

Running head: INDIRECT AGENCY

Dirty Work, Clean Hands: The Moral Psychology of Indirect Agency

Neeru Paharia

Karim S. Kassam

Joshua D. Greene

Max H. Bazerman

Harvard University

ABSTRACT

When powerful people cause harm, they often do so indirectly through other people. Are harmful actions carried out through others evaluated less negatively than harmful actions carried out directly? Four experiments examine the moral psychology of *indirect agency*. Experiment 1 reveals effects of indirect agency under conditions favoring intuitive judgment, but not reflective judgment, using a joint/separate evaluation paradigm. Experiment 2 demonstrates that effects of indirect agency cannot be explained by perceived lack of foreknowledge or control on the part of the primary agent. Experiment 3 indicates that reflective moral judgment is sensitive to indirect agency, but only to the extent that indirectness signals reduced foreknowledge and/or control. Experiment 4 indicates that effects of indirect agency result from a failure to automatically consider the potentially dubious motives of agents who cause harm indirectly.

Keywords: *ethics, indirect agency, moral psychology, decision-making*

Dirty Work, Clean Hands: The Moral Psychology of Indirect Agency

In 2005, Merck, a major pharmaceutical company, sold the rights to two slow-selling cancer drugs, Mustargen and Cosmegen, to smaller and lesser-known Ovation Pharmaceuticals. Merck's decision to sell the drugs had dire consequences for dependent users, some of whom saw their monthly drug costs jump from \$160 to \$1,100 after Ovation raised the price (Berenson, 2006). The price increase was not based on changes in cost, and Merck continued to manufacture the drugs. Though Merck did not directly increase the price, it is likely that Merck foresaw the price increase and shared indirectly in the resulting profit. Were Merck's actions ethical? And did the indirect nature of their actions affect how they were perceived?

When powerful people do morally questionable things, they rarely interact directly with their putative victims. Mobsters have hit men. CEO's have vice presidents, lawyers, and accountants. More specifically, the powerful are likely to carry out their intentions through the actions of other agents, with varying degrees of explicit direction and control. The present investigation explores the effects of indirect agency on moral judgment.

Recent work in moral psychology examines the roles played by various forms of physical indirectness in the evaluation of harmful actions (Royzman & Baron, 2002; Waldmann & Dieterich, 2007; Cushman, Young, & Hauser, 2006; Greene, Sommerville, Nystrom, Darley & Cohen, 2001; Greene, Lindsay, Clarke, Nystrom, & Cohen, 2008). Royzman & Baron (2002), for example, contrast moving someone in front of oneself in order to block an incoming bullet (direct) with jumping out of a bullet's path, exposing someone else as a target (indirect). The present research differs from the foregoing in two key respects. First, it is not focused on *physical* indirectness. Had Merck raised the prices of its drugs *directly*, rather than selling the

rights to Ovation, the physical chain of events connecting Merck to its disappointed customers would not necessarily have been any shorter. The indirectness in question is conceptual, not physical. Second, the indirectness in question concerns the role of *secondary agents*. Merck (the *primary agent*) acted indirectly in the present sense because it acted through Ovation (the *secondary agent*).

Milgram (1974) examined the role of secondary agents in a variant of his classic obedience experiment. He found that subjects were even more likely to shock a victim (when directed by an authority figure) when the subjects were allowed to act through a secondary agent (another confederate of the experimenter) who could be directed to administer shocks to the victim from a remote location. Milgram proposed that the presence of a secondary agent created distance between the subject and the victim, and that this distance reduced the strain on the participant and made disobedience less likely. Milgram's experiments suggest that the presence of secondary agents can reduce strain for the actor, but do not address the question of how observers perceive harmful actions that are carried out through secondary agents.

We conducted four experiments to examine the moral psychology of indirect agency. Experiment 1 aims to document effects of indirect agency under conditions favoring intuitive vs. reflective judgment. Experiments 2 and 3 aim to determine whether effects of indirect agency result from uncertainty concerning the primary agent's foreknowledge of, or control over, harmful outcomes. Experiment 4 tests the hypothesis that effects of indirect agency result from a failure to automatically consider the potentially dubious motives of agents who cause harm indirectly.

EXPERIMENTS 1A-C

In Experiment 1, we examined the effects of indirect agency under two distinct modes of evaluation: separate evaluation and joint evaluation. Under joint evaluation, multiple alternatives are simultaneously available for consideration, while under separate evaluation only one alternative is considered (Bazerman, Loewenstein, & White, 1992). For example, Bazerman, Schroth, Shah, Diekmann, & Tenbrunsel (1994) asked MBA students which of two options they would prefer: a job paying an annual salary of \$75,000 while their classmates also made \$75,000 (*the fair option*), or a job paying \$85,000 while their classmates made \$95,000 (*the money option*). Students exhibited different preferences based on how the scenarios were presented. When presented separately, participants reported that they would be more likely to accept the *fair option* than the *money option*, despite the reduced salary. When the two options were evaluated jointly, however, the relationship was reversed, with participants more likely to accept the *money option* even though it seemed less fair. Students who saw the scenarios one at a time based their judgments more on fairness, an emotionally salient factor that can be evaluated in the absence of information about alternative outcomes. Those who saw scenarios together focused on absolute salary level, a factor made salient when both options are presented together and that most people, upon reflection, regard as the more important factor.

