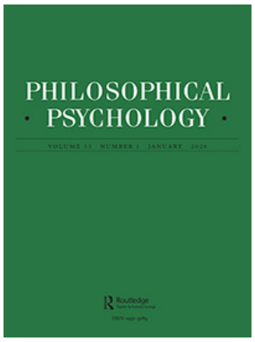

Philosophical Psychology

ISSN: 0951-5089 (Print) 1465-394X (Online) Journal homepage: https://www.tandfonline.com/loi/cphp20

Are the folk historicists about moral responsibility?

Matthew Taylor & Heather M. Maranges

To cite this article: Matthew Taylor & Heather M. Maranges (2020) Are the folk historicists about moral responsibility?, Philosophical Psychology, 33:1, 1-22, DOI: 10.1080/09515089.2019.1695045

To link to this article: https://doi.org/10.1080/09515089.2019.1695045

Published online: 06 Dec 2019.

PHILOSOPHICAL PSYCHOLOGY
2020, VOL. 33, NO. 1, 1–22
https://doi.org/10.1080/09515089.2019.1695045

ARTICLE

Are the folk historicists about moral responsibility?

Matthew Taylor[a] and Heather M. Maranges[b]

[a]Philosophy Department, Florida State University, Tallahassee, USA; [b]Psychology Department, Florida State University, Tallahassee, USA

ABSTRACT

Manipulation cases have figured prominently in philosophical debates about whether moral responsibility is in some sense deeply historical. Meanwhile, some philosophers have thought that folk thinking about manipulated agents may shed some light on the various argumentative burdens facing participants in that debate. This paper argues that folk thinking is, to some extent, historical. Across three experiments, a substantial number of participants did not attribute moral responsibility to agents with manipulation in their histories. The results of these experiments challenge previous research indicating that folk thinking is not historicist. Furthermore, perceptions of reduced free will, but not of a change in personal identity of the agent, account for the attenuation of moral responsibility when the agent was manipulated. To the extent that folk thinking is relevant to philosophical debates about the nature of moral responsibility, the results helpfully illuminate the dialectical burdens facing competing conceptions of responsible agency.

ARTICLE HISTORY
Received 9 April 2018
Accepted 26 February 2019

KEYWORDS
Free will; moral responsibility; compatibilism; experimental philosophy

1. Introduction

There has been substantial philosophical attention to manipulation cases and their relevance to accounts of free and responsible action (Kane, 1996; McKenna, 2012; Mele, 2006, 2008; Pereboom, 2001, 2014; Todd, 2013). In manipulation scenarios, the manipulator M does not force the victim V to act as M wants. Instead, M covertly implants certain desires, values, beliefs, and other mental states in V and then allows V to act upon those implanted mental states in the way that M plans. Some philosophers, incompatibilists, have often appealed to manipulation cases to support their view over compatibilism – the view that it is possible for agents to perform free and responsible actions in deterministic universes (Pereboom, 2014; Todd, 2013).[1] Manipulation cases have also figured prominently in debates between advocates of different views of compatibilism (McKenna, 2012; Mele, 2008). That debate concerns whether moral responsibility is in some sense deeply historical.[2]

CONTACT Matthew Taylor mct15@my.fsu.edu; mattaylo87@gmail.com Philosophy Department, Florida State University, Tallahassee, USA

This article has been republished with minor changes. These changes do not impact the academic content of the article.

According to historical compatibilists, it is possible to have two agents that are near psychological duplicates – relevantly similar with respect to their mental contents (e.g., desires, values, beliefs, etc.) – yet differ with respect to their status as free and responsible agents as a consequence of facts about their individual histories. Non-historical compatibilists deny this claim. Historicists have argued against non-historicism by appealing to specific sorts of manipulation cases: agents in some manipulation cases are not morally responsible because of what happened to them in the past (e.g., they were victims of heavy duty value engineering; Mele, 2008).[3] If this is correct, then non-historicism begins to look increasingly untenable in light of manipulation cases (more on this below).

2. Method: Lay intuitions

Meanwhile, some philosophers have thought that empirical findings on lay thinking about manipulation cases can be used to settle some debates between competing conceptions of free and responsible action (Björnsson, 2016; Sripada, 2012). This methodology might strike some philosophers as exceedingly unpromising. What relevance could the sentiments of non-philosophers have about nuanced and highly sophisticated philosophical positions in the metaphysics of free will and moral responsibility? It is unlikely that these experimental subjects have the ability to draw important philosophical distinctions about free will and responsibility. Furthermore, why suppose that they can attend to the salient details and form accurate representations of manipulation scenarios? These are all legitimate concerns, and we certainly do not take folk thinking about moral responsibility to be the final word in these debates.

The above concerns do not mean that folk thinking about manipulation is completely uninteresting or irrelevant to philosophical debates. First, there are ways to mitigate the sorts of problems noted above. As other experimental philosophers have pointed out, there are steps researchers can take to avoid misrepresentations of hypothetical scenarios (Björnsson, 2016, p. 640). For example, experimenters can describe the cases in substantial detail and employ comprehension checks to increase the likelihood that folk are thinking about the hypothetical scenarios in the correct manner. Second, there are reasons for believing that folk thinking is not entirely irrelevant to philosophical theorizing. There is a genuine worry that the philosopher's intuitions about hypothetical scenarios are theoretically motivated. Incompatibilists may be more inclined to focus on features of cases that support their judgments about free will and moral responsibility, whereas compatibilists of varying stripes focus on other features that support their opposing judgments (McKenna, 2008, p. 144; Björnsson, 2016, p. 641). By collecting data from ordinary people, theoretically motivated intuitions or confirmation bias are less likely to occur. Furthermore, folk thinking can have an effect on the degree of dialectical burden shouldered by participants in philosophical debate. The more divergent a philosophical theory is from ordinary thinking, the harder it will be for proponents of that theory to convince others of the truth of their view (Mele, 2013, p. 183). Hence, we think it a worthwhile project to investigate ordinary intuitions about moral responsibility in response to manipulation cases.

The literature on experimental approaches to manipulation cases has focused entirely on the debate between incompatibilism and compatibilism (Björnsson, 2016; Sripada, 2012). This is somewhat surprising, given that manipulation cases have also appeared in debates between historicist and non-historicist forms of compatibilism. In this paper, we aim to correct for this lack of attention: we are interested in whether the folk concept of moral responsibility is historical or non-historical. Sripada (2012) briefly acknowledges historical compatibilist approaches to moral responsibility. However, he claims that "the results of the studies reported in this paper suggest that these extra historical conditions are actually unnecessary [for an agent to perform free and responsible action]" (Sripada, 2012, p. 567). We disagree. Sripada has only conducted experiments on one kind of manipulation case (i.e., indoctrination cases), and he has not tested the kinds of manipulation cases that historicists have typically appealed to in support of their view. In that case, Sripada has been too quick to dismiss the plausibility of historical conditions on moral responsibility. We maintain that historical compatibilism deserves more rigorous testing than has been conducted thus far. Our aim is to examine whether folk intuitions about moral responsibility are sensitive to the histories of agents in manipulation cases presented by historicists. Only then would we be justified in drawing the same sorts of non-historicist conclusions as Sripada. Our thesis in this paper is that – *pace* Sripada – folk thinking about moral responsibility is, to some extent, historical.

