**Commonsense Morality and the Bearable Automaticity of Being[[1]](#footnote-1)\***

Samuel Murraya,b,c and Thomas Nadelhoffer d

a Laboratorio de Emociones y Juicios Morales, Universidad de Los Andes, Bogotá, Colombia

b Philosophy Department, Providence College, Providence, RI

c Neuroscience Program, Providence College, Providence, RI

d Philosophy Department, College of Charleston, Charleston, SC

***Forthcoming in Consciousness & Cognition***

*Abstract*

Some research suggests that moral behavior can be strongly influenced by trivial features of the environment of which we are completely unaware. Philosophers, psychologists, and neuroscientists have argued that these findings undermine our commonsense notions of agency and responsibility, both of which emphasize the role of practical reasoning and conscious deliberation in action. We present the results of four vignette-based studies (*N* = 1,437) designed to investigate how people think about the metaphysical and moral implications of scientific findings that reveal our susceptibility to automaticity and situational influences. When presented with lightly fictionalized narratives about these findings, participants exhibit no tendency toward changing judgments of freedom and responsibility compared to control groups. This suggests that people seem unwilling to adopt skeptical attitudes about agency on the basis of these scientific findings.

1. **Introduction**

Some research suggests that moral behavior can be strongly influenced by trivial features of the environment of which we are completely unaware. Consider, for instance, the Watching Eye Effect. Multiple studies (both in the laboratory and in the field) have found that priming people with images of eyes can make them more likely to be charitable (Kelsey, Vaish, & Grossmann, 2018; Fathi, Bateson, & Nettle, 2014), cooperative (Ernest-Jones et al., 2011), and generous (Burnham & Hare, 2007; Haley & Fessler, 2005) and less likely to litter (Ernest-Jones, Nettle, and Bateson, 2011) or steal a bicycle (Nettle, Nott, & Bateson, 2012). Evidence for the Watching Eye Effect has even been found in young children (Kelsey, Grossmann, & Vaish, 2012).[[2]](#footnote-2)

 There is a lively ongoing debate among psychologists and philosophers about whether findings such as these might challenge or undermine our commonsense understanding of morality, including how we think about agency and responsibility. Some draw skeptical conclusions from the findings, arguing that the unwitting influence of environmental stimuli seems to diminish the role of conscious deliberation in action and thereby undermine individual autonomy. While responses have been offered to these arguments, both sides have proceeded on the basis of untested assumptions about commonsense conceptions of agency and responsibility, such as the importance of conscious deliberation in practical reasoning. However, if these assumptions about commonsense moral commitments turn out to be incorrect, then it is unclear how these scientific findings would *challenge* such commitments. Thus, rather than add to the discussion of the implications, we conducted four vignette-based studies (*N* = 1437) to investigate how people think about the metaphysical and moral implications of scientific findings like the ones mentioned above. This provides an indirect measure of the commitments underlying commonsense morality that are central to debates about what these empirical results mean. These results show that most participants do not draw skeptical conclusions about agency and responsibility when presented with evidence for phenomena like the Watching Eye Effect. Before discussing our studies—which we believe are the first to explore this issue (although see Kozuch & Nichols, 2011 for related results)—we will set the stage with an overview of the relevant scientific findings and the associated philosophical arguments.

1. **The Rise of Situationism and Automaticity**

We often explain our own behavior and the behavior of others by appealing to character traits—e.g., “Jenny told the truthin the face of possible personal repercussionsbecause she is a courageous and honest person.” In doing so, we assume that these traits shed light on who we are and why we do what we do. We also assume that having virtuous traits like courage and honesty leads us to behave consistently from one context to the next. One cannot be a genuinely virtuous agent if one regularly (or even semi-regularly) fails to act appropriately when the opportunity arises. Finally, we expect virtuous agents to do the right thing even when doing so may be difficult or come with potentially significant costs. So, when Jenny speaks truth to power in the face of possible retaliation, this is precisely what we expect her to do given her virtuous character. This package of commitments is part of how people think about the relationship between character and action.

Given this commonsense commitment about the role of character in action, it is unsurprising that the focus on character traits is commonplace in the philosophical literature on moral agency. On the one hand, both ancient (e.g., Aristotle, 1994; Plato, 1997) and contemporary (e.g., Annas, 1999; 2011; Anscombe, 1958; Driver, 2001; Foot, 1978; Hursthouse, 1996; 1999; Slote, 1992; 2001; Taylor, 1991) virtue ethicists have made character the centerpiece of their ethical theories. According to Aristotle, virtuous character is “firm and unchangeable” (1984, *EN*, 1105b1). As such, he claims that virtuous agents never knowingly or intentionally behave viciously (1984, *EN*, 1100b32), and that they do what virtue requires even in the most demanding and difficult situations (1984, *EN*, 1104a8-10). On this view, right actions like Jenny’s speaking truth to power are those that flow from virtuous character traits like courage and honesty, traits that we rightly hold in high esteem precisely because they consistently issue in right action.