The tendency for separate evaluation to engender affective or intuitive judgment, and for joint evaluation to engender more cognitive or reflective judgment is not specific to this particular scenario (Hsee, Rottenstreich & Xiao, 2005; Bazerman, Tenbrunsel, & Wade-Benzozi, 1998). In another study, Hsee (1996) asked participants to evaluate second-hand music dictionaries either separately or jointly. In separate evaluation, participants stated they would be willing to pay more for a dictionary with 10,000 entries whose condition was 'like new', whereas

in the joint evaluation condition participants said they would pay more for a 20,000 entry dictionary whose cover was torn. The dictionary's aesthetic properties were intuitively evaluable to participants and thus dominated separate evaluation. But in joint evaluation, the difference in number of entries was salient and became the more important factor. In general, joint evaluation allows for the consideration of attributes that are difficult to assess in isolation and is therefore more likely to result in reason-based choice and less likely to be influenced by affective factors (Hsee 1998; Bazerman et al., 1998).

In Experiment 1, we used a joint/separate evaluation paradigm to document the effects of indirect agency, pitting the effect of indirectness against another relevant factor, the magnitude of harmful consequences. An effect of indirect agency may naturally be seen as a bias. For example, few would explicitly argue that a murderous mobster is less culpable if he acts through a hit man rather than pulling the trigger himself. Thus, we predict that effects of indirect agency, insofar as they exist, are more likely to be present under conditions favoring intuition (separate evaluation). But, if the comparison between indirect and direct agency is made more transparent under conditions favoring reflection (joint evaluation), the blame reduction for indirect agency will be reduced or eliminated. More specifically, we predict that under conditions of separate evaluation, people's judgments will be dominated by the effect of indirect agency and demonstrate a preference for a greater indirect harm. Under conditions of joint evaluation, on the other hand, people's judgments will be dominated by the effect of magnitude and demonstrate a preference for a lesser direct harm.

Experiment 1A employed hypothetical scenarios modeled after the case of Merck and Ovation. In the direct scenario, the larger company directly raised the price of the drug, imposing hardship on patients. In the indirect scenario, the larger company sold the drug rights to the

smaller company that went on to raise the price, imposing even greater hardship on patients. We predicted that participants would find indirect action (with more severe consequences) to be less unethical when alternatives were presented separately, but that people would reverse their preferences when the alternatives were presented jointly. Experiments 1B and 1C tested the same hypotheses using alternative scenarios.

EXPERIMENT 1A

Method

Participants

One-hundred and fifty people (128 females $M_{\text{age}} = 32.7$, $SD = 10.0$) recruited from an online subject pool completed the survey and a package of unrelated surveys in exchange for \$5.

Procedure

Participants were randomly assigned to one of three conditions: *separate evaluation – direct sale* (SE-direct), *separate evaluation – indirect sale* (SE-indirect), or *joint evaluation* (JE).

In the SE-direct condition, participants read the following scenario:

A major pharmaceutical company, X, had a cancer drug that was minimally profitable. The fixed costs were high and the market was limited. But, the patients who used the drug really needed it. The pharmaceutical was making the drug for \$2.50/pill (all costs included), and was only selling it for \$3/pill. [The pharmaceutical firm raised the price of the drug from \$3/pill to \$9/pill, thus increasing the value of the drug to company X by \$10 million.]

In the SE-indirect condition, the text in brackets was replaced with the following: “The major pharmaceutical X sold the rights to a smaller pharmaceutical, Y, for \$12 million. In order to re-coop costs, company Y increased the price of the drug to \$15/pill.”

All participants were asked “*On a scale of 1 (not at all unethical) - 10 (very unethical), how unethical do you think company X's behavior was in this decision?*” In the JE condition, the text in brackets was replaced with the following:

Consider two cases in which the firm had either:

Case A) Raised the price of the drug from \$3/pill to \$9/pill, thus raising the value of the drug to company X by \$10 million,

or

Case B) Sold the rights to produce the drug to a smaller company Y for \$12 million. In order to re-coop costs, company Y raised the price of the drug to \$15/pill.

Participants were then asked the following question: “In which case would the behavior of company X have been more unethical?” Participants responded using an eleven point scale anchored at -5 with “In terms of ethics, Case A is much worse than Case B”, at 0 with “In terms of ethics, the two cases are equally ethical/unethical”, and at 5 with “In terms of ethics, Case B is much worse than Case A”. We simultaneously manipulated indirectness and magnitude of harm in this experiment in order to generate a preference reversal. In Experiments 2-3 we examine the effect of indirectness while controlling for magnitude of harm.

Following previous research (Bazerman et al., 1992), we asked participants in the joint condition an explicitly comparative question in order to enable a natural comparison of the two options, rather than using two independent rating scales for each option which would parallel the separate evaluation conditions. The use of different scales necessitates a more conservative test

of our hypothesis than the standard interaction that could have been tested if the same scales had been used in both conditions (i.e. it requires significant effects in opposing directions).

