Title

Rational Hypocrisy: A Bayesian Analysis Based on Informal Argumentation and Slippery Slopes

Journal

Cognitive Science, 38(7)

ISSN

0364-0213

Authors

Rai, Tage S
Holyoak, Keith J

Publication Date

2014-09-01

DOI

10.1111/cogs.12120

Peer reviewed

COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 38 (2014) 1456–1467

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12120

Rational Hypocrisy: A Bayesian Analysis Based on Informal Argumentation and Slippery Slopes

Tage S. Rai,[a] Keith J. Holyoak[b]

[a]*Kellogg School of Management, Northwestern University*
[b]*Department of Psychology, University of California, Los Angeles*

Received 14 February 2013; received in revised form 11 July 2013; accepted 22 July 2013

Abstract

Moral hypocrisy is typically viewed as an ethical accusation: Someone is applying different moral standards to essentially identical cases, dishonestly claiming that one action is acceptable while otherwise equivalent actions are not. We suggest that in some instances the apparent logical inconsistency stems from different evaluations of a weak argument, rather than dishonesty per se. Extending Corner, Hahn, and Oaksford's (2006) analysis of slippery slope arguments, we develop a Bayesian framework in which accusations of hypocrisy depend on inferences of shared category membership between proposed actions and previous standards, based on prior probabilities that inform the strength of competing hypotheses. Across three experiments, we demonstrate that inferences of hypocrisy increase as perceptions of the likelihood of shared category membership between precedent cases and current cases increase, that these inferences follow established principles of category induction, and that the presence of self-serving motives increases inferences of hypocrisy independent of changes in the actions themselves. Taken together, these results demonstrate that Bayesian analyses of weak arguments may have implications for assessing moral reasoning.

Keywords: Moral reasoning; Hypocrisy; Bayesian inference; Argumentation

1. Introduction

"I respect the jury's verdict. But I have concluded that the prison sentence given to Mr. Libby is excessive." With these words, former President George W. Bush commuted the sentence of I. Lewis "Scooter" Libby, Jr., for obstruction of justice and leaking the identity of CIA operative Valerie Plame. Critics of the decision noted that Libby had

Correspondence should be sent to Tage S. Rai, Ford Center for Global Citizenship, Kellogg School of Management, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208. E-mail: tage.rai@gmail.com

actually received the minimum sentence allowable for his offense under the law and that many of Libby's supporters, including the Bush administration, were actively pressing for mandatory minimum sentencing laws at a national level (Goodman, 2007). Accordingly, critics of the decision saw it as a textbook case of moral hypocrisy: Different rules were being applied to Bush's underling, Libby, than to everyone else in the United States.

Moral hypocrisy is typically thought to occur when an actor employs Different standards for identical cases (Valdesolo & DeSteno, 2007). Viewed within this framework, accusations of hypocrisy entail an inherent *logical inconsistency* with *moral implications*, in that the morally hypocritical action has the same logical structure as previous cases, and thus should not be treated any differently according to some norm of justice. The assumption (at least implicitly) is that the hypocrite is being dishonest, or at least self-deceptive, because the hypocrite must be aware (or should be aware) of the logical inconsistency and is therefore committing a falsehood (Batson, Kobrynowicz, Dinnerstein, Kampf, & Wilson, 1997; for a review of hypothesized bases for moral disagreements, see Rai & Fiske, 2011).

Although people certainly sometimes engage in such genuine hypocrisy, in some instances there may be an alternative explanation of *apparent* hypocrisy, one that leaves open the possibility that the (alleged) hypocrite is both honest and rational. This possibility arises as an extension of recent Bayesian analyses of informal argumentation, in particular, analyses of slippery slope arguments.