Further, the capacity for practical reasoning also plays a ubiquitous role in ethical theories related to agency (e.g., Brink, 2013; Dennett, 1984; Fingarette, 1972; Fischer, 1987; 1994; Fischer & Ravizza, 1998; Haji, 1998; MacIntyre, 1957; McKenna, 2013; Morse, 2004; 2008; Neely, 1974; Nelkin, 2011; Nozick, 1981; 2011; Pink, 2009; Vargas, 2013; Velleman, 1980; Wallace, 1994; Wolf, 1990). This is, presumably, because the ability to deliberate and make choices is central to responsible agency. For instance, Brink (2013) claims that responsibility requires “that agents not simply act on their strongest desires, but be capable of stepping back from their desires, evaluating them, and acting for good reasons” (p. 131), and Vargas (2013) claims that responsible agents possess “a suite of basic agential capacities implicated in effective self-directed agency . . . and [are] also possessed of the relevant capacity for (a) detection of suitable moral considerations and (b) self-governance with respect to [those considerations]” (p. 335).

Philosophers and cognitive scientists often emphasize that the kind of practical reasoning at issue is conscious. Some philosophers count something as practical reasoning only if the relevant contents are manifested in consciousness (Audi, 2006: 116).[[3]](#footnote-3) Others claim that we take ourselves to be consciously aware of at least some of the reasons that drive our decision-making[[4]](#footnote-4) (Bargh, 1997; Clark et al., 2023; Prinz, 2003; Wegner, 2002). Thus, philosophers and cognitive scientists attribute two different positions to people when theorizing about commonsense morality: (1) what people do across different situations is largely a function of their stable character traits, and (2) conscious deliberation is the primary cause of most of our intentional actions.[[5]](#footnote-5) Notably, discussions of these commonsense commitments rarely provide evidence that people think or act in ways that align with these positions. Claims about character and conscious deliberation are *portrayals* of a commonsense view of morality, though it is an open question whether this portrayal is accurate.

Yet, despite the central role played by character and practical reasoning in both commonsense morality and ethical theory, a large body of empirical research in psychology paints a different picture of human agency: First, some findings support the thesis of situationism. According to more modest versions of this thesis, while character traits may play some limited role in human behavior, environmental stimuli play a much more significant role than we ordinarily assume. The Watching Eye Effect is a case in point. We tend to think that behavior like charitable donation or refraining from littering might stem from compassion or a conscious concern for the environment; instead, research suggests that situational factors like the presence of eyes are considerably more important than we initially think and perhaps overemphasize the importance of intrinsic properties in causing such behavior (Nisbett & Ross, 1980; Ross & Nisbett, 1991; Ross, 1977). Second, there are findings that support the closely related thesis of automaticity—that is, the view that in response to environmental stimuli, automatic mental processes can sometimes reflexively lead to behavior in the absence of conscious awareness or control (e.g., Anderson, 1992; Bargh, 1994; 1996; Greenwald & Banji, 1995). These automatic processes can influence not only our perceptual and motor skills but also our judgments, goals, motivations, and even our moral behavior (Amaya & Doris, 2015; Bargh, 1997; Doris 2015).

In light of research on situational influences and automaticity, some researchers have raised doubts about the role that character traits and practical reason play in moral behavior. For instance, empirically-minded skeptics have appealed to the situationist literature to challenge the psychological plausibility of the kind of character-based moral psychology adopted by Aristotle and contemporary virtue ethicists (Alfano, 2013; Doris, 1998; 2002; 2005; 2006; 2008; 2009; 2010; Doris & Stich, 2005; 2006; Harman, 1999, 2000b; Machery, 2010; Merritt, 2000; 2009; Merritt, Doris, & Harman, 2010; Vranas, 2005). For instance, Harman (1999) claims:

On this occasion I discuss a different kind of rejection of folk morality, one that derives from contemporary social psychology. It seems that ordinary attributions of character traits to people are often deeply misguided, and it may even be the case that there is no such thing as character, no ordinary character traits of the sort people think there are, none of the usual moral virtues and vices (p. 316).

The most alarming aspect of the situationist findings is that “both disappointing omissions and appalling actions are readily induced through seemingly minor situations.” (Merritt, 2000, p. 357). We do not fancy ourselves to be the sorts of agents whose moral behavior is significantly influenced by the smell of fresh baked goods, whether we are in a rush, whether we find loose change in a phone booth, whether the victim in a criminal trial is physically unattractive, etc. (Latané, & Darley, 1970; Vrij, & Firmin, 2001). Instead, according to the standard portrayal of the commonsense view, we assume that our moral character and conscious deliberation drives our behavior and largely shields us from the sorts of morally insignificant situational influences discussed above. However, critics of virtue ethics claim that empirical results call into questions these commonsense commitments about character and conscious deliberation (again, assuming this portrayal of the commonsense view is correct).

 Unsurprisingly, those philosophers who place stock in the explanatory value of character traits have not taken the situationist challenge lying down (see, e.g., Adams, 2006; Annas, 2005; Appiah, 2008; Arpaly, 2005; Badhwar, 2009; Besser-Jones, 2008; Kamtekar, 2004; Knobe & Leiter, 2007; Kupperman, 2001; Miller, 2003; 2013a; 2013b; Montmarquet, 2003; Russell, 2009; Samuels & Casebeer, 2005; Sarkissian, 2010; Snow, 2010; Solomon, 2003; 2005; Sreenivasan, 2002; 2020; Swanton, 2003; Upton, 2009; and Webber, 2006a; 2006b; 2007a; 2008b). One strategy for responding to the situationist challenge is to argue that the critics of virtue ethics have misrepresented or misunderstood the nature of virtue (e.g., Annas, 2005; 2011; DePaul, 1999; Kamtekar, 2004; Kupperman, 2001).