We had participants address questions of “ethicality” rather than “morality,” because the former term is more often used in business contexts such as those described in our scenarios. Students in professional schools of business or medicine might take a course in “ethics,” but are unlikely to take a course in “morality.”

Results

Results – Separate Evaluation

SE results revealed a significant effect of *directness*, $t(77) = 3.13$, $p < .005$, indicating that the *SE-direct* price increase was rated as significantly more unethical ($M = 6.85$, $SD = 2.47$) than the *SE-indirect* price increase ($M = 5.05$, $SD = 2.84$).

Results – Joint Evaluation

The JE condition was analyzed using a one-sample t-test with zero (scale midpoint) as the test value. The results revealed a significant effect $t(70) = 2.88$, $p = .005$, indicating that the indirect price increase was rated as more unethical ($M = .94$, $SD = 2.76$) when the two cases were seen together.

EXPERIMENT 1B

We investigated whether the same pattern observed in Experiment 1A would replicate in other business contexts. In Experiment 1B, instead of looking the direct and indirect implication of a drug sale, we looked at toxic waste clean up. In the *direct* condition, a well-known real estate company X decided to only clean up 40% of the toxic waste on the land of a new housing development. In the *indirect* condition, they sell the land to a lesser-known company, Y, who

does not engage in any clean up effort. We predicted that the *direct* condition would be viewed less favorably in separate evaluations, but that the *indirection* condition would be viewed less favorably in the *joint* evaluation.

Methods

Participants

One hundred and sixty-five participants (84 females, $M_{\text{age}} = 23.62$, $SD = 11.01$) were recruited on or around Harvard University campus. A candy bar was offered as incentive.

Procedure

Participants were randomly assigned to one of three conditions: *separate evaluation – direct price increase* (SE-direct), *separate evaluation – indirect price increase* (SE-indirect), or *joint evaluation* (JE). In the SE-direct condition, participants read the following scenario:

A well-known real estate developer, X, owned a piece of property that they wished to construct new housing units on. The property contained some health-threatening toxic substances that would require a substantial amount of clean-up, and was worth \$50 million dollars as is. It would require \$30 million to fully clean the land, but the value would only go up to \$60 million. [The housing developer decided to only invest \$12 million in a 40% clean up effort, and the value of the land went up to \$54 million. They built housing units on the land, all of which have now been sold.]

In the SE-indirect condition, the text in brackets was replaced with the following: “The housing developer sold the land to a lesser known developer, Y, without cleanup. The lesser-known developer invested no money in any clean up effort and built housing units on the land, all of which have now been sold.”

All participants were asked “*On a scale of 1 (not at all unethical) - 10 (very unethical), how unethical do you think company X's behavior was in this decision?*” In the JE condition, the text in brackets was replaced with the following:

Consider two cases in which the developer had either:

Case A) Invested only \$12 million in a 40% clean up effort and the value of the land went up to \$54 million. They built housing units on the land, all of which have now been sold.

Case B) The housing developer sold the land to a lesser known developer, Y, without cleanup. The lesser-known developer invested no money in any clean up effort, and built housing units on the land, all of which have now been sold.

Participants were then asked: “In which case would the behavior of company X have been more unethical?” Participants responded using the same scale used in Experiment 1A.

Results – Separate Evaluation

The results revealed a significant effect, $t(86) = 2.38, p < .05$, indicating that the direct price increase was rated as significantly more unethical ($M = 7.09, SD = 2.40$) than the indirect price increase ($M = 5.80, SD = 2.68$).

Results – Joint Evaluation

The results revealed a significant effect $t(76) = 3.47, p < .01$, indicating that the indirect price increase was rated as more unethical ($M = 1.17, SD = 2.96$) when the two cases were seen together.

EXPERIMENT 1C

Experiment 1C another followed the format of Experiments 1A and 1B but in an air pollution domain. In the *direct* condition, a well-known electricity producer company X turned down their pollution controls on a coal factory by 50% to increase their profit from \$1 to \$2,

resulting in 25% more pollution. In the *indirect* condition company X sold the factory to a lesser-known company Y who turned down the pollution controls by 75% to increase profit from \$1 to \$4, resulting in 75% more pollution. We predicted that the *direct* condition would be viewed less favorably in separate evaluations, but that the *indirect* condition would be viewed less favorably in the *joint* evaluation.

Methods

Participants

Sixty-nine participants (35 females, $M_{\text{age}} = 25.62$, $SD = 13.56$) were recruited on or around Harvard University campus. A candy bar was offered as incentive.

Procedure

Participants were randomly assigned to one of three conditions: *separate evaluation – direct* (SE-direct), *separate evaluation – indirect* (SE-indirect), or *joint evaluation* (JE). In the SE-direct condition, participants read the following scenario:

A major electricity producer, company X, owned a factory that produced power out of coal. The factory was using a variety of pollution controls that helped the environment but reduced profitability. Company X was earning a profit of \$1.00 per unit of energy produced. [Company X reduced the use of pollution controls by 50% and increased their profit margin from \$1.00 to \$2.00 per unit of energy produced. The value of their factory went up to \$8 million. However, by turning down the pollution controls, the amount of pollution emitted from the factory increased by 25%.]