3. Non-historicism vs historicism

It will be helpful to present some preliminary definitions before proceeding. We use the term 'moral responsibility' to refer to an agent's deserving of moral praise and blame for performing good and bad intentional actions, respectively. We also follow Derk Pereboom (2014) in using 'moral responsibility' in the basic desert-entailing sense of the term – as opposed to other senses of the term that involve consequentialist or contractual considerations.

The aim of this section is to briefly consider the debate between historicists and non-historicists. We focus on Harry Frankfurt's (2002) non-historicist view in particular. According to Frankfurt, it does not matter (with respect to moral responsibility) how a moral agent became the way that he is with respect to his mental contents – so long as he is in a particular psychological condition, namely that he identifies with or reflectively endorses his desires, values, and other mental states:

> If someone does something because he wants to do it, and if he has no reservations about that desire but is wholeheartedly behind it, then – so far as his moral responsibility for doing it is concerned – it really does not matter how he got that way. (p. 27)

Frankfurt's view of moral responsibility is not historical – all that matters is whether the agent's psychological condition satisfies the endorsement criteria presented by Frankfurt. For any two agents that are near psychological duplicates, if one of them satisfies Frankfurt's conditions, then the other must as well – and they cannot differ with respect to their status as morally responsible agents – *pace* historicists.

Al Mele (2006, 2008, 2013)) has provided a compelling argument against Frankfurt's non-historicist view. Mele presents a pair of cases intended to elicit intuitive responses that conflict with the Frankfurt's account of moral responsibility. We henceforth follow Mele in using the term 'intuition' to refer to our pre-theoretical inclinations to form beliefs or judgments about particular cases. The pair of examples involve moral agents that are near psychological duplicates, yet have come to acquire their current psychologies in radically different ways. The first example involves an evil man named Chuck who expended great effort to become a thoroughly vicious agent who enjoys killing others. In the second example, we are asked to imagine a woman named Beth who expended effort to become an extremely kind person. When Beth goes to sleep, a team of psychologists remove Beth's values, character, and desires, and replace them with Chuck's psychological attributes. When Beth awakens, she notices a desire to kill her neighbor George. She reflects on this new desire and is wholeheartedly behind it. Beth satisfies Frankfurt's conditions for moral responsibility when she kills her neighbor George. Mele reports that, despite the fact that Beth and Chuck are near psychological duplicates when they kill people, he is inclined to blame Chuck but not Beth. This kind of manipulation case has been taken to challenge Frankfurt's account, as well as any other non-historicist view that neglects the importance of an agent's history.

How might non-historicists respond to Mele's challenge? The non-historicist may respond that she does not share Mele's intuition about Beth, and attempt to convince historicists that they are making an error. Mele (2013) identifies two possible persuasive strategies:

> People who differ in their intuitions (or lack thereof) about Beth's killing George may try to persuade each other that they are making an error. One strategy features exploring additional cases. Another is to use the methods of experimental philosophy. If it were to turn out that a great majority of lay folk strongly agree with the assertion that Beth does not – or does – deserve to be blamed for killing George, that might

> sway the other side to some degree. But these are topics for another occasion. (pp. 172–173)

Some defenders of non-historicism have taken up the first strategy (McKenna, 2012). The purpose of this paper is to explore the second strategy identified by Mele. We will investigate the extent to which ordinary people attribute moral responsibility to agents manipulated in the ways described in Mele's examples.

4. Previous research on moral responsibility

While there is currently no systematic evidence on lay responses to manipulation cases used to support historicism over non-historicism, the social and moral psychological literature provides important insights. Judgments about responsibility and blame are driven by the following psychological mechanisms: (M1) tracking the agent's desire to act, (M2) belief that the act will result in the desired outcome, (M3) the agent's intention to act, and (M4) the causal connection between the act and the relevant outcome (e.g., Alicke, 2000; Knobe, 2003; Malle & Knobe, 1997; Pizarro, Uhlman, & Bloom, 2003; Pizarro, Uhlman, & Salovey, 2003). The agent's current moral character provides insight into the likelihood that he or she desires, believes, intends, and acts now and in the future. Thus, unsurprisingly, people care deeply about the moral character of others, and a myriad of work demonstrates the centrality of moral character in person perception (e.g., Brambilla & Leach, 2014; Goodwin, Piazza, & Rozin, 2014). The cases we employ in our experiments describe agents who are identical in the features which elicit the psychological mechanisms listed above.

Additional work has examined the role personal histories play in lay people's moral perception, responsibility ratings, and blame judgments. Work by Nadler and colleagues (e.g., Nadler, 2012; Nadler & McDonnell, 2011) focuses on the history of the moral agent's character and their patterns of behavior prior to the relevant immoral act. A history of immoral conduct in the past amplifies responsibility and blame judgments of the agent's act, for example. In other work, Alicke and Davis (1990) found that non-external manipulations such as epileptic seizures and psychotic delusions attenuate responsibility and blame judgments of bad actors.

Gill and Cerce's (2017) work assessed the effect of developmental historicist narratives (i.e., explanations for the upbringing of agents that explains why they are the way they are) in responsibility and blame judgments. They theorized and empirically confirmed that that lay people have two conceptions of free will that attenuate judgments of moral responsibility in light of some aspect of an agent's history that may explain present behavior. *Freedom of action* captures the ability to do otherwise, the unconstrained volitional will, or the choice making mechanism at the time of action. *Freedom of self-formation* tracks the locus of moral character control "within the protracted period over which the enduring character of the wrongdoer was forged," and concerns whether the agent is the genuine originator of her own morally defective character (Gill & Cerce, 2017, p. 363).

The present empirical investigation differs from this work in that we use manipulation cases to test whether the folk have intuitions of the historicist flavor in other kinds of scenarios. Manipulation cases are temporally different than Gill and Cerce (2017) prolonged developmental history, in that the former's relevant historical event is discrete. We expect that the present manipulation cases will evoke intuitions that the agent's capacity for freedom of action is constrained, whereas it is an open question whether they will evoke concerns about the freedom of self-formation, given the briefness of character manipulation.