For instance, Annas (2005) claims that situationists overlook or neglect a central element of the virtue ethics tradition—namely, the role played by practical reason in developing virtues and acting virtuously. Annas notes that, on Aristotle’s view, “the more you have developed a virtue through reflection and reasoning, the more, not less, aware you become of what is important in different situations” (2005, p. 638). As such, a properly virtuous agent will not be subject to the sorts of situational influences that critics of virtue ethics cite, so the situationist challenge poses problems only for a “radically unintellectual version of virtue” (2005, p. 639).

But this response to the situationist challenge opens the door to the previously mentioned worries about automaticity. After all, the very capacity for practical reasoning that Annas and others appeal to in defending the role played by character traits in moral behavior is threatened by the findings that suggest that much of human behavior is driven by automatic cognitive mechanisms that sometimes bypass conscious awareness—processes that are initiated by environmental stimuli and that can influence our moral behavior without being influenced by our practical reasoning.[[6]](#footnote-6) As Bargh (1997) points out: “Automaticity pervades everyday life, playing an important role in creating the psychological situation from which subjective experience and subsequent conscious and intentional processes originate. Our perceptions, evaluations, and the goals we pursue can and do come under environmental control” (p. 50).

That said, while the situationist challenge to virtue ethics has received the lion’s share of the attention in the philosophical literature, some philosophers have also discussed the potential threat that this challenge might pose to free will and moral responsibility (e.g., Brink, 2013; Doris, 2002; 2015; Mele, 2014; Mele & Shepherd, 2013; Miller, 2016; Nahmias, 2007; Nelkin, 2005; McKenna & Warmke, 2017; Vargas, 2013). For instance, Nahmias (2007) points out that while we normally take ourselves to be self-controlled creatures who can “deliberate to reach some consistency among [our] competing desires and values” (p. 1)—the gathering findings on situationism and automaticity potentially call our autonomy into question by showing that “we do not govern our behavior according to principles we have consciously chosen” (p. 1). While Nahmias acknowledges that these findings do reveal that our autonomy may be more limited than we have assumed, he does not believe that the threat is global in scope or that skepticism about moral agency is warranted.

Nelkin (2013) develops a similar view when it comes to the potential threat that the findings on situationism and automaticity may pose to our capacity for reasons-responsiveness. Take, for instance, views whereby morally responsible agents must “act on a mechanism that is regularly receptive to reasons, some of which are moral reasons, and at least weakly reactive to reason” (Fischer & Ravizza, 1998, p. 82) or where, “an agent is responsible if and only if the agent can do the right thing for the right reasons (Wolf, 1990, p.68). As Nelkin (2013) points out, “one way of seeing the situationist cases—or at least some of them—as troubling is this: simply put, the subjects seem to be acting for bad reasons, or at least not acting for good reasons, and they seem stuck doing so. At the same time, having the ability to act for good reasons is essential to freedom and/or responsibility” (p. 199).

For instance, in the studies on the Watching Eye Effect, many (if not most) of the participants failed to act for the right kinds of reasons even when they ultimately did the right thing—that is, they did not act upon reasons that they would presumably endorse if given the opportunity to do so. If these participants were told that their behavior was unconsciously influenced by images of eyes rather than their characterological disposition to be charitable, cooperative, helpful, honest, etc., they would likely deny it (as predicted by the fundamental attribution error). These sorts of findings reveal not only that our actions might often be influenced by reasons that we eschew, but that we might often be unaware of the environmental stimuli that prevent us from acting in accordance with the reasons that we endorse. This lack of self-knowledge and situational awareness makes it difficult for us to guard ourselves against situational influences that get us to do the wrong thing or that get us to do the right thing for the wrong reasons.

As Vargas (2013) rightly points out, the findings on situationism and automaticity represent a potential dual threat. On the one hand, they could threaten the various theories of ethics and action that we have been discussing. On the other hand, these findings could threaten the dictates of commonsense morality about the role of character and conscious deliberation in action.

It is important to recall the structure of the argument. According to a common portrayal of the commonsense view of morality, there is an important role for character and conscious deliberation in practical agency. But, according to some interpretations of the empirical evidence on automaticity and situational influences on decision-making, conscious deliberation seems unimportant for some relevant forms of practical reasoning. Thus, if this common portrayal is correct, then these interpretations of the results challenge commonsense commitments about morality.

This paper assesses whether the portrayal of commonsense commitments is accurate such that these interpretations threaten such commitments. Some results have also called this portrayal into question. Vekony et al. (2020) showed that people attribute intentionality to actions that occur without knowledge or awareness. However, the vignettes portrayed habitual or skilled actions that stem from character (e.g., a proficient free throw shooter making a free throw in a dark gym where they cannot see the basket). Kozuch & Nichols (2011) also found that people do not always presume that we are aware of why mental events or intentional actions are happening as they are happening. However, Kozuch & Nichols found that people tend to think that the reasons influencing decision-making *are* accessible to introspection (see their Experiment 4).[[7]](#footnote-7) Thus, some evidence suggests that character and conscious deliberation are central to decision-making even if people accept the possibility of intentional action in the absence of awareness or accessibility of reasons motivating such action.

We examined whether evidence for situationism and automaticity influence people’s attributions of agency and responsibility in the way that skeptics might predict. To test this, we presented people with lightly fictionalized versions of experiments from the empirical literature. As we will now discuss, across four vignette-based studies we found that participants do not tend to draw skeptical conclusions about agency and responsibility in the face of evidence that our moral behavior is more strongly influenced by environmental stimuli and automatic processes than we may have traditionally assumed.

