & Powell, 2016; Simon et al., 2015). At least two types of cognitive processes may have played a role in the coherence shifts observed in the present study. One explanation for the family favoritism observed in factual and unethicity judgments is motivated reasoning (Ditto et al., 2009; Kahan et al., 2012; Kunda, 1990). That is, participants' motivation to favor and protect their in-group member (brother) may have led them to judge the "facts of the case" in a way that favored that in-group member. Motivated reasoning account has often been associated with constraint-satisfaction models of reasoning and an intuitionist view of moral judgment (Haidt, 2012). Haidt (2012) argued that strong desires and immediate reactions toward objects or agents often shape people's moral judgment, whereas a deliberate reasoning process simply rationalizes it.

However, we cannot confidently attribute the observed family favoritism to motivation because our experiments did not include a direct measure of participants' motivation to protect their brother in our experiments. The observed coherence shifts for factual and unethicity judgments might also be understood in terms of the person-centered approach of "virtue ethics" (Pizarro & Tannenbaum, 2011; Uhlmann, Pizarro, & Diermeier, 2015) coupled with rational Bayesian

inference. People are often predisposed to evaluate a person's virtue or quality (whether a person is good or bad), and to treat that evaluation as a source of prior knowledge to use in interpreting that person's behavior and intentions in the future. For example, harmful choices are perceived to be more blameworthy or intentional when the agents were characterized as immoral, dislikeable, or unfair as opposed to moral, likeable, or fair (Alicke & Zell, 2009; Kliemann, Young, Scholz, & Saxe, 2008; Siegel, Crockett, & Dolan, 2017). From this perspective, in-group favoritism—differential factual and moral judgments for different agents—might be understood as a consequence of Bayesian inference operating on differing prior beliefs about the moral character of agents, rather than as a simple bias based on double standards.

These two accounts are not mutually exclusive, and both may have contributed to our pattern of findings. The two potential factors at play—personal relationship with the target person (presumably leading to stronger motivation to care and protect) and knowledge about the target person's character—are strongly correlated in real-world situations, which makes it very difficult to disassociate their effects. A possible direction for future research might be to create artificial situations in which participants have to make judgments about a closely related person whom they do not know well (e.g., a long-lost sibling), versus a person who is not personally related but is well-known as being moral (e.g., a popular benefactor).

Morality presumably evolved to reward cooperation and loyalty in relatively small groups (Greene, 2014; Haidt, 2012). Perhaps as a consequence, many members of today's societies often react to increasing levels of demographic or cultural diversity with heightened in-group favoritism and parochialism (e.g., ethnocentrism). In this context, our findings have potential implications for finding practical remedies for sharp disagreement in moral judgment between opposing parties in the real world. For instance, people holding contrasting prior beliefs (e.g., political liberals vs. conservatives; Slothuus, Leeper, & Druckman, 2018; or death penalty supporters vs. opposers; Jern, Chang, & Kemp, 2014) are inclined to favor their own groups and agendas, often producing drastically different moral evaluations toward the same event, person, or policy. If these conflicting moral judgments hinge at least in part on people's factual understanding of an event, then reducing ambiguity in interpreting the situation may help reduce the gap between people's differing moral judgments. Nevertheless, in many cases of real-world controversies (such as political disputes), the inherent complexity of the problem may impede reasoners from easily reducing the factual ambiguity.

It would be useful in future studies to investigate the impact of in-group favoritism on moral reasoning process in different cultural contexts (Graham, Meindl, Beall, Johnson, & Zhang, 2016). Understanding cross-societal and within-societal

differences in moral norms and reasoning is growing more crucial given the fierce clashes of cultures that we face today. For example, Chen, Brockner, and Katz (1998) found evidence that people from individualistic (e.g., United States) versus collectivistic (e.g., China) cultures may have fundamentally different rationales for favoring in-group members (self-enhancement versus unconditional dedication to the group).

In summary, we found that people favored their brother when making consequential moral decisions. When rationally possible (i.e., when the situation was ambiguous), they did so after coherence-based reasoning had paved the way by creating a justification for leniency. But even when the situation was unambiguous, and their brother's guilt was clear, participants still often refused to report him to the police. Family obligation appears to be a powerful constraint that affects people's decisions. The reason some of our participants stated—“Because he's my brother”—is often sufficient justification for a moral decision.

Acknowledgements We thank Zach Home, Derek Powell, and Dan Simon for helpful discussions.

Author's note This research was supported by NSF Grant BCS-1827374. A preliminary report of part of this study was presented at the 40th Annual Conference of the Cognitive Science Society (Madison WI, July 2018). The data and materials for all experiments and description of supplementary mediation analyses are available as supplemental online material at Open Science Framework (https://osf.io/qrhdz/?view_only=bcc932afa88c47bb80a2863ae027b1f6). None of the experiments was preregistered.