5. Experiments

The aim of this section is to report empirical evidence gathered on folk thinking about manipulation and moral responsibility. Study 1 tests whether people's moral intuitions respond to manipulation cases in the way historicists suggest they do. Study 2 tests whether the folk's attributions of moral responsibility and blame are driven by the relevant historical details. Finally, Study 3 improves upon the design of the first two studies by employing vignettes designed to avoid some potential confounds. Study 3 also seeks to: (a) rule out the possibility that perceptions of numerical identities are driving effects; and (b) provide a mechanistic account of lay people's subscription to historicism, namely whether freedom of self-formation or freedom of action mediate the relationship between manipulation (vs. no manipulation) and moral responsibility.

5.1. Study 1

The aim of the first study was to test whether folk intuitions about moral responsibility mimicked those of historical compatibilists. Subjects (final N = 212; 101 women, 109 men, 2 undisclosed; age range 20 to 83 years, mean age = 39, SD = 13.19) were recruited using Amazon's Mechanical Turk, an online service connecting survey takers to researchers. Each vignette was quite long, and so we have reported these in their entirety in the Appendix.

Subjects in the "manipulation" group read a story about a woman named Sally who was a nice person as a result of her own effort. Everyone in Sally's community believes that she is a nice person. Unfortunately, a team of scientists break into Sally's house at night and remove all of her nice values, desires, and so on. They then implant vicious character traits, values, and desires (including a desire to kill her neighbor, George). The scientists do this because they want George killed. When Sally wakes up, she reflects on her new desires. Given her newly implanted values, she comes to the conclusion that morality is for weaklings, and she is fully behind her desire to kill George. She kills George when she knows she can get away with it. The scientists revisit Sally the night after the killing and reverse the manipulation procedure – they return all of her old values and desires, and remove the ones they implanted.

Subjects in the "no manipulation" group read a vignette about a woman named Sally who is fully behind her murderous desires. She works hard to become a cruel person and succeeds in this task. Everyone in Sally's community (falsely) believes that she is a good person. One night, a team of scientists wanted George killed. They wrongly believed that Sally was a nice person and designed an elaborate plan to attempt to implant murderous values in Sally after erasing her original values. Before they perform this procedure, they learn that Sally already has murderous values (including the desire to murder George), and so they never execute their plan. They do not implant any values in Sally nor do they erase her already held values. They simply allow Sally to kill George herself. The next day, Sally stalks and kills George.

All participants were then asked to indicate the extent to which they agreed with various statements about Sally, on a scale of 1 to 7 (Strongly Disagree/Strongly Agree). We were primarily interested in their responses to the following statements:

> Sally is morally responsible for killing George.
>
> [(Strongly Disagree – 1) (Strongly Agree – 7)]
>
> (2) Sally deserves to be punished for killing George.
>
> [(Strongly Disagree – 1) (Strongly Agree – 7)]

Subjects completed the questionnaire by reporting various demographics and received debriefing.

Results showed that the manipulation-vignettes produced the expected effects on subjects' intuitions about Sally's moral responsibility and basic desert. Participants were significantly less likely to agree that Sally was morally responsible and deserving of punishment in response to the manipulation scenario when compared to the "no manipulation" group. That is, the "no manipulation" group attributed much more moral responsibility to Sally than did the "manipulation" group, $F(1, 210) = 189.03$, $p < .001$, $\eta_p{}^2 = .48$. In the "no manipulation" group, 96.1% of subject agreed with the statement that Sally was morally responsible for killing George.[4] In the "manipulation" group, 32.7% of subjects agreed with the statement that Sally was morally responsible for killing George.[5] Likewise, there was a significant effect of manipulation condition such that the “no manipulation” group agreed that Sally deserves punishment much more than the “manipulation” group, $F(1, 210) = 191.71$, $p < .001$, $\eta_p{}^2 = .48$. In the “no manipulation” group, 94.2% of subject agreed with the statement that Sally deserved punishment for killing George[6], whereas in the “manipulation” group, 33.3% of subjects agreed with that statement.[7]

The women named Sally were meant to be described as near psychological duplicates of each other across all vignettes – qualitatively similar with respect to (among other things) valuing cruelty and acting on their reflectively endorsed desire to kill George. Despite the fact that the agents across the vignettes are near psychological duplicates during the killing, subjects were significantly more likely to disagree with attributions of Sally’s responsibility in the vignettes involving heavy-duty value engineering in her past. This pattern of responses mimics the pattern reported by historicists in response to these manipulation cases.

Someone might worry that it remains somewhat unclear whether historical features or non-historical features are driving the intuitions in these cases. Perhaps participants are (wrongly) thinking that there is some other non-historical feature present in the manipulation vignette that is not present in the non-manipulation scenario. Another limitation was that all participants were given both a “moral responsibility” and “deserving punishment” item. However, the presence of one of these items may have influenced their responses to the other. Finally, we did not ensure that participants attended to the most important details of each vignette. One way of addressing this issue is to present comprehension checks to ensure that subjects are forming the right kind of mental representations of the scenarios presented. We corrected for these issues in the next study.

5.2. Study 2

Subjects (final $N = 250$; 131 women, 118 men, 1 undisclosed; age range 20 to 81 years, mean age = 38, $SD = 12.30$) were again recruited using Amazon’s Mechanical Turk. The second experiment was a 2 × 2 design – subjects were first presented with either a “manipulation” scenario or “no manipulation” scenario, and they were then asked to report the extent to which they agreed with statements that involved either attributions of moral responsibility or attributions of deservingness of moral blame. We updated the vignettes to improve on the limitations of Study 1 and to correct for possible misinterpretations from Sripada (2012) (see Appendix). All subjects were presented with comprehension checks after reading each vignette:

Did psychologists implant values and desires in Sally? [YES/NO]

Does Sally understand morality and the difference between right and wrong? [YES/NO]

Is Sally fully behind her desire to kill George? [YES/NO]

Subjects completed the comprehension check and then were instructed to indicate the extent to which they agreed with a statement about Sally (on a Likert scale from 1 to 7). The worry in the previous study was that subjects' intuitions may have been sensitive to some other non-historical feature of Sally during the killing, despite our best attempts at describing each Sally as near psychological duplicates. In this study, we connected judgments about Sally's moral responsibility to historical facts about her. Subjects in the "moral responsibility" group read the following statement:

> In light of what happened to Sally the night before, Sally is not morally responsible for killing George.
>
> [(Strongly Disagree – 1) (Strongly Agree – 7)]

Subjects in the "basic desert" group read the following statement instead:

> In light of what happened to Sally the night before, Sally does not deserve to be blamed for killing George.
>
> [(Strongly Disagree – 1) (Strongly Agree – 7)]

These two items constituted the dependent variable in the experiment – we expected people in the "manipulation" group to agree more than the "no manipulation" group. Subjects then completed an attention check procedure (see Appendix), reported various demographics, and were debriefed.