1. **Exploring Folk Intuitions**
	1. *Study 1*

If the way that situationists and skeptics have construed commonsense views about agency and responsibility is correct, then providing people with empirical evidence of morally insignificant environmental stimuli influencing moral behavior in laboratory and/or real-world settings should diminish their attributions of agency and responsibility in those settings. In order to test whether people draw skeptical conclusions when presented with the results of situationist studies, one would need to ensure that they understand (a) that the environmental stimuli is the primary cause (or “difference maker”) of the behavior in the studies, (b) that the participants depicted in the vignettes are unaware of the influence of the environmental stimuli, and (c) that the participants depicted in the vignettes would reject the suggestion that the environmental stimuli was the primary cause of their behavior (especially when the behavior is praiseworthy).[[8]](#footnote-8) So, we designed Study 1 to test three hypotheses about how people interpret empirical findings about the causal efficacy of situational influences and unconscious primes on moral behavior.

H 1.1: People will judge that the environmental stimuli or unconscious primes (as described in situationist experiments) are the primary cause of behavior for participants in situationist experiments.

H 1.2: People will judge that the participants in situationist experiments are unaware of the primary causes of their behavior (i.e., the environmental stimuli or prime).

H 1.3: People will judge that participants in situationist experiments would reject the claim that environmental stimuli or unconscious primes are the driving influence behind their behavior.

3.1.1 Methods and participants

*Sample size.* 540 participants were recruited on Amazon’s Mechanical Turk. To determine sample size, we conducted an *a priori* power analysis using G\*Power software (Faul et al., 2007). For a two-tailed one sample *t*-test to have 95% power to detect a minimal effect size of interest (*d* = 0.50), 54 participants were recommended per condition. Because we were uncertain about potential effects of vignette or ordering effects, we wanted to collect enough participants for each of 8 possible conditions to be adequately powered (*N* = 432). To account for exclusions and attrition, we over-recruited by 25%.

Participants had to be in the United States and to have completed 500 surveys with a 97% approval rate to qualify for the study. 80 participants were excluded for failing a simple attention check (“If Monday is the first day of the week, what is the third day of the week?”). 11 participants were excluded for failing a pre-registered comprehension check. The wording of the comprehension question varied depending on condition, but in each case the check concerned the percentage of people who acted congruently with the expected effect of the prime (full materials are available in the *Supplementary materials* §4). 125 participants were excluded for providing invalid responses to an open-ended question (*N* = 324, *M*age = 39.7, *SD*age = 12.1, 45% female). No data were analyzed prior to stopping collection and applying exclusions.

3.1.2 Procedure

Participants were randomly assigned to read about one of four possible situationist experiments. Participants read about how the presence or absence of some situational factor inclined people to act in different ways. For example, some participants read about a study where agents were primed to cheat on a task because they were prompted to think about achievement goals[[9]](#footnote-9):

There is gathering evidence from psychology that human decision-making and behavior are influenced by factors of which we are not consciously aware.

In a recent study, Robert Bard and colleagues of Yale University subliminally primed participants in the experimental condition (but not in the control condition) to think about the importance of achieving one’s goals. Participants were then asked to perform a task that required them to identify and write down as many words as they could, based on a set of seven “Scrabble” letter tiles. A few minutes later, they were told over an intercom to stop. Hidden video cameras recorded the participants’ behavior throughout to see how many of them continued to search for and write down words after they were told to stop doing so.

Whereas only 21% of the participants in the control condition cheated, more than half of the participants in the “achievement” condition (55%) ignored the instructions to stop.

John was a participant who was subliminally primed to think about the importance of achievement goals. He cheated in the experiment.

Half of the stimuli described primed behaviors that were morally good (helping), while the other half described primed behaviors that were morally bad (cheating). After reading the vignette, participants answered a comprehension check: “How many participants in [the experimental group] [cheated/helped] in the experiment?” Participants were then asked to respond to several items about the vignette:

* *Reason*: What caused John to cheat? (Participants were provided an open response box to type answers; they were also told: “If you think there are several relevant causes, please write what you think is the primary cause that explains why John [cheated/helped]”)
* *Aware*: Consider your answer above. Was John aware of the primary cause influencing his behavior? (1 = Definitely NOT aware, 5 = Unsure, 9 = Definitely IS aware)
* *Avowal*: If you told John the primary cause of why he cheated, would he believe you or not? (1 = Definitely WOULD NOT believe, 5 = Unsure, 9 = Definitely WOULD believe)

Participants were also given a 9-pt. scale (1 = Strongly disagree, 5 = Unsure, 9 = Strongly agree) to answer the following items:

* *Free*: John had free will
* *Responsibility*: John is [blameworthy/praiseworthy] for [cheating/helping]
* *Character*: John is a [dishonest/honest] person
* *Reliability*: Findings from psychology studies like these are reliable.[[10]](#footnote-10)

The order of these items was counterbalanced to control for potential order effects of asking about primary reasons. That is, some participants answered Reason, Awareness, and Avowal questions first, while others answered them second.