1.1. Bayesian analyses of informal argumentation and the slippery slope

Hahn and Oaksford (2007) have argued that many seeming fallacies of informal reasoning (e.g., arguments from ignorance, circular arguments, and slippery slope arguments) should not necessarily be dismissed based on their logical structure. For example, slippery slope arguments are characterized by opposition to the adoption of a proposed action A on the grounds that even if it were deemed acceptable in isolation, it should ultimately be rejected because its acceptance would lead to the eventual adoption of an undesirable action B (Corner, Hahn, & Oaksford, 2006, 2011). An example of a common slippery slope argument is the assertion by opponents of gay marriage that if laws are passed that allow same-sex marriage, then there would be no grounds to restrict marriages between people and animals or adults and minors (Johnson, 2013). Critically, the proponent of the slippery slope argument is implying that B represents an action that has the same logical structure as A and therefore it will come to pass if A is accepted.

Corner et al.'s (2006, 2011) insight was that the acceptability of a slippery slope argument does not depend on its logical structure, but rather hinges on the conditional probability that undesirable outcome B will in fact occur if action A is approved. In practice, the relative undesirability of outcome B is based on the ethical values and perceptions of observers. Meanwhile, the conditional probability value depends in part on perceptions of the prior probability that action A will expand a category boundary far enough to include B in the future, based on the exemplars that observers have been exposed to previously. The closer that A is perceived to be to a category boundary, the more likely it is that A

will expand the category boundary. More precisely, the relevant quantity would presumably be a function of the difference between the probability of B being a category member conditional on A being a member versus A not a member (intuitively, a judgment of the probability that acceptance of A as a member will causally lead to acceptance of B; see Holyoak & Cheng, 2011). For example, the claim that legalizing marijuana use might lead to the eventual legalization of heroin may seem less persuasive than the claim that legalizing cocaine might lead to the legalization of heroin. Both claims have the same logical structure, but because cocaine is perceived as a harder drug than marijuana and this criteria is perceived as important for category inclusion, it can reasonably be claimed that the inclusion of cocaine into the realm of legal behavior increases the probability of heroin's inclusion in the future, and consequently, the slippery slope argument is more convincing.

Corner et al. (2011) asked participants to evaluate slippery slope arguments against designating a first location as having "outstanding natural beauty" because it would lead to a second location also receiving the designation. The designation "outstanding natural beauty" was supposed to be reserved for particularly beautiful locations and is based on the number of large animals living in the location. The second location had 179 large animals. Participants rated the strength of the slippery slope argument as higher when the first location had 194 large animals than when it had 218 animals, indicating that participants believed there was a higher probability of inclusion of the second location when the number of animals in the first location was more similar. Importantly, the number of animals in the first location must be higher than the number of animals in the second location. If the first location had 178 animals, similarity would further increase, but a slippery slope argument against its inclusion would not make sense because the second location (with 179 animals) would not have any greater negative utility than the first location. To the extent that the similarity of A and B leads people to believe that the inclusion of A will increase the probability of the eventual inclusion of B, *and* that B represents a highly negative outcome, slippery slope arguments will be deemed acceptable.

1.2. Inverse relationship between the slippery slope and moral hypocrisy

The logical structure of an allegation of moral hypocrisy is related to the structure of slippery slope arguments. A critic of a slippery slope argument believes that cases A and B do not (and should not) belong to the same category because they are highly dissimilar and the utility of B is much more negative, such that the inclusion of A will not lead to the inclusion of B in the future. In contrast, a critic who alleges moral hypocrisy when someone treats a new case B more favorably than a precedent case A believes that cases A and B do in fact belong to the same category because they are highly similar and the utility of B is equal or more negative, and hence should be viewed and treated similarly (or less favorably) from a moral perspective. By extension, it may be possible that just as with slippery slope arguments, perceptions of moral hypocrisy depend on inferences of prior probabilities and utilities related to content. Specifically, increasing the likelihood of cases A and B belonging to the same category, and increasing the negative utility of B, will

increase perceptions of moral hypocrisy if case B is treated more favorably than case A. Just as with slippery slope arguments, similarity is used to estimate the probability of shared category inclusion. In this paper, we use the term *moral utility* to refer to the extent to which observers view an action as morally negative, reprehensible, or evil *relative* to the treatment it receives. Just as with Corner et al.'s (2011) notion of undesirability, moral utility is based on the ethical values and perceptions of observers. Although moral utility is a criterion upon which actions can be more or less similar, it is distinct from other criteria in that *greater* dissimilarity should increase inferences of moral hypocrisy if the two actions are treated similarly. Thus, if action B is significantly worse than action A, but receives proportionally worse treatment, arguments for hypocrisy will be weakened. And just as with slippery slope arguments, increasing the negative utility of one of the cases may decrease the similarity of the cases yet *increase* perceptions of moral hypocrisy.