Results indicated that the manipulation vignettes produced the expected effects on subjects' intuitions about Sally's moral responsibility and basic desert. Note that we reverse coded Dependent Variable (DV) responses so that higher values reflect agreement with Sally's being morally responsible and blameworthy, making results comparable to Study 1. Replicating Study 1, participants were significantly less likely to agree that Sally was morally responsible or blameworthy in response to the "manipulation" scenario when compared to the "no manipulation" group. That is, there was a significant effect of condition on the extent to which participants ascribed moral responsibility or blame to Sally such that the "no manipulation" group held Sally much more responsible or blameworthy than the "manipulation" group, $F(3, 246) = 153.76, p < .001, \eta_p{}^2 = .39$. In the "no manipulation" group, 94.1% of subjects agreed with the statement that Sally was morally responsible or blameworthy for killing George.[8] In the "manipulation" group, 46.5% of subjects agreed with the statement that Sally was morally responsible or blameworthy for killing George, whereas 44% disagreed with that statement in light of facts about Sally's history.[9]

There was no effect of whether participants responded to a question about moral responsibility or deservingness of moral blame, $F(3, 246) = 2.75$, $p = .31$, $\eta_p{}^2 = .004$. Collapsed across "manipulation" groups, 73.4% of participants agreed that Sally was morally responsible for killing George and 70.5% of participants agreed that Sally was blameworthy for killing George. There was no interaction between manipulation and response question (i.e., moral responsibility or basic desert), $F(3, 246) < 0.001$, $p = .99$, $\eta_p{}^2 < .001$. These preliminary results suggest that a substantial minority agree that an agent's history can excuse their actions.

Someone might complain that subjects may have been non-historicists but withheld judgments of moral responsibility in manipulation scenarios because of a severance of Sally's personal identity. Subjects may have said that Sally is not morally responsible for killing George because the manipulation resulted in someone else killing George – the murderer was not Sally. Perhaps the folk think that there is an important connection between personal identity and the agent's values and character traits. There is some research showing that radical deteriorations in an agent's character (from good to bad) result in ordinary people disagreeing with statements that the agent (after the deterioration) is the same person (Strohminger & Nichols, 2014; Tobia, 2015). Does this evidence support the competing hypothesis about personal identity? Admittedly, Sally does undergo a very severe form of deterioration in the manipulation case.

We maintain that it is unclear whether experimenters in the above studies tested for the kind of "numerical identity" that would support such a hypothesis. Statements about whether an agent is the "same person," could be eliciting folk judgments about identity in the self-concept sense of the term used in psychology, as opposed to the numerical identity the sense of the term in philosophy (Descombes, 2016; Schwenkler, 2017[10]).[11] One way to tease apart these different senses of identity would be to ask about personal identity more indirectly. For example, experimenters could rework the manipulation scenario and add in the following detail: Sally donated money to Oxfam when she was younger. They could then ask the subjects to indicate the extent to which they agree with the following statement: "the woman who killed George donated to Oxfam when she was younger." If the folk tend to disagree with this statement to a significant extent, then we are in trouble. This result would be evidence that manipulation cases of the sort we are presenting produce severances in personal identity, and non-historicists can explain away our results that appear to favor historicism (to some extent).[12] In the next experiment, we rule out this competing hypothesis.

5.3. Study 3

Studies 1 and 2 provide evidence that external manipulation attenuates attributions of moral responsibility and blame. Nevertheless, stronger support for our conclusions can be drawn by ruling out potential confounds.[13] The previous experiments did not control for potential forward-looking considerations (e.g., Sally's future character) that may have been driving participants' intuitions. A crucial difference between the control and experimental groups is that Sally remains deeply vicious and evil in the control group (but her character is improved in experimental groups). This forward-looking difference may have amplified moral responsibility judgments in the control group in a misleading way. Given that we are focused on backward-looking basic desert, we decided to add the following detail in the control group: the psychologists visit Sally after the murder and change her character from bad to good. We ruled out any influences of forward-looking considerations by making Sally's future character and conduct identical across groups.

We also investigated two potential mechanisms driving the split in folk intuitions here. First, the non-historicist might argue that the results from experiment 2 are inconclusive because subjects may have judged that Sally does not exist after the manipulation, and so concluded that Sally is not responsible for the murder. Second, it could be that people's concerns with Sally's free will are pushing their intuitions one way or the other. Gill and Cerce (2017) present two conceptions of free will that lay people may track in manipulation cases: *freedom of action* and *freedom of self-formation*. Hence, we test the role of both senses of personal freedom in the current study. The "manipulation" group read the following vignette:

> When Sally crawled into bed last night, she was one of the kindest, most gentle people on Earth. She was not always that way, however. She worked hard to change her character, and she succeeded. Before Sally's efforts to change her character, she donated money to Oxfam with the goal of helping others.
>
> Sally does something awful the next day. She awakens with a desire to stalk and kill a neighbor, George. What happened is that, while Sally slept, a team of psychologists implanted murderous values in Sally after erasing her original values. They removed all of her original character traits also – so that she was no longer kind and gentle. Their procedure included implanting a desire to murder George the next day.
>
> Sally is still like anyone else in many respects. Sally understands morality, the difference between right and wrong, and various ways she might conduct her life. Given her current values, she currently does not want to live a moral life because she views morality as a system for weaklings. Additionally, Sally was not simply fed lies about George – she knows the truth about who he is and she knows exactly why she dislikes him. Sally is not a robot who simply does as others instruct. Nor is she under the grip of an irresistible impulse. Rather, Sally is a person with desires, values, hopes, and dreams, just like anyone else. But Sally's desires include killing George. And her core values and character traits recommend killing George.
>
> The desire to kill George is not in conflict with any of her other preferences – it is well integrated with her other desires. Sally reflects on her new desire to kill George.

> Among other things, she thinks that this new desire does not conflict with her new system of values. Upon reflection, Sally has no reservations about her desire to kill George and is wholeheartedly behind it. Sally devises a plan for killing him, and she executes it – and him – that afternoon, once she is confident that the killing would go undetected. Sally slit George's throat. She likes her values and she kills George because she wants to do it.
>
> When Sally falls asleep after her cruel deed, the team of psychologists change her character and values so that Sally is extremely kind and incapable of harming others. The psychologists only wanted George dead, and so – after the murder – they change Sally in the relevant ways so that she is no longer a threat to others. Sally goes on to live a moral life.