3.1.3 Analysis plan

In addition to our pre-registered analyses, we supplemented our analytic approach with Bayesian statistical tests to quantify the amount of evidence in favor of the null hypothesis. This was done to mitigate concerns about the high attrition rate across our studies and to better understand the null results that we find across our studies. It is not possible to quantify the evidence for a null effect using conventional frequentist testing. A Bayes Factor is a continuous measure of the relative strength of evidence and can quantify the degree to which the data are compatible with the null over the alternative hypothesis. All Bayesian analyses were performed using the R package BayesFactor and Bayes Factors (BF) were calculated using default settings (Morey & Rouder, 2018). BF01 indicates evidence in favor of the null hypothesis. As per previous recommendations, we refer to a BF > 3 as “moderate” and BF > 10 as “strong” evidence (Rouder et al., 2009).

3.1.4 Results

We sorted responses about reasons into three categories: Internal, Situation, Hybrid. Internal reasons referred to intrinsic states of the agent as the primary reasons for acting (e.g., decisions, choices, etc.). Situation reasons referred to the situational factor as the primary cause (e.g., finding a dime). Hybrid reasons referred to situational factors that prompted internal states relevant to the decision (e.g., finding a dime put John in a good mood, which made him more likely to help). Some internal responses were:

* I honestly have no idea, but if I had to guess it’s because he’s a good person. Not because of any dime bullshit.
* I don’t believe there is any correlation to find a dime and helping a woman. He helped because he is a decent person.
* John is an inherently honest person who would not cheat for any reason. It is part of who he is.

Some situational reasons were:

* He was primed to think it was important to achieve a goal.
* John saw a dime on the phone.
* He was rushed.

Hybrid responses included the following:

* John was in a good mood because he found the dime.
* I would think that he believed it was morally wrong to cheat and the word on the poster (“GODS”) caused him to think about other worldly things that would [*sic*] notice if he was cheating.

**Table 1** summarizes count data and other descriptives for Study 1.

*Table 1. Counts, means, and SDs of dependent variables for each condition in Study 1*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Condition** | **Motive** | **Count** | **Aware** | **Avowal** | **Free** | **Responsible** | **Character** | **Reliable** |
| *Achievement* | *Internal* | 29 | 4.79 (*2.6*) | 6.00 (*2.0*) | 7.55 (*1.4*) | 6.59 (*1.8*) | 6.10 (*2.0*) | 6.31 (*1.3*) |
| *Achievement* | *Situation* | 26 | 5.81 (*2.4*) | 6.58 (*1.6*) | 7.35 (*1.5*) | 6.96 (*1.3*) | 6.50 (*1.7*) | 6.58 (*1.4*) |
| *Gods* | *Internal* | 49 | 6.10 (*2.2*) | 6.76 (*2.0*) | 7.94 (*1.6*) | 7.27 (*2.0*) | 7.63 (*1.3*) | 6.08 (*1.7*) |
| *Gods* | *Situation* | 24 | 6.21 (*1.7*) | 6.79 (*1.4*) | 6.58 (*1.8*) | 6.21 (*1.6*) | 6.58 (*1.7*) | 6.67 (*1.2*) |
| *Phone* | *Internal* | 31 | 5.19 (*2.6*) | 6.23 (*2.5*) | 8.35 (*1.0*) | 7.65 (*1.3*) | 7.48 (*1.3*) | 5.29 (*2.4*) |
| *Phone* | *Situation* | 56 | 5.96 (*2.1*) | 6.50 (*1.7*) | 7.09 (*1.4*) | 6.73 (*1.5*) | 6.71 (*1.4*) | 6.91 (*1.3*) |
| *Samaritan* | *Internal* | 36 | 6.44 (*2.0*) | 6.81 (*1.7*) | 7.44 (*2.1*) | 6.00 (*2.3*) | 5.39 (*2.2*) | 6.39 (*2.1*) |
| *Samaritan* | *Situation* | 42 | 6.43 (*2.0*) | 7.43 (*1.4*) | 6.79 (*1.8*) | 5.83 (*1.9*) | 5.33 (*2.0*) | 6.02 (*1.8*) |

**Note.** This table does not include statistics for Hybrid responses. A full list of descriptives for Study 1 is available in the *Supplementary materials* §1.

**Figure 1** summarizes how participants thought about the cause of action described in the vignettes.



*Figure 1.* Proportion of motive responses by vignette in Study 1. \*\*\* = *p* < .001, ns = not significant at *p* = .05 level.

A test for equality of proportions found evidence for a difference in the proportion of motive responses across vignettes (c2(3) = 16.34, *p* < .001). Pairwise comparisons that used Holm correction for multiple comparisons showed that the proportion of motive responses in the Gods vignette differed significantly from the distribution in the phone vignette (*p* < .001). All other comparisons were not significant (all *p* > .27)*.* To see whether participants in any vignette selected significantly more of one type of motive, we conducted a series of one-sample proportion tests with continuity correction to see whether the proportion of situational motive responses was significantly different than chance (50%). We found no evidence for a significant difference from chance in the Achievement (*p* = .17; BF01 = 2.93), Phone (*p* = .27; BF01 = 2.98), and Samaritan vignettes (*p* = .65; BF01 = 2.67), though Bayesian analyses indicated that the evidence in favor of the null was weak-to-moderate. While we found evidence for a significant difference in the proportion of situational motive responses in the Gods vignette (c2(1) = 12.64, *p* < .001, Cohen’s *h* = .43, 95% *CI*[.13, .74]), this was because participants had significantly more *internal* motive responses (49) than situational responses (24).