References

- Alicke, M. D., & Zell, E. (2009). Social attractiveness and blame. *Journal of Applied Social Psychology*, 39(9), 2089–2105. <https://doi.org/10.1111/j.1559-1816.2009.00517.x>
- Baron, J., & Spranca, M. (1997). Protected values. *Organizational Behavior and Human Decision Processes*, 70(1), 1–16.
- Bloom, P. (2011). Family, community, trolley problems, and the crisis in moral psychology. *The Yale Review*, 99(2), 26–43. <https://doi.org/10.1111/j.1467-9736.2011.00701.x>
- Bonnefon, J. F., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles. *Science*, 352(6293), 1573–1576.
- 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. <https://doi.org/10.1037//0022-3514.67.5.773>
- Chen, Y. R., Brockner, J., & Katz, T. (1998). Toward an explanation of cultural differences in in-group favoritism: The role of individual versus collective primacy. *Journal of Personality and Social Psychology*, 75(6), 1490–1502. <https://doi.org/10.1037//0022-3514.75.6.1490>
- Cikara, M., Farnsworth, R. A., Harris, L. T., & Fiske, S. T. (2010). On the wrong side of the trolley track: Neural correlates of relative social valuation. *Social Cognitive and Affective Neuroscience*, 5(4), 404–413. <https://doi.org/10.1093/scan/nsq011>