The *"no manipulation"* group read the following vignette:

> When Sally crawled into bed last night, she was one of the meanest, most vicious people on Earth. She was not always that way, however. She worked hard to change her character, and she succeeded. Before Sally's efforts to change her character, she donated money to Oxfam with the goal of helping others.
>
> Sally does something awful the next day. Sally awakens with a desire to stalk and kill a neighbor, George. Last night, while Sally slept, a team of psychologists decided that they wanted George dead. They wrongly believed that Sally was a nice person, and designed an elaborate plan to implant murderous values in Sally after erasing her original character and values. Before they perform this implantation procedure, they learn that Sally already has vicious character traits and bad values – including the desire to murder George the next day – and so they never execute their plan. They do not implant any values, character traits, or desires in Sally, nor do they erase her already held values or character traits. They simply allow Sally to kill George without their intervention.
>
> Sally is still like anyone else in many respects. Sally understands morality, the difference between right and wrong, and various ways she might conduct her life. Given her current values, she currently does not want to live a moral life because she views morality as a system for weaklings. Additionally, Sally was not simply fed lies about George – she knows the truth about who he is and she knows exactly why she dislikes him. Sally is not a robot who simply does as others instruct. Nor is she under the grip of an irresistible impulse. Rather, Sally is a person with desires, values, hopes, and dreams, just like anyone else. But Sally's desires include killing George. And her core values and character traits recommend killing George.
>
> The desire to kill George is not in conflict with any of her other preferences – it is well integrated with her other desires. Sally reflects on her new desire to kill George. Among other things, she thinks that this new desire does not conflict with her system of values. Upon reflection, Sally has no reservations about her desire to kill George and is wholeheartedly behind it. Sally devises a plan for killing him, and she executes it – and him – that afternoon, once she is confident that the killing would go undetected. Sally slit George's throat. She likes her values and she kills George because she wants to do it.

> When Sally falls asleep after her cruel deed, the team of psychologists change her character and values so that Sally is extremely kind and incapable of harming others. The psychologists only wanted George dead, and so – after the murder – they change Sally in the relevant ways so that she is no longer a threat to others. Sally goes on to live a moral life.

After reading the vignette, responding to comprehension and attention check questions, and either the moral blame or responsibility items, all participants responded to the following item designed to unobtrusively measure judgments of personal identity:

> The woman who killed George donated to Oxfam when she was younger.
>
> [(Strongly Disagree – 1) (Strongly Agree – 7)]

Agreement with the statement reflects the belief that Sally is the same person throughout the entire story. If subjects disagree with the above statement, then they appear to think that Sally – the woman who donated to Oxfam earlier in life – no longer exists after the manipulation procedure.

We adapted each of the three freedom of action and freedom of self-formation items from Gill and Cerce (2017) work to reflect the details of Sally's case. These items were included in order to test whether folk conceptions of free will account for participants' intuitions about moral responsibility. Participants rated the extent to which they agreed with the following statements:

> Freedom of action:
>
> By using her capacity for free will, Sally could choose to STOP herself from killing George.
>
> [(Strongly Disagree – 1) (Strongly Agree – 7)]
>
> It is possible for Sally to use her free will to CHOOSE to behave differently.
>
> [(Strongly Disagree – 1) (Strongly Agree – 7)]
>
> Although Sally might have a strong inclination to kill George, she can use her capacity for free will to act differently.
>
> [(Strongly Disagree – 1) (Strongly Agree – 7)]
>
> Freedom of self-formation:
>
> Sally had free will in terms of initially becoming the type of person she is.
>
> [(Strongly Disagree – 1) (Strongly Agree – 7)]
>
> Throughout her life, Sally was always in control of her personality development.
>
> [(Strongly Disagree – 1) (Strongly Agree – 7)]

> Sally's killing George is purely a result of her freely choosing to become who she currently is.
>
> [(Strongly Disagree – 1) (Strongly Agree – 7)]

We collected data from 338 people. Before analyses, data were cleaned in the following way: participants who failed one or more of the comprehension questions (n = 87) or the attention check (n = 16; 9 of whom also failed the comprehension checks) were removed for analyses, leaving a final sample of 244 (125 females, 86 men, 1 unreported; age range 18 to 78, mean age = 39.59, SD = 13.71). For our analyses, we first tested whether Study 3 replicated the findings from Studies 1 and 2. Results indicated that the manipulation vignettes produced the expected effects on subjects' intuitions about Sally's moral responsibility and basic desert: there was a significant effect of condition on the extent to which participants ascribed moral responsibility or blame to Sally such that the "no manipulation" group held Sally much more responsible or blameworthy than the "manipulation" group, F(1,108) = 218.82, p < .001, η_p^2 = .475. In the "no manipulation" group, 93.1% of subjects[14] agreed with the statement that Sally was morally responsible or blameworthy for killing George. In the "manipulation" group, 29.5% of subjects agreed, whereas 58.3% excused Sally in light of facts about her history[15].

We also tested whether participants' personal identity responses differed between the "manipulation" and "no manipulation" groups. The groups did not differ in their agreement that Sally was the same person over time, F(1, 242) = .13, p = .723, η_p^2 = .001: *no manipulation* (M = 6.17, SD = 1.40), *manipulation* (M = 6.23, SD = 1.42). We then tested whether responses to the personal identity item predicted responsibility and blame judgments. Given the continuous nature of both the predictor and outcome variables, we used regression. Responses to the personal identity question did not predict responsibility and blame judgments, b = .08, SE = .10, t(243) = .74, p = .462. The null results here preclude personal identity as a reasonable mediator between the manipulation cases and moral responsibility and blame responses. The current findings challenge any non-historicist attempt to explain away our findings by appeal to severances of personal identity.

Finally, we tested whether freedom of action and a backward-looking freedom of self-formation mediate the relationship between condition and responsibility and blame judgments. Our initial analysis focused on whether these freedom ratings differed between the groups. Groups differed with respect to ratings on freedom of action, F(1, 242) = 56.79, p < .001, η_p^2 = .19, and freedom of formation, F(1, 242) = 47.08, p < .001, η_p^2 = .16. The "manipulation" group rated Sally as having much less freedom of action (M = 4.68, SD = 1.96) and freedom of formation (M = 4.36, SD = 1.25) relative to the "no manipulation" group (*action:* M = 6.20, SD = 1.12; *formation:* M = 5.49, SD = 1.32)

We then tested whether these ratings were associated with responsibility and blame judgments using regression analyses. Freedom of action ratings predicted attributions of responsibility and blame, $b = .83$, $SE = .06$, $t(242) = 12.89$, $p < .001$. Likewise, freedom of formation ratings predicted attributions of responsibility and blame, $b = .89$, $SE = .09$, $t(242) = 10.20$, $p < .001$.