1.3. Rational hypocrisy

To illustrate the close link between slippery slope arguments and allegations of moral hypocrisy, let us return to the Scooter Libby commutation. We could turn the critique of Bush's action into a slippery slope argument by claiming that commuting Libby's sentence would morally require us to commute everyone else convicted of similar crimes, because Libby's crime was no less heinous than theirs. The stronger this slippery slope argument seems, the weaker is Bush's perceived defense against the allegation of moral hypocrisy. More specifically, as the similarity of Libby's crime to those of other convicted felons and the negative utility of Libby's crime increases, the strength of both the slippery slope argument and the allegation of hypocrisy increases. However, perhaps Bush in fact viewed Libby's offense as dissimilar to and less negative in utility than the crimes of others convicted of the same charges (because, for example, Libby was a "patriot" motivated by "noble" motives, such as increasing the security of the United States). Given such prior beliefs, Bush could rationally decide that commuting Libby (based on his exculpatory circumstances and exemplary character) was rationally and morally consistent with not commuting the sentences of others convicted of what appeared to him to be dissimilar crimes.

By analyzing accusations of moral hypocrisy in Bayesian terms of informal argumentation, we treat the critical observer of a potentially hypocritical moral action as being presented with two competing hypotheses. The first hypothesis is that the alleged hypocrite has acted genuinely hypocritically for selfish reasons. By *genuine* hypocrisy, we mean that the moral actor believes that a precedent case A and a current case B are highly similar and that the utility of case B is more negative, but is choosing to treat case B more favorably for selfish reasons. The moral actor's actions are thus *internally inconsistent* with their held moral beliefs. The alternative hypothesis is that the moral actor has acted in an *apparently* hypocritical way due to differences in the prior beliefs of the moral actor and the critical observer. From this perspective, no two cases are completely identical, and the critical observer's inference of moral hypocrisy may actually reflect disagreement between the observer and the moral actor over the prior probability of the two cases having a shared category membership.

Given this Bayesian framework of competing hypotheses, there are two primary routes through which inferences of moral hypocrisy will be strengthened or weakened. First, to the extent that case B is viewed as highly similar and as equally or more negative in its moral utility than case A, this will count as negative evidence against the apparent hypocrisy hypothesis, according to which the alleged hypocrite perceives that the two cases are sufficiently different as to merit preferential treatment of B. Second, to the extent that a moral actor is perceived to have selfish reasons to treat two cases differently, this will be viewed as positive evidence of genuine hypocrisy independent of any changes in the similarity or utility of the cases. Thus, for the critical observer faced with Bush's decision, the two competing hypotheses are that Bush had different prior beliefs as to whether Libby's actions were sufficiently different and less negative in their moral utility from those of other criminals as to warrant better treatment, or that Bush believed that all criminals should go to jail and that Libby was a criminal, but hypocritically commuted his sentence for selfish reasons. Viewed in a Bayesian framework, evidence that Bush may have had self-serving reasons to commute Libby's sentence would count as positive evidence in support of the hypothesis that Bush was a hypocrite, whereas evidence that Libby's case differed from that of other convicts, or that his crime carried less negative utility, would count as positive evidence in favor of the apparent hypocrisy hypothesis.