Even though participants did not uniformly perceive the relevant situational factor as the primary cause of behavior, it still might be the case that the participants who *did* perceive the situational factor as the primary cause attributed less awareness to the agent and thought the agent would disavow the situational factor as a cause of behavior. To test this, we fitted separate linear models to predict awareness and avowal from motive response, condition, and their interaction. We also statistically controlled for vignette in each model. Hybrid responses were excluded from the data because there were too few to make reliable statistical inferences.

There was no evidence for an effect of vignette on judgments of reliability (*F*(3, 285) = 0.36, *p* = .78, h2p = .004, 90% *CI*[.00, .01]), but there was evidence for an interaction between vignette and motive type (*F*(3, 285) = 4.79, *p* = .003, h2p = .05, 90% *CI*[.01, .09]). However, the interaction was due to a significant difference in perceived reliability of the experiment describing a person helping someone because they found extra change in a payphone. People who perceived the situational cue as the primary cause saw the study as more reliable compared to those who perceived the primary cause as an internal disposition (*M*diff = 1.62, *t*(285) = -4.33, *p* < .001, *d* = -0.97, 95% *CI*[-1.42, -0.52]). To account for this difference, including the variability of perceived reliability across vignettes, we statistically controlled for reliability in all models summarized below.

ANOVAs found no evidence for an effect of motive on judgments of awareness (*F*(1, 284) = 1.53, *p* = .22, h2p = .01, 90% *CI*[.00, .03]) or on judgments of avowal (*F*(1, 284) = 1.84, *p* = .18, h2p = .01, 90% *CI*[.00, .03]). There was a trending effect of motive on judgments of awareness, with people recognizing the *situational cue* as the primary cause attributing marginally more awareness than people who recognize an *internal disposition* as the primary cause (*M*diff = .36, *t*(284) = -1.35, *p* = .18, *d* = -0.17, 95% *CI*[-0.41, 0.08], BF01 = 2.85). Across motive types, participants attributed awareness significantly above the uncertainty point (internal: *M* = 5.69, 95% *CI*[5.27, 6.11]; situational: *M* = 6.05, 95% *CI*[5.62, 6.48]). A similar trend emerged for judgments of avowal, with the situational cause prompting greater certainty in avowal relative to internal causes (*M*diff = .29, *t* (284) = -1.32, *d* = -0.16, 95% *CI*[-0.41, 0.08], BF01 = 2.62) (see **Figure 2**). Bayesian t-tests indicated weak-to-moderate evidence for the null hypothesis that there was no effect of motive type on attributions of awareness or avowal.



*Figure 2.* Judgments of awareness (left panel) and avowal (right panel) by motive type and vignette. Hybrid responses were excluded due to low response rates. Error bars represent standard errors.

 We explored how character judgments and attributions of free will and responsibility changed based on perceived motive. To do this, we fit linear models to predict judgments of free will, responsibility, and character from motive response, awareness, vignette and all interactions while statistically controlling for vignette and reliability.

 An ANOVA found evidence for an effect of motive response on judgments of free will (*F*(1, 276) = 23.90, *p* < .001, h2p = .08, 90% *CI*[.04, .14]), responsibility (*F*(1, 276) = 8.63, *p* = .004, h2p = .03, 90% *CI*[.01, .07]), and character (*F*(1, 276) = 10.74, *p* = .001, h2p = .04, 90% *CI*[.01, .08]) (see **Figure 3**). When participants perceived an internal motive compared to a situational motive, they attributed more free will (*M*diff = .89, *t*(276) = 4.37, *p* < .001, *d* = 0.55, 95% *CI*[0.30, 0.81]), responsibility (*M*diff = .59, *t*(276) = 2.74, *p* = .007, *d* = 0.35, 95% *CI*[0.10, 0.60]), and character (*M*diff = .60, *t*(276) = 3.04, *p* = .003, *d* = 0.38, 95% *CI*[0.13, 0.64]). However, there was no evidence for an interaction between motive type and awareness for these judgments (all *p* > .08).



*Figure 3*. Judgments of free will (left panel), responsibility (middle panel), and character (right panel) by motive type and vignette. Hybrid responses were excluded for low response rate. Error bars represent standard errors.

3.1.5 Discussion

Recall our three hypotheses about how people would interpret the findings of situationist experiments:

H 1.1: People will judge that environmental stimuli or unconscious primes (as described in situationist experiments) are the primary cause of behavior for participants in situationist experiments.

H 1.2: People will judge that the participants in situationist experiments are unaware of the primary causes of their behavior (i.e., the environmental stimuli or prime).

H 1.3: People will judge that participants in situationist experiments would reject the claim that environmental stimuli or unconscious primes are the driving influence behind their behavior.

We found no evidence to support these hypotheses. In some cases, we found some evidence *against* these hypotheses. Participants were not more likely to cite a situational factor as the primary cause of behavior in the experiment. Moreover, those who *did* perceive the situational cue as the primary cause of behavior attributed marginally more awareness to the agent and were more certain the agent would accept the situational cue causes their behavior. Moreover, these results cannot easily be explained in terms of participants thinking the experiments are unreliable, as these findings held even when statistically controlling for judgments of reliability.