Our final analysis focused on whether differences between groups' intuitions about Sally's freedom to act or freedom to develop as she willed accounted for the effect of condition (manipulation vs. no manipulation) on responsibility and blame judgments. We conducted a 10,000 bootstrapping resample mediation analyses using Model 4 in the PROCESS Macro with freedom of action and freedom of formation ratings entered as simultaneous mediators (Preacher & Hayes, 2004). Condition (manipulation vs. no manipulation) predicted decreased responsibility and blame attributions through perceptions of both lower freedom of action, $b = -.66$, $SE = .15$, CI_{95} [−.998, −.418], and freedom of formation, $b = -.332$, $SE = .12$, CI_{95} [−.608, −.137]. That is, the "manipulation" group viewed Sally as more constrained in her ability to stop herself from killing George and in having had the opportunity to develop morally, and therefore as less responsible and blameworthy for his murder compared to the "no manipulation" group.

This research adds to a growing body of literature in philosophy and moral psychology suggesting that history matters in judgments of free will and moral responsibility. In our work, the agent's history matters when it involves heavy-duty value engineering, whereas prior research discovered that history matters with respect to developmental inputs (e.g., Gill & Cerce, 2017); but in both experimental designs, ordinary perceptions that an agent's free will is constrained by their history attenuates responsibility and blame judgments. Building on Gill and Cerce (2017) work, we find that both lay conceptions of free will come into play when people consider a manipulation case. A perceived reduction in freedom of self-formation, or opportunity for the agent to form her own moral character, accounts for the effect of manipulation on reduced blame. In contrast to Gill and Cerce (2017) findings, freedom of action is also seen as reduced in the case of manipulated Sally, and therefore she is judged as less morally responsible and blameworthy.

There are also important philosophical implications to be drawn from the present studies. Recall that, according to Sripada (2012), two experiments on folk thinking about moral responsibility suggested that "extra historical conditions are actually unnecessary [for agents to perform free and responsible actions]" (Sripada, 2012, p. 567). Sripada's claim about folk thinking is inconsistent with the results reported in this paper. In Study 2, a substantial minority (i.e., 44% of subjects) agreed with the statement excusing Sally in light of her agential history. While it is also true that a substantial minority (i.e., 46.5% of subjects) also disagreed with the historical claim, the folk appeared to be mostly split on the truth of historicism. However, even more damaging findings emerged in Study 3, where most of the participants (i.e., 58%) agreed that Sally is not morally responsible for killing George in light of what happened to her the night before. Given that the folk do not overwhelmingly come down firmly in favor of non-historicism, Sripada's claim that folk thinking about moral responsibility is non-historicist is challenged by the present evidence.

Another philosophical implication is this: Mele (2013) has suggested that non-historicists might employ experimental philosophy to try to persuade historicists of the truth of their view by showing that most ordinary people agree with them about manipulation cases. If non-historicists were hoping to show an error in their opponents thinking by appeal to experimental results on folk thinking, they will be met with great disappointment. The results indicated by the three studies suggest that many people do not agree with non-historicist views of moral responsibility.

Suppose that historicists are correct in claiming that the cost of accepting non-historicism is roughly the time and effort it would take to convince others of the truth of their view (Mele, 2013, p. 183). It would seem that non-historicists face the burden of producing arguments of sufficient strength to convince a substantial portion of the populace (i.e., 44 to 58%) of the truth of non-historicism. Historicists, on the other hand, face the burden of producing arguments of sufficient strength to convince a different portion of the populace (i.e., 28 to 46%) in manipulation scenarios. While some philosophers have tried to make the costs of historicism explicit by appealing to other hypothetical cases or arguments (McKenna, 2012), no one has appealed to manipulation cases that purportedly support historicism to illuminate the potential costs of accepting a historical condition on moral responsibility. The present research helpfully illuminates the dialectical burden facing proponents of both camps in the debate about historicism.

Notes

1. For the purposes of this paper we will understand *determinism* as "the thesis that there is at every moment exactly one physically possible future" (van Inwagen, 1983, p. 3).
2. Consider the following historical concept 'sunburn.' Suppose there are two people with skin burns that are qualitatively indistinguishable. Now imagine that one of these burns was caused by exposure to the sun, whereas the other was produced by a heating lamp. We would only call the former a sunburn because it must be connected to some causal history that traces back in time to an event involving the person's exposure to the sun. Historicists claim that the concept moral responsibility is similar to sunburns with respect to historical properties.
3. McKenna (2012) points out that the important dispute between historicists and non-historicists concerns whether there is any historical criterion for directly free (and

responsible) action. Some acts gain their status as free (or responsible) in virtue of a direct exercise of certain abilities by the agent at that time. Other acts gain their status as free (or responsible) in virtue of the proper kinds of causal relations to prior direct exercises. The former are identified as directly free and responsible acts, whereas the latter are indirectly free and responsible acts.

4. In the “no manipulation” group, 2.9% disagreed, and 1% neither agreed nor disagreed with attributions of Sally’s moral responsibility.
5. In the “manipulation” group, 58% disagreed, and 9.3% neither agreed nor disagreed with attributions of Sally’s moral responsibility.
6. In the “no manipulation” group, 3.9% disagreed, and 1.9% neither agreed nor disagreed with Sally’s deservingness of punishment.
7. In the “manipulation” group, 53.7% disagreed, and 13% neither agreed nor disagreed with Sally’s deservingness of punishment.
8. In the “no manipulation” condition, 2.2% neither agreed nor disagreed, and 3.7% disagreed with Sally’s status as morally responsible in light of her history.
9. In the “manipulation” condition, 9.5% neither agreed nor disagreed.
10. Schwenkler, J. (2017). *How Do the Folk Think of Seeing?* Unpublished manuscript.
11. Roy Baumeister (1999) offers the following definition of self-concept: “the individual’s belief about himself or herself, including the person’s attributes and who and what the self is” (p. 247). Given this definition, we might think that agents that undergo radical changes are not the same person in the self-concept sense, because they no longer have the same attributes that make up who and what the self is.
12. Another point worth mentioning is that it appears that only deterioration in one’s character or values severs the link of numerical identity. Subjects that read vignettes involving improvements in one’s character or value in the experimental studies stated above claimed that the agent was the same person. We suspect that if subjects are told a similar story about no manipulation (cruel) Sally who had nice values implanted via the psychologists and subsequently performed good acts (e.g., giving to charity), we would not be inclined to praise her for what she does. If the evidence on personal identity is correct, most people would still claim that Sally is the same person (because her character improved).
13. Someone might complain that there remain crucial differences between Sally across groups with respect to her past character. Sally exerts effort to mold herself into a good person in the experimental group, whereas Sally becomes a bad person in the control group. The current literature suggests that prior intentions (Pizarro et al., 2003), prior effort (Bigman & Tamir, 2016), and prior character (Nadler, 2012; Nadler & McDonnell, 2011) matter in responsibility and blame attributions. From an experimental control perspective, it would benefit our design to make equal the agents’ histories on these dimensions. However, controlling for these influences was not possible given the philosophical scenarios focused on in this paper (viz., the examples historicists have presented as evidence against non-historicism). We needed to test intuitions in response to cases where a manipulated agent’s character was radically reversed (from good to bad) by psychologists, and the manipulation resulted in their being a near psychological duplicate of the non-manipulated agent described in the control group. Thus, the radical reversal examples required differences in past intentional action and direction of character change so that the agents described in both control and experimental groups could be near psychological duplicates after the radical reversal manipulation. The details about an agent’s past efforts matter for moral responsibility according to historicists but not according to their opponents. The non-historicist will deny that these past efforts matter with respect to an agent’s

moral responsibility. Keeping these historical differences in our stories allows us to test for whether folk thinking is, to some extent, historicist.