1.4. Overview of studies

In the present paper, we test our proposal that perceptions of moral hypocrisy depend on underlying probabilities related to content and moral utilities. Experiments 1 and 2 demonstrate that manipulating the probability of shared category membership and respective moral utilities for two cases affects perceptions of hypocrisy (Experiment 1) and that participants are selective in the attributes to which they attend when making accusations (Experiment 2). Experiment 3 demonstrates that the existence of potential self-serving motives for moral actors increases perceptions of moral hypocrisy, independent of changes in the similarity or moral utility of the two cases.

2. Experiment 1

2.1. Method

Participants in all three experiments were recruited via the Internet, and completion of the questionnaires was voluntary. Questionnaires for Experiments 1 and 3 were posted on the Craigslist website under their "volunteers" section in Los Angeles, New York, and Chicago. The questionnaire for Experiment 2 was posted on Amazon's Mechanical Turk website. After giving consent to participate, each participant was randomly assigned to a condition. The IP addresses of participants' computers were recorded to ensure that participants did not participate in the study multiple times.

In Experiment 1, participants ($N = 59$) were presented with one variant of an embezzlement scenario and one variant of a scenario about pregnancy laws. Scenario order was counterbalanced across participants. Each vignette was varied to experimentally manipulate differences in similarity and moral utility between a precedent action A and a proposed action B, such that the proposed action was either similar and of equal moral utility to the precedent action or highly dissimilar and of reduced negative utility. Each participant saw one high similarity/equal utility case and one low similarity/reduced utility case. We hypothesized that perceptions of moral hypocrisy due to preferential treatment would be higher when similarity between the cases was lower and the negative utility of the case in question was higher.

In one vignette, participants read about a man who was caught embezzling from his company and had been sentenced to 5 years in prison. Participants were told that his lawyer had requested the man's sentence to be reduced to 2 years. Participants in the high similarity/equal utility condition were told that other convicted felons were incarcerated for monetary crimes, such as theft and money laundering. Participants in the low similarity/reduced negative utility condition were told that other criminals were incarcerated for violent crimes, such as homicide and assault. Following the manipulation, all participants were asked to evaluate how morally hypocritical it would be for the judge to reduce John's sentence to 2 years when other convicts are sentenced to more time, using a Likert scale ranging from 1 (*not hypocritical at all*) to 7 (*completely hypocritical*). In the other vignette, participants read about a new law that had passed in Norway making it illegal for pregnant women to drink alcohol because it is passed directly to infants and can lead to severe side effects. Participants were then told that lawmakers were considering a law to ban women from consuming caffeine during pregnancy. Participants in the high similarity/equal utility condition were told that caffeine is passed directly to infants and can cause severe side effects, while participants in the low similarity/reduced negative utility condition were told that caffeine is passed indirectly and can lead to minor side effects. All participants were then asked how hypocritical it would be for the law to permit caffeine consumption while restricting alcohol consumption.

According to our hypothesis, reducing John's sentence to 2 years should be seen as more hypocritical when his crime is more similar and carries equal moral utility to the crimes of other convicts. Similarly, it should be seen as more hypocritical to allow pregnant mothers to consume caffeine but not alcohol when the effects of caffeine are similar to those of alcohol (direct rather than indirect causal pathway) and equally as harmful (severe side effects).

2.2. Results

The pattern of ratings is shown in Fig. 1. A between-subjects analysis of variance revealed that ratings of hypocrisy were significantly higher when the precedent action A and the proposed action B were highly similar based on surface and utility similarity ($M = 4.54$) than when the actions were more dissimilar ($M = 3.27$), $F(1, 55) = 7.69$, $p < .01$. Ratings did not differ across the two vignettes, $F(1, 55) < 1$. There was no