In short, we had a hard time getting participants to take the skeptical interpretation of situationist experiments. When reading about the results of psychological experiments on automaticity and priming, many participants interpret the agent’s behavior in the vignette in terms of internal motives rather than situational motives even when the explicit focus of the experiments being described was the latter. Moreover, participants tend to (mistakenly) attribute a high degree of *awareness* of the causal influence of the situational factor to the agents in the vignettes even when it should be clear that they lack awareness. Finally, they do not tend to think that agents in the vignettes would *disavow* being caused to act by a situational factor if they were so informed. Thus, our results cut against the kinds of skeptical interpretation of these kinds of experiments that are favored by the situationists.[[11]](#footnote-11)

 However, exploratory analyses identified an effect of motive response on attributions of free will, responsibility, and character. People who thought the primary cause of behavior was situational rather than internal tended to attribute significantly less free will and responsibility for action and thought the actions said less about the person’s character. These results should be interpreted cautiously for two reasons. First, there was no evidence for an interaction between motive type and judgments of awareness. Thus, whatever the effect of motive type is on these judgments does not reflect underlying differences in attributions of awareness. Second, the effects were small and our sample was not adequately powered to detect such small effects.

 These results cannot distinguish between two possible interpretations of participant behavior. On the one hand, participants might simply misunderstand the experiments and not see them as providing evidence for the situationist thesis. On the other hand, the vignettes might elicit cognitive biases that function to preserve attributions of agency and responsibility by distorting judgments concerning the strength of the influence of situational cues or the degree of the agent’s awareness of the stimuli. This kind of defensive thinking about agency and responsibility has been identified in other studies (Murray & Nadelhoffer, 2023). Study 1 provides some evidence that supports the defensive interpretation, insofar as most people attributed greater awareness and avowal of situational cues relative to internal causes. However, the vignettes we used confound the two interpretations because participants are not explicitly told about the cause of behavior (the situational cue), the lack of awareness, or the possibility of disavowal. If these aspects of the vignettes were made explicit and participants still exhibited the same pattern of responses, this would provide good evidence for the underlying ‘stickiness’ of the dispositionist tendencies of commonsense morality and the operation of cognitive biases in interpreting the scientific evidence. To test this, we conducted another experiment.

*3.2 Study 2*

3.2.1 Methods and materials

*Participants*. 674 participants were recruited on Academic Prolific.[[12]](#footnote-12) To determine sample size, we conducted an *a priori* power analysis using G\*Power software (Faul et al., 2007). For a two-tailed *t*-test to have 95% power to detect a minimal effect size of interest (*d* = 0.3), 580 participants were recommended. We over-recruited by 15% to account for exclusions and attrition. Minimal effect size of interest was calculated from effects observed in Study 1. To account for possible inflation, we subtracted the standard error from the observed effect in the power analysis.

The collected sample was representative on age, gender identity, and ethnicity based on US census data. We pre-registered two exclusion criteria. 45 participants were excluded for providing invalid responses to an open-ended question. 0 participants were excluded for failing a pre-registered comprehension check. 7 participants were excluded for self-reported distraction (*N* = 622, *M*age = 45.92, *SD*age = 15.5, 51.3% female, 48.7% male, 11.6% Black, 10.4% mixed, 7.4% other, 61.2% white). No data were analyzed prior to stopping collection and applying exclusions.

3.2.2 Procedure

Materials were adapted from Study 1. The “Experimental” vignettes described an experiment that examined the effect of situational priming on either cheating or helping behavior. An example vignette is included below (full vignettes are included in the *Supplementary materials* §4):

|  |  |
| --- | --- |
| ***Control*** | ***Experimental*** |
| In a recent study, Robert Bard and colleagues of Yale University explored cheating behavior. Participants were asked to perform a task that required them to identify and write down as many words as they could, based on a set of seven “Scrabble” letter tiles. A few minutes later, they were told over an intercom to stop. Hidden video cameras recorded the participants’ behavior throughout to see how many of them continued to search for and write down words after they were told to stop doing so. John was one of the participants in the study who cheated. | There is gathering evidence from psychology that human decision-making and behavior are influenced by factors of which we are not consciously aware.In a recent study, Robert Bard and colleagues of Yale University subliminally primed participants in the experimental condition (but not in the control condition) to think about the importance of achieving one’s goals. Participants were then asked to perform a task that required them to identify and write down as many words as they could, based on a set of seven “Scrabble” letter tiles. A few minutes later, they were told over an intercom to stop. Hidden video cameras recorded the participants’ behavior throughout to see how many of them continued to search for and write down words after they were told to stop doing so.Whereas only 21% of the participants in the control condition cheated, more than half of the participants in the “achievement” condition (55%) ignored the instructions to stop. The participants who ignored the instructions were **not aware** that being primed to think about their goals caused them to cheat. When researchers told participants about the priming effect after the experiment, most participants **denied** that the prime caused them to ignore the instructions.John was a participant who was subliminally primed to think about the important of achievement goals. He cheated in the experiment. |

Unlike Study 1, the Experimental vignettes explicitly noted three things: (1) that the situational prime was the cause of the relevant behavior, (2) that the person was unaware of the influence of the situational prime on their behavior, and (3) when experimenters told participants about the influence of the prime after the experiment, the participants rejected the idea. For each “Experimental” vignette, we developed a “Control” vignette that described the same experiment but did not mention anything about situational priming or automaticity. Participants were randomly assigned to read one of 4 vignettes. Afterward, they answered similar items to those used in Study 1. The *awareness*, *avowal*, and *reliability* items were the same, but the following items were changed:

* *Free will*: How much free will does John have?
* *Responsibility*: How much blame (praise) does John deserve for [condition-dependent action]?
* *Character*: How dishonest (generous) is John?