- Ditto, P. H., Pizarro, D. A., & Tannenbaum, D. (2009). Motivated moral reasoning. In B. H. Ross (Series Ed.) & D. M. Bartels, C. W. Bauman, L. J. Skitka, & D. L. Medin (Eds.), *Psychology of learning and motivation, Vol. 50: Moral judgment and decision making* (pp. 307–338). San Diego, CA: Academic Press. [https://doi.org/10.1016/S0079-7421\(08\)00410-6](https://doi.org/10.1016/S0079-7421(08)00410-6)
- Ferreira-Garcia, R., Fontenelle, L. F., Moll, J., & de Oliveira-Souza, R. (2014). Pathological generosity: An atypical impulse control disorder after a left subcortical stroke. *Neurocase*, 20(5), 496–500. <https://doi.org/10.1080/13554794.2013.826681>
- Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford, CA: Stanford University Press.
- Foot, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review*, 5(1), 5–15. <https://doi.org/10.1093/0199252866.003.0002>
- Goldin, G., & Darlow, A. (2013). TurkGate (Version 0.4.0) [Computer software]. Retrieved from <http://gideongoldin.github.com/TurkGate/>
- Graham, J., Meindl, P., Beall, E., Johnson, K. M., & Zhang, L. (2016). Cultural differences in moral judgment and behavior, across and within societies. *Current Opinion in Psychology*, 8, 125–130. <https://doi.org/10.1177/106939718602000109>
- Greene, J. D. (2014). *Moral tribes: Emotion, reason, and the gap between us and them*. New York, NY: Penguin Press. <https://doi.org/10.1163/17455243-01402003>
- Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, 293(5537), 2105–2108. <https://doi.org/10.1126/science.1062872>
- Haidt, J. (2007). The new synthesis in moral psychology. *Science*, 316(5827), 998–1002. <https://doi.org/10.1126/science.1137651>
- Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. New York, NY: Pantheon Books.
- Haidt, J., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research*, 20(1), 98–116. <https://doi.org/10.1007/s11211-007-0034-z>
- Hamilton, W. D. (1964). The genetical evolution of social behavior: II. *Journal of Theoretical Biology*, 7(1), 17–52. [https://doi.org/10.1016/0022-5193\(64\)90039-6](https://doi.org/10.1016/0022-5193(64)90039-6)
- Holyoak, K. J., & Powell, D. (2016). Deontological coherence: A framework for commonsense moral reasoning. *Psychological Bulletin*, 142(11), 1179–1203. <https://doi.org/10.1037/bul0000075>
- Holyoak, K. J., & Simon, D. (1999). Bidirectional reasoning in decision making by constraint satisfaction. *Journal of Experimental Psychology: General*, 128(1), 3–31. <https://doi.org/10.1037/0096-3445.128.1.3>
- Jern, A., Chang, K. M. K., & Kemp, C. (2014). Belief polarization is not always irrational. *Psychological Review*, 121(2), 206. <https://doi.org/10.1037/a0035941>
- Kahan, D. M., Hoffman, D. A., Braman, D., Evans, D., & Rachlinski, J. J. (2012). They saw a protest: Cognitive illiberalism and the speech-conduct distinction. *Stanford Law Review*, 64, 851–906.
- Kahane, G., Wiech, K., Shackel, N., Farias, M., Savulescu, J., & Tracey, I. (2012). The neural basis of intuitive and counterintuitive moral judgment. *Social Cognitive and Affective Neuroscience*, 7(4), 393–402. <https://doi.org/10.1093/scan/nsr005>
- Kahane, G., Everett, J. A., Earp, B. D., Caviola, L., Faber, N. S., Crockett, M. J., & Savulescu, J. (2018). Beyond sacrificial harm: A two-dimensional model of utilitarian psychology. *Psychological Review*, 125(2), 131–164. <https://doi.org/10.1037/rev0000093>
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3), 480–498. <https://doi.org/10.1037/0033-2909.108.3.480>
- Kliemann, D., Young, L., Scholz, J., & Saxe, R. (2008). The influence of prior record on moral judgment. *Neuropsychologia*, 46(12), 2949–2957. <https://doi.org/10.1016/j.neuropsychologia.2008.06.010>
- Liu, B. S., & Ditto, P. H. (2013). What dilemma? Moral evaluation shapes factual belief. *Social Psychological and Personality Science*, 4(3), 316–323. <https://doi.org/10.1177/1948550612456045>
- Nagel, T. (1979). *Mortal questions*. Cambridge, UK: Cambridge University Press.
- Nagel, T. (1986). *The view from nowhere*. New York, NY: Oxford University Press.
- Petrinovich, L., O'Neill, P., & Jorgensen, M. (1993). An empirical study of moral intuitions: Toward an evolutionary ethics. *Journal of Personality and Social Psychology*, 64(3), 467.
- Pizarro, D. A., & Tannenbaum, D. (2011). Bringing character back: How the motivation to evaluate character influences judgments of moral blame. In M. Mikulincer & P. R. Shaver (Eds.), *The social psychology of morality: Exploring the causes of good and evil* (pp. 91–108). Washington, DC: American Psychological Association. <https://doi.org/10.1037/13091-005>
- Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, 40(3), 879–891.
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36. Retrieved from <http://www.jstatsoft.org/v48/i02/>
- Russo, J. E., Carlson, K. A., Meloy, M. G., & Yong, K. (2008). The goal of consistency as a cause of information distortion. *Journal of Experimental Psychology: General*, 137(3), 456–470. <https://doi.org/10.1037/a0012786>
- Satorra, A., & Bentler, P. M. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika*, 75(2), 243–248. <https://doi.org/10.1007/s11336-009-9135-y>
- Shackelford, T. K., & Hansen, R. D. (Eds.). (2016). *The Evolution of Morality*: Springer International Publishing.
- Siegel, J. Z., Crockett, M. J., & Dolan, R. J. (2017). Inferences about moral character moderate the impact of consequences on blame and praise. *Cognition*, 167, 201–211. <https://doi.org/10.1016/j.cognition.2017.05.004>
- Simon, D., Krawczyk, D. C., & Holyoak, K. J. (2004). Construction of preferences by constraint satisfaction. *Psychological Science*, 15(5), 331–336. <https://doi.org/10.1111/j.0956-7976.2004.00678.x>
- Simon, D., Stenstrom, D. M., & Read, S. J. (2015). The coherence effect: Blending cold and hot cognitions. *Journal of Personality and Social Psychology*, 109(3), 369–394. <https://doi.org/10.1037/pspa0000029>
- Simon, D., Snow, C. J., & Read, S. J. (2004). The redux of cognitive consistency theories: Evidence judgments by constraint satisfaction. *Journal of Personality and Social Psychology*, 86(6), 814–837. <https://doi.org/10.1037/0022-3514.86.6.814>
- Singer, P. (1979). *Practical ethics*. Cambridge, UK: Cambridge University Press.
- Slothuus, R., Leeper, T. J., & Druckman, J. N. (2018). Motivated responses to political communications: Framing, party cues, and science information. In H. Lavine & C. S. Taber (Eds.), *The feeling, thinking citizen: Essays in honor of Milton Lodge* (pp. 125–150). New York, NY: Routledge.
- Uhlmann, E. L., Pizarro, D. A., Tannenbaum, D., & Ditto, P. H. (2009). The motivated use of moral principles. *Judgment and Decision Making*, 4(6), 476–491.
- Uhlmann, E. L., Pizarro, D. A., & Diermeier, D. (2015). A person-centered approach to moral judgment. *Perspectives on Psychological Science*, 10, 72–81. <https://doi.org/10.1177/1745691614556679>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.