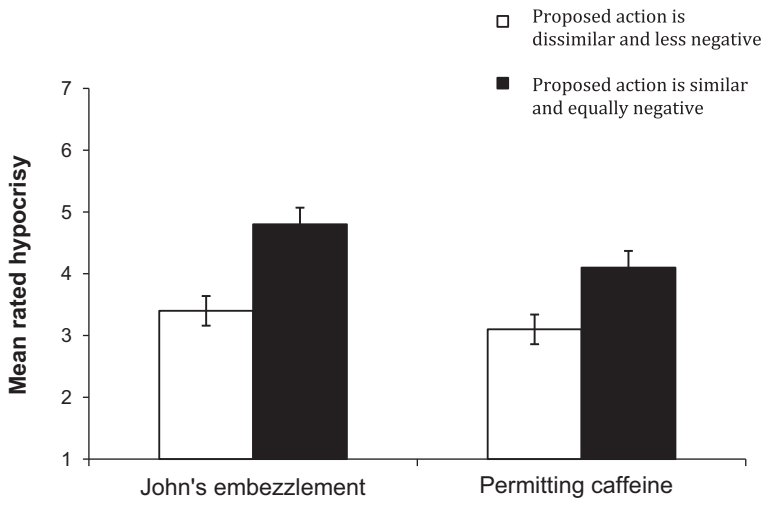

Fig. 1. Mean ratings of moral hypocrisy as a function of relative similarity and moral utility between cases, for each of two vignettes (Experiment 1). A rating of 1 corresponded to "not hypocritical at all" while a rating of 7 corresponded to "completely hypocritical." Error bars indicate one standard error of the mean.

interaction between vignette and experimental manipulation, $F(1, 55) = 1.85$, $p = .18$, indicating that the basic effect was observed for both vignettes. There was also no effect of presentation order, $F(1, 55) = 1.43$, $p = .24$.

3. Experiment 2

The purpose of Experiment 2 was to tease apart the effects of similarity and moral utility while explicating more clearly the nature of moral utility. In Experiment 1, actions differed in both similarity and underlying moral utility. In terms of a rational analysis of moral hypocrisy, it is moral utility (i.e., the costs of a proposed action for affected individuals or society as a whole) that is directly relevant, rather than overall similarity. Work on categorization has shown that although overall similarity serves as a default guide in making category judgments, specific knowledge about causally relevant features takes precedence when it is available (Rehder, 2006, 2009). In theory, the primary reason to attend to similarity is that it is useful for inferring underlying moral utility. Accordingly, we designed a set of vignettes that were intended to separately manipulate similarity and underlying moral utility in assessments of moral hypocrisy.

3.1. Method

In Experiment 2, participants ($n = 168$, distributed approximately equally across four conditions) read a single vignette about a man who destroyed $10,000 worth of

electronics and asked for his sentence to be reduced due to mitigating circumstances. Participants were then told about how the judge had given the same sentence in a previous case that varied in both similarity and moral utility. Specifically, participants were either told that the defendant in the previous case had destroyed property that had similar features (computer parts) or dissimilar features (designer clothes) and that it had resulted in either greater damage ($15,000) or lesser damage ($5,000).

Observers ultimately determine the features that are relevant for assessing moral utility versus similarity. In a pilot study, we confirmed that participants identify the amount of monetary damages as more morally relevant than the type of item that was stolen for determining how morally wrong a crime was. Participants were asked two questions: "When deciding how morally wrong a theft is, how important is it to take into account what kind of item was stolen, such as whether the thief stole electronics or clothes," and an identical question that asked about the importance of taking into account "how much the items were worth, such as a difference between $5,000 and $15,000." Order was counterbalanced across participants, and participants reported their responses on a five-point Likert scale ranging from "rather unimportant" to "rather important." Within subjects t tests revealed that participants believed that monetary damages were more important to take into account ($M = 3.17$, $SD = 1.45$) than the type of item stolen ($M = 2.91$, $SD = 1.44$), $t(120) = 2.08$, $p = .04$.

Thus, this design yielded four between-subjects conditions: high similarity/high negative utility, high similarity/low negative utility, low similarity/high negative utility, and low similarity/high negative utility. Participants then rated how hypocritical it would be to reduce the man's sentence without reducing the sentence in the previous case. We predicted that when damages are made explicit they should ignore similarity, as it carries no diagnostic value for inferring moral utility. Instead, participants should only be sensitive to the damages caused, such that it is more hypocritical to reduce the man's sentence when the previous case caused less damage, regardless of the similarity between the cases.