In the SE-indirect condition, the text in brackets was replaced with the following: “Company X sold the factory to a lesser-known company Y for \$12 million dollars. In order to re-coop costs, company Y reduced the use of pollution controls by 75% and increased their profit margin from

\$1.00 to \$4.00 to per unit of energy produced. However, by turning down the pollution controls, the amount of pollution emitted from this factory increased by 75%.”

All participants were asked “*On a scale of 1 (not at all unethical) - 10 (very unethical), how unethical do you think company X's behavior was in this decision?*” In the JE condition, the text in brackets was replaced with the following:

Consider two cases in which the firm had either:

Case A) Reduced the use of pollution controls by 50% and increased their profit margin from \$1.00 to \$2.00 per unit of energy produced. The value of their factory went up to \$8 million. However, by turning down the pollution controls, the amount of pollution emitted from the factory increased by 25%.

Case B) Sold the factory to a lesser-known company, Y, for \$12 million dollars. In order to re-coop costs, company Y reduced the use of pollution controls by 75% and increased their profit margin from \$1.00 to \$4.00 to per unit of energy produced. However, by turning down the pollution controls, the amount of pollution emitted from this factory increased by 75%.

Participants were then asked: “In which case would the behavior of company X have been more unethical?” Participants responded using the same scale used in Experiments 1A and 1B.

Results – Separate Evaluation

The results revealed a significant effect, $t(46) = 2.20$, $p < .05$, indicating that the direct price increase was rated as significantly more unethical ($M = 6.92$, $SD = 2.0$) than the indirect price increase ($M = 5.21$, $SD = 3.243$).

Results – Joint Evaluation

The *joint* evaluation condition was analyzed using a one-sample t-test with zero (scale midpoint) as the test value. The results revealed a significant effect $t(20) = 5.273, p < .001$, indicating that the indirect price increase was rated as more unethical ($M = 2.81, SD = 2.44$) when the two cases were seen together.

Discussion

When participants evaluated alternatives separately, participants were significantly more lenient when harm was caused indirectly. When alternatives were assessed jointly, the effect of consequences replaced or dominated the effect of indirectness. This preference reversal suggests that the use of indirect agency has a significant impact on intuitive moral judgment, but that when participants are allowed to compare alternatives and reflect on their decisions, the indirectness of an action is seen as less important than the magnitude of its harmful consequences.

EXPERIMENT 2

Indirect agency may attenuate moral condemnation because it is unclear whether the primary agent foresaw the resulting harm. We might assume that Merck foresaw dramatically increased prices when they sold their drugs to Ovation, but it's possible that they did not. If Merck failed to foresee the harm they caused, they may be held less accountable for that reason (Young, Hauser & Saxe, 2007; Cushman, 2008). A second relevant factor is the primary agent's perceived level of control over the secondary agent. A secondary agent may be seen as acting autonomously, or alternatively, as serving as an instrument of the primary agent (willingly or otherwise). For example, even if Merck correctly anticipates that Ovation will raise the price

(foreknowledge), the decision to raise the price may nevertheless be Ovation's. A decrease in the perceived control of the primary agent may result in a perceived diffusion of responsibility (Darley and Latane, 1968) from the primary to the secondary agent.

In Experiments 2 and 3, we included conditions in which the agent clearly had foreknowledge of the consequences, and in which the secondary agent was transparently an instrument of the primary agent. (That is, the primary agent had both foreknowledge of and control over the harmful outcome. Note that it is possible for the primary agent to have foreknowledge without control, but not vice versa.) If indirect action attenuates blame as a result of a perceived lack of foreknowledge or control of the primary agent, judgments in these conditions should mirror judgments of direct action. We examined the effects of these factors on relatively intuitive versus relatively reflective judgments using separate evaluation in Experiment 2 and joint evaluation in Experiment 3. In Experiments 2-3, unlike Experiment 1, the magnitude of the harm was held constant across conditions/scenarios in order to better isolate the targeted effects.

Method

Participants

Five hundred and eighteen Kentucky residents (301 females $M_{\text{age}} = 44.3$, $SD = 13.85$) were recruited through an online survey service to complete a group of unrelated surveys.

Procedure

Participants were randomly assigned to one of four separate evaluation conditions: *direct*, *indirect*, *indirect — foreknowledge*, and *indirect — control* (foreknowledge + control). The *direct* and *indirect* conditions were nearly identical to the separate conditions run in Experiment 1A, except that in the *indirect* conditions, company Y only raised the price to \$9 rather than \$15, and

company X's profit only increased by \$10 million, thus holding the consequences constant between the *direct* and the *indirect* conditions at an increase to \$9 per pill. The *indirect – foreknowledge* condition was identical to the *indirect* condition but with the addition of the following sentence to indicate foreknowledge: “Company X was aware that company Y would raise the price to \$9/pill before the sale of the drug.” In the *indirect – control* condition, the text in brackets (see Experiment 1) was replaced by the following to indicate both foreknowledge and control:

Rather than brand and distribute the drug themselves incurring a cost of \$100,000 to company X, they made a contract with company Y for this service. Under the contract, company Y agreed to sell the product under company Y's name and through their distribution channels for \$9/pill. Company X paid company Y \$100,000 for this service and increased the value of the drug to company X by \$10 million.