14. In the "no manipulation" group, 4.7% of participants rated her as not responsible and 2.3% were undecided.
15. In the "manipulation" group, 12.2% of participants were undecided.

Acknowledgments

We would like to thank Alfred R. Mele, Ryan M. McManus, and two anonymous reviewers at this journal for their helpful feedback on this project and manuscript.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Matthew Taylor http://orcid.org/0000-0002-4050-7290

References

Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological Bulletin*, *126*(4), 556-574. doi: 10.1037/0033-2909.126.4.556

Alicke, M. D., & Davis, T. L. (1990). Capacity responsibility in social evaluation. *Personality and Social Psychology Bulletin*, *16*, 465–474.

Baumeister, R. F. (1999). Self-concept, self-esteem, and identity. In V. J. Derlega, B. A. Winstead, & W. H. Jones (Eds.), *Nelson-hall series in psychology. Personality: Contemporary theory and research* (pp. 339–375). Chicago, IL, US: Nelson-Hall Publishers.

Bigman, Y. E., & Tamir, M. (2016). The road to heaven is paved with effort: Perceived effort amplifies moral judgment. *Journal of Experimental Psychology: General*, *145*, 1654–1669.

Björnsson, G. (2016). Outsourcing the deep self: Deep self discordance does not explain away intuitions in manipulation arguments. *Philosophical Psychology*, *29*, 637–653.

Brambilla, M., & Leach, C. W. (2014). On the importance of being moral: The distinctive role of morality in social judgment. *Social Cognition*, *32*, 397–408.

Descombes, V. (2016). *Puzzling Identities*. Trans. S.A. Cambridge, MA: Schwartz. Harvard University Press.

Frankfurt, H. (2002). Reply to John Martin Fischer. In S. Buss & L. Overton (Eds.), *Contours of agency* (pp. 27–31). Cambridge, MA: MIT Press.

Gill, M. J., & Cerce, S. C. (2017). He never willed to have the will he has: Historicist narratives, "civilized" blame, and the need to distinguish two notions of free will. *Journal of Personality and Social Psychology*, *112*, 361–382.

Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology*, *106*, 148–168.

Kane, R. (1996). *The significance of free will*. USA: Oxford University Press.

Knobe, J. (2003). Intentional action in folk psychology: An experimental investigation. *Philosophical Psychology*, *16*, 309–324.

Malle, B. F., & Knobe, J. (1997). The folk concept of intentionality. *Journal of Experimental Social Psychology*, *33*, 101–121.

McKenna, M. (2008). A hard-line reply to Pereboom's four-case manipulation argument. *Philosophy and Phenomenological Research*, *77*, 142–159.

McKenna, M. (2012). Moral responsibility, manipulation arguments, and history: Assessing the resilience of nonhistorical compatibilism. *The Journal of Ethics*, *16*, 145–174.

Mele, A. R. (2006). *Free will and luck*. Oxford: Oxford University Press.

Mele, A. R. (2008). Manipulation, compatibilism, and moral responsibility. *The Journal of Ethics*, *12*, 263–286.

Mele, A. R. (2013). Manipulation, moral responsibility, and bullet biting. *The Journal of Ethics*, *17*, 167–184.

Nadler, J. (2012). Blaming as a social process: The influence of character and moral emotion on blame. *Law and Contemporary Problems*, *75*, 1–31.

Nadler, J., & McDonnell, M. (2011). Moral character, motive, and the psychology of blame. *Cornell Law Review*, *97*, 255–304.

Pereboom, D. (2001). *Living without free will*. Cambridge: Cambridge University Press.

Pereboom, D. (2014). *Free will, agency, and meaning in life*. Oxford: Oxford University Press.

Pizarro, D., Uhlmann, E., & Salovey, P. (2003). Asymmetry in judgments of moral blame and praise: The role of perceived metadesires. *Psychological Science*, *14*, 267–272.

Pizarro, D. A., Uhlmann, E., & Bloom, P. (2003). Causal deviance and the attribution of moral responsibility. *Journal of Experimental Social Psychology*, *39*, 653–660.

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods*, *36*, 717–731.

Schwenkler, J. (2017). How Do the Folk Think of Seeing?

Sripada, C. S. (2012). What makes a manipulated agent unfree? *Philosophy and Phenomenological Research*, *85*, 563–593.

Strohminger, N., & Nichols, S. (2014). The essential moral self. *Cognition*, *131*, 159–171.

Tobia, K. P. (2015). Personal identity and the Phineas Gage effect. *Analysis*, *75*, 396–405.

Todd, P. (2013). Defending (a modified version of) the Zygote argument. *Philosophical Studies*, *164*, 189–203.

van Inwagen, P. (1983). *An essay on free will*. Oxford: Oxford University Press.

Appendix

In Study 1, subjects in the "no manipulation" group read the following vignette:

> Sally enjoys killing people, and she is wholeheartedly behind her murderous desires, which are well integrated with her other desires and values. When she kills, she does so because she wants to do it, and she identifies herself with her vicious values. When she was much younger, Sally enjoyed torturing animals, but she was not wholeheartedly behind this. These activities sometimes caused her to feel guilty, she experienced bouts of squeamishness, and she occasionally considered abandoning animal torture. However, Sally valued being the sort of person who does as she pleases.
>
> She also wholeheartedly rejects conventional morality as a system designed for and by weaklings. She freely set out to ensure that she would be wholeheartedly behind her torturing of animals and related activities, including her merciless bullying of

> vulnerable people, and she was morally responsible for so doing. One strand of her strategy was to perform cruel actions with increased frequency in order to harden herself against feelings of guilt and squeamishness and eventually to extinguish the source of those feelings. Sally strove to ensure that her psyche left no room for mercy. Her strategy worked. One night, while Sally slept, a team of psychologists decided that they wanted George, Sally's neighbor, dead. They wrongly believed that Sally was a nice person, and designed an elaborate plan to attempt to implant murderous values in Sally after erasing her original values. Before they perform this procedure, they learn that Sally already has murderous values (including the desire to murder George) and so they never execute their plan. They do not implant any values in Sally nor do they erase her already held values. They simply allow Sally to kill George herself. The next day, Sally stalked and killed George. She waited until she was confident that the killing would go undetected and then slit George's throat. Sally likes her values and she kills George because she wants to do it.