3.2. Results

The pattern of ratings is shown in Fig. 2. A between-subjects analysis of variance tested accusations of hypocrisy when precedent cases had involved greater or lesser damage and were either highly similar or highly dissimilar. When damages were lower in the precedent case, ratings of how hypocritical it would be to were significantly higher than when damages were higher in the precedent case, $F(1, 164) = 11.242$, $p < .001$. There were no significant differences in ratings of hypocrisy between cases that were highly similar on the surface and those that were highly dissimilar, $F(1, 164) = .077$, $p = .78$). The interaction between the damages and the similarity of the precedent case was also nonsignificant, $F(1, 164) = 1.021$, $p = .31$. Thus, whereas underlying utilities had a pronounced impact on accusations of hypocrisy, surface similarity alone had no effect when it was explicitly devoid of diagnostic value.

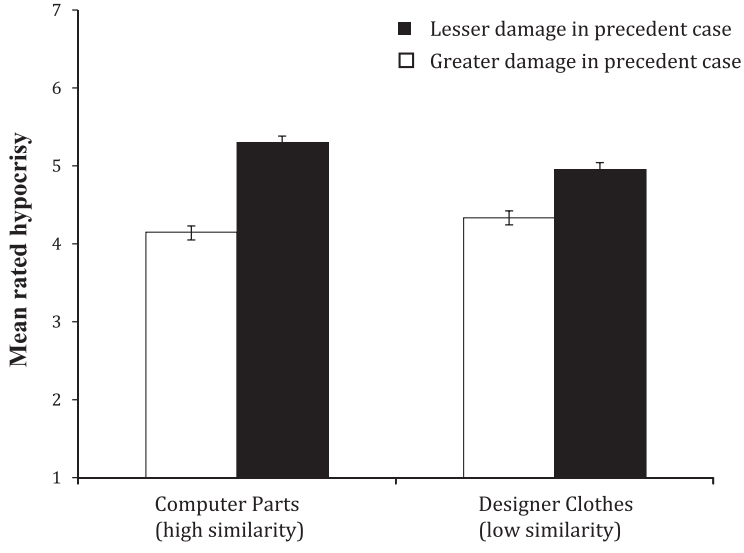

Fig. 2. Mean ratings of moral hypocrisy as a function of relative moral utility and surface similarity between the current case and the precedent case (Experiment 2). A rating of 1 corresponded to "not hypocritical at all" while a rating of 7 corresponded to "completely hypocritical." Error bars indicate one standard error of the mean.

4. Experiment 3

In Experiment 3, we aimed to demonstrate that the introduction of self-serving motives would count as positive evidence in support of the *genuine* hypocrisy hypothesis. Specifically, we varied the likelihood of self-serving motives while keeping similarity and utility between cases constant, to examine the role that self-serving motives play in accusations of hypocrisy.

4.1. Method

In Experiment 3, participants ($N = 47$, distributed approximately equally across two conditions) were presented with one vignette that was varied to experimentally manipulate the benefit an actor would likely receive from performing an action that could be construed as morally hypocritical. Specifically, participants read a vignette about James, who had been sentenced to prison for vehicular homicide in a highly publicized case. There were extreme mitigating factors, but the minimum sentencing laws required that James go to prison for several years. James requested that the governor reduce his sentence. Participants were either told that James came from a rich family that would be likely to donate to the Governor's campaign, or that he came from a poor family that would be unlikely to donate. We hypothesized that reducing James' sentence to 3 years would be seen as more hypocritical when his family is rich because observers will perceive that the governor has a self-serving interest in taking the action. The perception of

a selfish motive will count as positive evidence in favor of the hypothesis that the Governor is genuinely hypocritical, rather than believing that he perceived James' case as legitimately different enough from precedent cases so as to warrant differential treatment.