The SE rating scale was as in Experiment 1.

Results

Ratings of ethicality were analyzed using a one-way between subjects ANOVA with four conditions (*direct* vs. *indirect* vs. *indirect – foreknowledge* vs. *indirect – control*). The analysis revealed a significant effect, $F(3, 514) = 5.36, p < .01$. Planned comparisons revealed that company X was rated as significantly more unethical in the *direct* condition ($M = 7.45, SD = 3.10$) than in the *indirect* condition, ($M = 5.85, SD = 3.37$), the *indirect – foreknowledge* condition, ($M = 6.35, SD = 3.05$), and *indirect – control* condition ($M = 6.62, SD = 3.25$), $t(514) = 3.88, p < .01$, $t(514) = 2.77, p < .01$, and $t(514) = 2.13, p = .03$ respectively (see Figure 1). The difference between the *indirect* and *indirect – control* conditions was marginally significant

$t(514) = 1.93, p = .054$. The difference between the *indirect* and *indirect — foreknowledge* conditions was not significant $t(514)=1.22, p = .22$.

[Insert Figure 1 about here]

Discussion

As in Experiment 1, actions carried out indirectly through secondary agents were deemed less unethical than actions carried out directly in separate evaluations favoring intuitive judgment. The results further suggest that indirectness is not merely a signal of reduced foreknowledge or control. Indirect actions were judged more unethical than direct actions even when the primary agent's foreknowledge was made explicit and when the secondary agent was transparently an instrument of the primary agent, acting under contract such that the primary agent had full foreknowledge of and control over the outcome.

In Experiment 3, we examined the influence of the same two factors on reflective moral judgment by using joint evaluation.

EXPERIMENT 3

Method

Participants

Forty-four participants (25 females $M_{\text{age}} = 27.7, SD = 12.4$) from the Cambridge, MA community participated in exchange for a candy bar. One participant failed to answer all questions and was excluded from the analysis.

Procedure

Participants rated the ethicality of company X in all four scenarios used in Experiment 2 (*direct*, *indirect*, *indirect – foreknowledge*, and *indirect – control*) in a joint evaluation on the same ten point scales. The order of scenarios was fully randomized across subjects.

Results

Ratings of ethicality were analyzed using a repeated measures ANOVA with *condition* (*direct* vs. *indirect* vs. *indirect – foreknowledge* vs. *indirect – control*) as the within-subjects factor. The analysis revealed a significant effect, $F(3, 126) = 6.40, p < .01$. Planned comparisons show that company X was rated as less unethical in the *indirect* condition ($M = 4.56, SD = 2.91$) than in the *direct* ($M = 5.92, SD = 2.81$), $t(42) = 3.54, p < .01$, *indirect – foreknowledge* ($M = 5.77, SD = 2.87$), $t(42) = 3.41, p < .01$, and *indirect – control* conditions ($M = 5.76, SD = 2.71$), $t(42) = 2.91, p < .01$. There were no significant differences among the latter three conditions, all t 's < 1 , all p 's $> .6$ (see Figure 2).

[Insert Figure 2 about here]

Discussion

Under conditions favoring reflective judgment (joint evaluation), participants did not distinguish between harm caused directly and harm caused indirectly when foreknowledge and control were explicitly attributed to the primary agent. However, indirect action was judged more leniently when the primary agent's foreknowledge and/or degree of control were ambiguous. This suggests reflective moral judgment is sensitive to indirect agency, but only to the extent that indirectness signals reduced foreknowledge and/or reduced control on the part of the primary

agent. In Experiment 1, where the price of the pill was raised to \$15 in the *indirect* condition, as opposed to only \$9 in the *direct* condition, the action in the *indirect* condition was judged to be more unethical in a joint evaluation. In that case, the effect of consequences dominated the effect of indirectness. Here, however, with the consequences held constant across conditions, *indirect* action was seen as less unethical than direct action when the primary agent's degree of foreknowledge/control were ambiguous.

EXPERIMENT 4

Experiment 4 tested an additional hypothesis concerning the mechanism behind the indirect agency effect: People who evaluate indirect actions in isolation (SE) respond relatively favorably to these actions because they are relatively unlikely to question the primary agent's motives. Joint evaluation, in contrast, prompts people to consider the primary agent's motives by drawing attention to the more straightforward alternative of direct action. This hypothesis predicts that the effect of indirect agency can be reduced or eliminated by first drawing people's attention to the possibility of dubious motives on the part of the primary agent.

Method

Participants

One-hundred and forty-two participants (79 females $M_{\text{age}} = 40.2$, $SD = 21.1$) were recruited on or around the Harvard University campus.