Subjects in the "manipulation" group in Study 1 read the following vignette:

> When Sally crawled into bed last night, she was one of the kindest, gentlest people on Earth. She was not always that way, however. When she was a teenager, Sally came to view herself, with some justification, as self-centered, petty, and somewhat cruel. She worked hard to improve her character, and she succeeded. When she dozed off, Sally was incapable of doing harm to others: her character left no place for a desire to do such a thing. Moreover, she was morally responsible, at least to a significant extent, for having the character she had. But Sally awakens with a desire to stalk and kill a neighbor, George. Although she had always found George unpleasant, she is very surprised by this desire. What happened is that, while Sally slept, a team of psychologists implanted murderous values in Sally after erasing her original values. They did this while leaving her memory intact, which helps account for her surprise.
>
> Sally reflects on her new desire. Among other things, she thinks that this new desire does not conflict with her newly implanted system of values. She also judges that she finally sees the light about morality – that it is a system designed for and by weaklings. Upon reflection, Sally has no reservations about her desire to kill George and is wholeheartedly behind it. Furthermore, the desire is not in conflict with any of her other preferences – it is well integrated with her other desires. Seeing nothing that she regards as a good reason to refrain from stalking and killing George, provided that she can get away with it, Sally devises a plan for killing him, and she executes it – and him – that afternoon, once she is confident that the killing would go undetected. She slit George's throat. Sally likes her values and she kills George because she wants to do it. The team of psychologists are proud that they have once again succeeded in their task to implant wicked values in innocent people, as they have succeeded five other times on other unwitting victims. When Sally falls asleep at the end of her horrible day, the manipulators reverse everything they had done to her. When she awakes the next day, she is just as sweet as ever.

The main difference between these vignettes and those presented in Study 2 is that there are additional details about Sally (e.g., that she is like everyone else in many respects, that she isn't a robot that simply does as others command), and that there are more similarities between the two vignettes in Study 2.

In Study 2, the "manipulation" group read the following vignette:

> When Sally crawled into bed last night, she was one of the kindest, most gentle people on Earth. She was not always that way, however. She worked hard to improve her character, and she succeeded. When she dozed off, Sally was incapable of doing harm to others: her character left no place for a desire to do such a thing. Everyone in Sally's community believes that she is a nice person. But Sally awakens with a desire to stalk and kill a neighbor, George. Although she had always found George unpleasant, she is very surprised by this desire. What happened is that, while Sally slept, a team of psychologists implanted murderous values in Sally after erasing her original values. They removed all of her original character traits also – so that she was no longer kind and gentle. They did this while leaving her memory intact, which helps account for her surprise.
>
> Sally is still like anyone else in many respects. Sally understands morality, the difference between right and wrong, and various ways she might conduct her life. Given her current values, she currently does not want to live a moral life because she views morality as a system for weaklings. Additionally, Sally was not simply fed lies about George – she knows the truth about who he is and she knows exactly why she dislikes him. Sally is not a robot who simply does as others instruct. Nor is she under the grip of an irresistible impulse. Rather, Sally is a person with desires, values, hopes, and dreams, just like anyone else. But Sally's desires include killing George. And her core values and character traits recommend killing George.

The remainder of the vignette was identical to the last paragraph presented in Study 1. The "no manipulation" group read a strikingly similar vignette:

> When Sally crawled into bed last night, she was one of the meanest, most vicious people on Earth. She was not always that way, however. She worked hard to change her character, and she succeeded in becoming a vicious person. Not only did her strategy work, but she was also very good at hiding her vicious character from others. Everyone in Sally's community believes that she is a nice person. However, when she dozed off last night, Sally was incapable of showing mercy to others: her character left no place for a desire to do such a thing. Sally awakens with a desire to stalk and kill a neighbor, George. Sally's new desire does not surprise her in the least bit. She has a kill list and George is the next victim in her murder spree.
>
> Last night, while Sally slept, a team of psychologists decided that they wanted George dead. They wrongly believed that Sally was a nice person, and designed an elaborate plan to implant murderous values in Sally after erasing her original character and values. Before they perform this implantation procedure, they learn that Sally already has vicious character traits and bad values (including the desire to murder George the next day) and so they never execute their plan. They do not implant any values, character traits, or desires in Sally, nor do they erase her already held values or character traits. They simply allow Sally to kill George without their intervention.
>
> Sally is still like anyone else in many respects. Sally understands morality, the difference between right and wrong, and various ways she might conduct her life. Given her current values, she currently does not want to live a moral life because she views morality as a system for weaklings. Additionally, Sally was not simply fed lies about George – she knows the truth about who he is and she knows exactly why she dislikes him. Sally is not a robot who simply does as others instruct. Nor is she under the grip of an irresistible impulse. Rather, Sally is a person with desires, values, hopes, and dreams,

> just like anyone else. But Sally's desires include killing George. And her core values and character traits recommend killing George.
>
> The desire to kill George is not in conflict with any of her other preferences – it is well integrated with her other desires. Sally reflects on her new desire to kill George. Among other things, she thinks that this new desire does not conflict with her system of values. Upon reflection, Sally has no reservations about her desire to kill George and is fully behind it. Sally devises a plan for killing him, and she executes it – and him – that afternoon, once she is confident that the killing would go undetected. Sally slit George's throat. She likes her values and she kills George because she wants to do it.

All subjects were also expected to complete the same attention check across both studies. Attention checks are especially important to ensure that the online survey taker is paying attention and reading each prompt carefully. Subjects were asked to read the following prompt:

> Most modern theories of decision-making recognize the fact that decisions do not take place in a vacuum. Individual preferences and knowledge, along with situational variables can greatly impact the decision process. In order to facilitate our research on decision making, we are interested in knowing certain facts about you, the decision maker. Specifically, we are interested in whether you actually take the time to read the directions; if not, then some of our manipulations that rely on changes in the instructions will be ineffective. So, in order to demonstrate that you have read the instructions, please ignore the question below, and simply select the highest possible choice. We salute you for reading carefully.

What is the approximate air temperature today (in Fahrenheit) where you currently live?
[Higher than 95]
[86–95]
[76–85]
[66–75]
[56–65]
[46–55]
[36–45]
[26–35]
[Below 26]

The data obtained from subjects that did not pass the attention check were not used in our analysis.