4.2. Results

Accusations of hypocrisy on the part of the Governor were significantly higher among participants who were told James came from a rich family ($M = 5.64$, $SD = 1.64$) than among participants who were told James came from a poor family ($M = 3.48$, $SD = 2.04$), $t(45) = 3.95$, $p < .001$. Thus, information about the potential for selfish gain supported accusations of hypocrisy, independent of the similarity between James' crime and the crimes of others.

5. General discussion

When presented with a potential case of moral hypocrisy, observers have two competing hypotheses for analyzing an actor's actions. Either the actor is a dishonest hypocrite lying for selfish reasons, or the similarity and utility of a proposed action B are sufficiently different from that of precedent action A to suggest that they warrant differential treatment. Here, similarity is a proxy for indexing the likelihood that two cases are included in the same category of treatment. We have defined the term "moral utility" to refer to the extent to which we assume that observers view an action as morally negative, reprehensible, or evil *relative* to the treatment it receives, where the more negative the action and the more preferential the treatment, the greater the injustice. We found that as a precedent action A and a proposed action B become more similar, and the difference in their relative utilities increases, perceptions of moral hypocrisy increase (Experiment 1). When the utilities of actions were made explicit, surface similarity of attributes that participants believe is not morally relevant did not influence judgments of hypocrisy (Experiment 2). Finally, perceptions of moral hypocrisy increase, independently of changes in the similarity or relative utility of two cases, if the potential for selfish gain is believed to be present (Experiment 3), because potential selfish motives count as positive evidence in support of the hypothesis that the actor is genuinely hypocritical.

The present findings extend analyses of other types of informal argumentation (Hahn & Oaksford, 2007) and represent the first attempt to analyze moral hypocrisy in Bayesian terms. The findings are novel in demonstrating that the logical structure of accusations of moral hypocrisy is conceptually related to that of slippery slope arguments, and that the same factors affecting the strength of slippery slope arguments also impact accusations of moral hypocrisy. Rather than being based on illogical motivated cognition, at least some apparent instances of moral hypocrisy may be rooted in Bayesian notions of argument strength. Differences in prior beliefs can lead to conflicting interpretations of argument strength, consequent accusations of hypocrisy, and attempts at rebuttal. By linking moral hypocrisy to Bayesian decision theory, we extend recent work that has argued for more

domain-general accounts of moral cognition (Cushman & Young, 2011; Rai & Holyoak, 2010; Waldmann & Dieterich, 2007; Waldmann, Nagel, & Wiegmann, 2012).

An intriguing question is whether genuine hypocrites may (explicitly or implicitly) adjust their beliefs regarding the similarity and utility between cases and regarding the potential selfish benefits they would accrue from taking the proposed action. Such instances raise the question of what counts as "genuine" hypocrisy. We have defined moral hypocrisy strictly in terms of internal inconsistency between beliefs and actions. However, if someone is self-deceptive, or heavily indoctrinated, he or she may believe that levels of similarity between cases and relative moral utilities are completely consistent with their actions, even if others would assess similarities and utilities very differently. From this perspective, the actor's reasoning would not be morally hypocritical in a strict sense, even if all rational observers agree that the actor's perception of the underlying similarities and utilities is flawed. Future studies should investigate how estimates of underlying probabilities and utilities vary in contexts related to possible hypocrisy, particularly when making judgments about one's own actions versus the actions of others (Valdesolo & Desteno, 2008).

In addition, future studies should investigate the role of similarity in greater detail. In particular, whereas our Experiment 2 demonstrated that utility takes precedence over similarity when utilities are made explicit, future studies should investigate whether people will use similarity between cases to make inferences of underlying moral utility (similar to work on categorization; Rehder, 2006, 2009). Finally, in Experiment 3 self-serving motives were confounded with the defendant's wealth, in that the Governor has greater self-serving motives when James' family is rich. It is possible that participants simply infer more moral hypocrisy when a rich rather than a poor family is involved. Future studies should hold wealth constant while only manipulating the likelihood of receiving financial benefit.