Procedure

Participants were randomly assigned to one of three conditions: *direct*, *indirect – motive-first*, or *indirect – motive-last*. The *direct* condition was identical to the direct condition in Experiment 2. The *indirect – motive-first* condition was identical to the *indirect – foreknowledge*

condition in Experiment 2, but with the addition of the following text, presented prior to the ethicality rating scale:

Company X decided to sell the drug to company Y rather than increase the price and market the drug themselves. One possible motive for selling to company Y is that company X was worried that raising the price would cause negative press. Selling to company Y avoids the negative reputation that might result from a large increase in drug price.

Next, participants indicated the extent to which they believed that this motive was influential in company X's decision. They used a ten point scale anchored by *Not at all Influential* and *Very Influential*. In the *indirect – motive-last* condition, the additional text and the motive-influence scale were presented after the ethicality scale. The *indirect-motive last* condition was thus similar to the *indirect-foreknowledge* conditions of Experiments 2 and 3, and was included to demonstrate that participants' perceptions of motive did not depend on question order.

Results

Ratings of ethicality were analyzed using a one-way, between-subjects ANOVA (*direct* vs. *indirect – motive-first* vs. *indirect – motive-last*). The analysis revealed a significant effect on ethicality ratings, $F(2, 141) = 6.88, p < .01$. Planned comparisons revealed that company X was rated as significantly more unethical in both the *direct* ($M = 6.36, SD = 2.44$) and *indirect motive-first* ($M = 5.42, SD = 2.64$) conditions than in the *indirect – motive-last* condition ($M = 4.40, SD = 2.60$), $t(141) = 3.71, p < .01$ and $t(141) = 1.99, p < .05$ respectively. The difference between the *direct* and *indirect motive-first* conditions was marginal, $t(141) = 1.78, p = .08$.

Ratings of the motive's influence did not differ significantly between the *indirect – motive-first* ($M = 6.60, SD = 2.17$) and *indirect – motive-last* ($M = 5.98, SD = 2.31$) conditions, $t(98) = 1.38, p > .16$.

Discussion

As predicted, participants who considered the possibility of dubious motives prior to evaluating company X's behavior (*indirect – motive-first*) rated the behavior as more unethical than those prompted to consider dubious motive afterwards (*indirect – motive-last*). The *motive-first* manipulation did not completely eliminate the effect of indirect agency, but did reduce the size of the effect.

GENERAL DISCUSSION

Four studies explored the moral psychology of indirect agency. Experiment 1 revealed a moral preference for indirect agency under conditions favoring intuitive judgment (separate evaluation). This preference for indirectness was reversed under conditions of joint evaluation, which favors reflective judgment. Experiment 2 demonstrated that the effect of indirect agency is not merely an effect of perceived foreknowledge or control on the part of the primary agent. Participants favored indirect action even when the secondary agent was explicitly described as an instrument of the primary agent, contracted to do its bidding. Results of Experiment 3 suggest that reflective moral judgment takes indirect agency into account, but only insofar as it signals reduced foreknowledge and/or control on the part of the primary agent. Experiment 4 tested the hypothesis that the effect of indirect agency is in part caused by a failure to automatically consider potentially dubious motives behind acts of indirect agency. Consistent with this hypothesis, highlighting such motives beforehand reduced the effect of indirect agency.

What, then, is the cause of the preference for indirect acts? Experiments 2 and 3 speak against two plausible explanations (attribution of reduced foreknowledge and/or control). The latter explanation posits a diffusion of responsibility (Darley and Latane, 1968) from the primary agent to the secondary agent when the primary agent's control over the secondary agent is in question. If the diffusion of responsibility explanation is correct, then diffusion requires no perceived lack of control on the part of the primary agent. That is, even when the primary agent is in complete control of the action, the *mere existence* of a secondary agent may be enough to create a perception of diffused responsibility. Note that the "diffusion" metaphor presupposes a fixed quantity of responsibility. A tendency to think of blame and responsibility as fixed quantities to be apportioned among agents may therefore be an important facilitating factor in the indirect agency effect.

Is the leniency toward indirect harmful agency an irrational bias? The data suggest that it is. The joint/separate reversals observed in Experiment 1 is logically inconsistent and cannot be rationally defended. The question is whether it is the separate evaluation or the joint evaluation (or both) that is the source of the irrationality. It does not automatically follow that joint evaluations are more rational or otherwise superior to separate evaluations. Joint evaluation may lead one to discount important factors that are best appreciated intuitively or to overemphasize relatively unimportant factors that are consciously accessible and amenable to comparison (Bazerman et al., 1998). Nevertheless, we maintain that the preference for indirect agency observed under separate evaluation is irrational for two reasons. First, Experiments 2 and 3 rule out what we regard as the two most plausible justifications for favoring indirect actions. Second, Experiment 4 suggests that the effect of indirect agency depends largely on a failure to attend to possible motivations for acting indirectly.

In terms of broader theory, the present results are consistent with Greene et al.'s dual-process theory of moral judgment (2001, 2004, 2007), according to which moral judgment is shaped by controlled cognitive processes that are highly sensitive to the consequences of actions as well as by automatic/intuitive processes that are often emotional and susceptible to bias. The current study adds to the dual process theory by specifying the conditions under which controlled cognitive versus automatic/intuitive processes are more likely to dominate. The present results are harder to reconcile with traditional rationalist theories of moral judgment (Kohlberg, 1969) because of the clear irrationality involved in joint/separate reversals. The present results are also in tension with Haidt's (2001) Social Intuitionist Model (SIM), according to which moral judgment is driven primarily by intuitive emotional responses. The challenge for the SIM is to explain the results of joint evaluation conditions, where judgments appear to be based primarily on the magnitude of harmful consequences. Judgments under joint evaluation appear to involve a conscious comparison involving an explicit normative metric ("Causing more harm is more unethical")—i.e. consequentialist moral reasoning. The SIM does make room for moral reasoning in exceptional cases, and one may argue that JE is exceptional.

A bias that favors indirect harms may have implications for how crimes are punished in our legal system. In the case of murder, the United States law does not differentiate between murder that is carried out directly and murder that is carried out indirectly (United States v. Simmons 2006). For example, if a mobster hires a hit man to kill his associate, he may be convicted of first degree murder based on the evidence that he solicited and facilitated murder. In this case, the law is no more lenient when a secondary agent is involved. The present research suggests that juries, on the other hand, may be biased to look more favorably on indirect harms and thus less likely to deliver guilty verdicts when harm is committed through a secondary agent.

Conclusion

It is possible that Merck's decision to sell its cancer drugs to Ovation was entirely innocent. But, if their decision was a deceptive strategy, it may have been more clever than they realized. Acting indirectly through another can hide the fact that one has caused harm, hide the fact that one knowingly chose to cause harm, and hide the extent of one's control over the harmful outcome. But the present results indicate that causing harm indirectly through another can protect harm-doers, and thus harm society, in a more subtle and insidious way. This is important to know, given that many of the greatest crimes against society are perpetrated by powerful people who carry out their intentions through others. The present results suggest that heightened awareness of people's sometimes dubious motivations for acting indirectly, and the organizational structures that facilitate them, may be a useful safeguard against the abuse of power.

REFERENCES

- Berenson, A. (2006, March 12th). A cancer drug's big price rise is cause for concern. *New York Times*, p. 25.
- Bazerman, M., Loewenstein, G., and Blount White, S. (1992). Reversals of preference in allocation decisions: Judging an alternative versus choosing among alternatives. *Administrative Science Quarterly*, 37, 220-240.
- Bazerman, M. H., Schroth, H. A., Shah, P. P., Diekmann, K. A., and Tenbrunsel, A. E. (1994). The Inconsistent Role of Comparison Others and Procedural Justice to Hypothetical Job Descriptions: Implications for Job Acceptance Decisions. *Organizational Behavior and Human Decision Processes*, 60 (3), 326-352.
- Bazerman, M., Tenbrunsel, A.E., and Wade-Benzozi, K.A. (1998). Negotiation with yourself and losing: Understanding and managing competing internal preferences. *Academy of Management Review*, 23, 225-241.
- Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition*, 108 (2), 353-380.
- Cushman, F., Young, L., and Hauser, M., (2006). The Role of conscious reasoning and intuition in moral judgment: testing three principles of harm. *Psychological Science*, 17, 1082–1089.
- Darley, J. M., and Latane, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology*, 8, 377-383.
- Greene, J.D., Lindsay, D., Clarke, A.C., Nystrom, L.E., and Cohen, J.D. (2008). Pushing moral buttons: The interaction between personal force and intention in moral judgment. Unpublished Manuscript.

- Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., and Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, 44, 389-400.
- Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., and Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, 293, 2105-2108.
- Greene, J. D. (2007). The secret joke of Kant's soul, in Moral Psychology, W. Sinnott-Armstrong (Ed.), *Moral Psychology Vol. 3: The Neuroscience of Morality: Emotion, Disease, and Development* (pp. 105-118). Cambridge, MA: MIT Press,
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108, 814-834.
- Hsee, C.K. (1996). The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluations of alternatives. *Organizational Behavior and Human Decision Processes*, 67(3), 247-257.
- Hsee, C.K. (1998). Less is better: When low-value options are valued more highly than high-value options. *Journal of Behavioral Decision Making*, 11, 107-121
- Hsee, C.K., Blount, S., Loewenstein, G., and Bazerman, M. (1999). Preference reversals between joint and separate evaluations of options a review and theoretical analysis. *Psychological Bulletin*, 125, 576-590.
- Hsee, C.K., Rottenstreich, Y., and Xiao, Z. (2005) When is more better? On the relationship between magnitude and subjective value. *Current Directions in Psychological Science*, 14(5), 234-237.

Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In D. A. Goslin (Ed.), *Handbook of socialization theory and research* (pp. 347-480). Chicago: Rand McNally.

Milgram, S. (1974). *Obedience to Authority: An Experimental View*. New York: Harpercollins

Royzman, E., and Baron J. (2002). The preference for indirect harm. *Social Justice Research*, 15, 165-184.

United States v. Simmons, 431 F. Supp. 2d 38, 4 (D.D.C. 2006)

Waldmann, M., and Dieterich, J. (2007). Throwing a bomb on a person versus throwing person on a bomb: Intervention myopia in moral intuitions. *Psychological Science*, 18, 247-253.