Finally, the Bayesian framework implies that rational analyses of competing hypotheses are occurring at all levels of judgment. Thus, in our Experiment 3, participants were informed of the Governor's potential self-serving motive. An interesting question for future research would be to consider how the source of the information affects its impact on inferences of hypocrisy. If the source of the information is perceived as one with its own selfish motives for casting doubt on a moral actor (e.g., a sensationalistic media personality seeking higher ratings), then observers may judge the actor's actions as less hypocritical, because they will perceive greater ambiguity as to whether the actor's motives are truly selfish. Rather than reflecting the operation of an entirely separate mental faculty, moral judgments may provide especially compelling examples of domain-general probabilistic reasoning operating on causal models of human action.

Acknowledgments

Preparation of this paper was supported by the UCLA Center for Society and Genetics (TR) and by ONR grant N000140810186 (KH). We thank Hongjing Lu and Chris Carroll

for comments on earlier drafts. A preliminary report of Experiment 1 was presented at the 33nd Annual Conference of the Cognitive Science Society (Boston, July 2011).

References

Batson, D., Kobrynowicz, D., Dinnerstein, J., Kampf, H., & Wilson, A. (1997). In a very different voice: Unmasking moral hypocrisy. *Journal of Personality and Social Psychology*, *72*, 1335–1348.

Corner, A., Hahn, U., & Oaksford, M. (2006). The slippery slope argument—Probability, utility & category reappraisal. In R. Sun, & N. Miyake (Eds.), *Proceedings of the 28th annual meeting of the cognitive science society* (pp. 1145–1150). Austin, TX: Cognitive Science Society.

Corner, A., Hahn, U., & Oaksford, M. (2011). The psychological mechanism of the slippery slope argument. *Journal of Memory and Language*, *64*, 133–152.

Cushman, F., & Young, L. (2011). Patterns of moral judgment derive from nonmoral psychological representations. *Cognitive Science*, *35*, 1052–1075.

Goodman, A. (2007). Commuting sentence, Bush spares Libby from 30 month jail-term. Democracy Now: The War and Peace Report. Available at http://www.democracynow.org/2007/7/3/commuting_sentence_bush_spares_libby_from. Accessed January 1, 2013.

Hahn, U., & Oaksford, M. (2007). The rationality of informal argumentation: A Bayesian approach to reasoning fallacies. *Psychological Review*, *114*, 704–732.

Holyoak, K. J., & Cheng, P. W. (2011). Causal learning and inference as a rational process: The new synthesis. *Annual Review of Psychology*, *62*, 135–163.

Johnson, L. (2013). Ben Johnson withdraws as Johns Hopkins commencement speaker after gay marriage outcry. *The Huffington Post*. Available at: http://www.huffingtonpost.com/2013/04/11/ben-carson-with draws-commencement_n_3060052.html. Accessed January 1, 2013.

Rai, T. S., & Fiske, A. P. (2011). Moral psychology is relationship regulation: Moral motives for unity, hierarchy, equality, and proportionality. *Psychological Review*, *118*, 57–75.

Rai, T. S., & Holyoak, K. J. (2010). Moral principles or consumer preferences? Alternative framings of the trolley problem. *Cognitive Science*, *34*, 311–321.

Rehder, B. (2006). When similarity and causality compete in category-based property generalization. *Memory & Cognition*, *34*, 3–16.

Rehder, B. (2009). Causal-based property generalization. *Cognitive Science*, *33*, 301–343.

Valdesolo, P., & DeSteno, D. (2007). Moral hypocrisy: Social groups and the flexibility of virtue. *Psychological Science*, *18*, 689–690.

Valdesolo, P., & Desteno, D. (2008). The duality of virtue: Deconstructing the moral hypocrite. *Journal of Experimental Social Psychology*, *44*, 1334–1338.

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

Waldmann, M. R., Nagel, J., & Wiegmann, A. (2012). Moral judgment. In K. J. Holyoak, & R. G. Morrison (Eds.), *The Oxford handbook of thinking and reasoning* (pp. 364–389). New York: Oxford University Press.