Young, L., Cushman, F., Hauser, M., and Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *PNAS*, 104, 8235-8240.

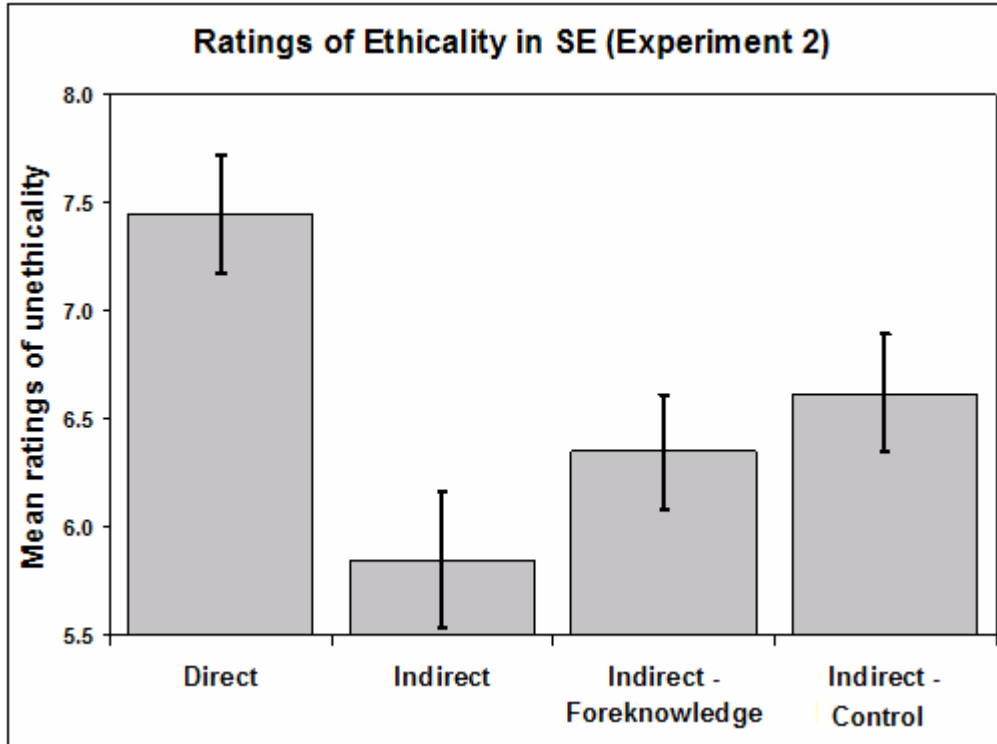


Figure 1. Ratings of Ethicality in Separate Evaluation (SE) from Experiment 2. Error bars represent standard error.

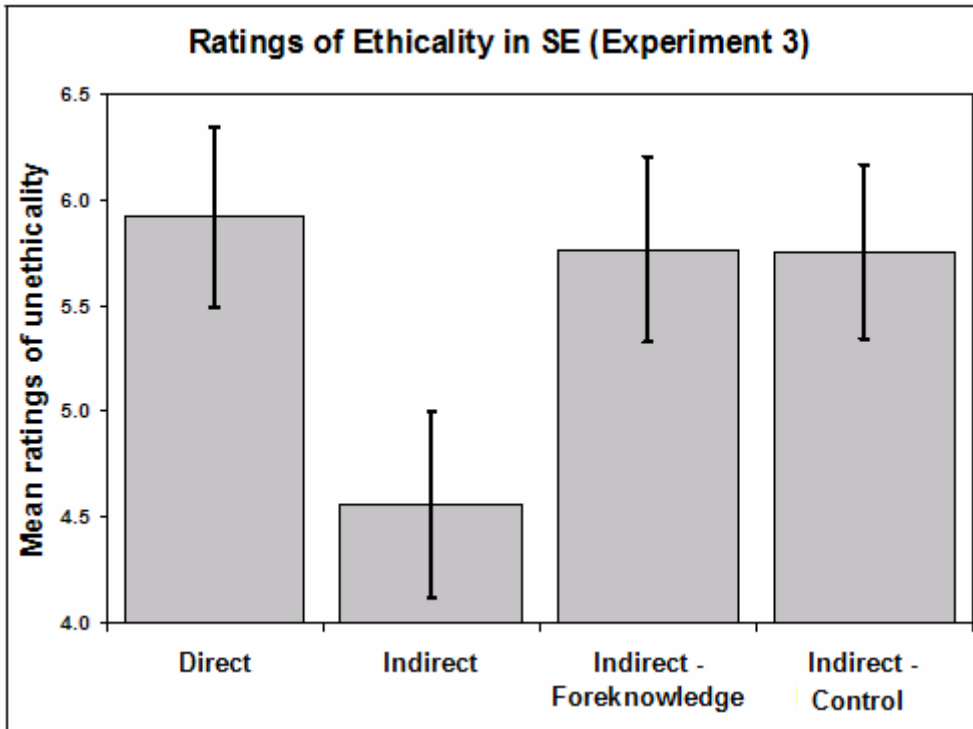


Figure 2. Ratings of Ethicality in Joint Evaluation (JE) from Experiment 3. Error bars represent standard error.