All items used a 9-pt. Likert scale (1 = none, 5 = some, 9 = a lot). These items were changed to address a possible confound of the items in Study 1. When asking about *agreement* (as in Study 1), participants might differ in terms of the degree to which they agree with a statement without altering the perceived degree of free will (e.g., a participant might agree less strongly that a person has a lot of free will). The new items address this confound by asking about magnitudes rather than agreement.

3.2.3 Results

Participants who read an “Experimental” vignette cited a situational influence significantly more often compared to an internal motive as the primary cause of action (64% vs. 36%; c2(1) = 24.3, *p* < .001). Participants who read a “Control” vignette cited an internal motive significantly more often compared to situational influences (94% vs. 6%; c2(1) = 257.6, *p* < .001). Bayesian one-sample proportion tests provided overwhelming evidence for both results (BF10 > 26000 in both cases).

 A 2(vignette)x2(condition) ANCOVA with reliability as a covariate found no evidence for an effect of reliability on judgments of awareness (*F*(1, 614) = 2.44, *p* = .12, h2g = .004). Because of this, we did not include reliability as a variable in the models reported below.

 Separate linear models were fit to predict perceptions of awareness and avowal from motive attribution. A term for vignette was also included. Because the manipulation successfully induced changes in the perceived motive (with control vignettes prompting internal causes and experimental vignettes prompting external causes), we did not include an additional term for condition. There was an effect of motive attribution on judgments of awareness (*t*(618) = -7.50, *p* < .001, *d* = -0.74, 95% *CI*[-0.94, -0.54]) and judgments of avowal (*t*(622) = -2.64, *p* = .01, *d* = -0.36, 95% *CI*[-0.63, -0.09]). In both cases, those who attributed an external motive perceived less awareness and less avowal (*M* = 3.64 and *M* = 4.76, respectively) compared to those who attributed an internal motive (*M* = 5.39 and *M* = 5.50, respectively).

 In Study 1, we found that seeing behavior as caused by a situational influence (compared to an internal motive) had an effect on judgments of free will, responsibility, and character inference. To assess whether these findings replicated, we fit linear models to predict judgments of free will, responsibility, and character from motive attribution. Terms for reliability and vignette were included in the model. Additionally, to ensure that our tests included participants who correctly interpreted the vignettes, we analyzed a subset of the data where participants who attributed an external motive also rated awareness below the indifference point (*M* < 5). This left 579 participants.[[13]](#footnote-13)

 There was no evidence for an effect of motive attribution on judgments of free will (*t*(577) = -0.13, *p* = .89), responsibility (*t*(577) = 0.95, *p* = .34), or character inference *t*(577) = 0.08, *p* = .93). Bayesian *t*-tests indicated moderate-to-strong evidence in each case for the null hypothesis (Free will: BF01 = 9.71; Responsibility: BF01 = 6.31; Character: BF01 = 9.75).

3.2.4 Discussion

Two important findings emerged from Study 2. When information about the causal influence of the situational cue is made explicit, participants tended to cite the appropriate cause as the primary reason for acting. This, in turn, was associated with differences in the perceived awareness of the causal factor and whether the agent would accept that causal factor as actually influencing their decision-making. However, there was no evidence that these differences in perceived awareness and avowal tracked meaningful differences in attributions of free will and responsibility or that situationally caused actions say less about a person’s character than internally caused actions. In other words, participants seem to understand the vignettes while continuing to make judgments that align with the dispositionist assumptions of commonsense morality.

 Unlike Study 1, we failed to replicate the effects of motive attribution on judgments of free will, responsibility, and character inference. This suggests that even when participants explicitly understand the causal influence of the situational cue on behavior, this does not reliably drive different attributions of agency (similar null results were found in a larger study that included more vignettes and comparisons with control conditions; see *Supplementary materials* §1). One explanation for this is that participants might think that the prime is not particularly strong. It could be the case that if the prime exerted a stronger influence on behavior, then people would alter their judgments of free will and responsibility. One metric for thinking about the strength of situational primes is in terms of the proportion of participants who act congruently with the prime. Thus, to replicate these null results and control for this possibility, we conducted another study where we varied the strength of the prime across conditions.

*3.3 Study 3*

3.3.1 Methods

*Participants*. 305 participants were recruited on Amazon’s Mechanical Turk.[[14]](#footnote-14) We calculated sample size with an *a priori* power analysis using G\*Power. To detect effects similar to those observed in Study 1 (*f* = .22), for an ANOVA test to have 80% power at standard error thresholds (*p* = .05), 230 participants were recommended. Because we planned to use stricter comprehension checks, we over-recruited by 35% (*N* = 305). We excluded participants who failed a simple attention check (“If Monday is the first day of the week, what is the third day of the week?”). Participants also needed to pass a strict comprehension check (described below). Additionally, we excluded any participants who completed the survey in less than half of the median time for the entire sample (242 seconds). All exclusion criteria were pre-registered. 56 participants were excluded based on the comprehension check and 26 participants were excluded for completing the survey too quickly (*N* = 223, *M*age = 36.93, *SD* = 10.4, 39% female).

3.3.2 Materials and procedure

Participants were randomly assigned to one of 6 conditions. They saw either of two vignettes used in the previous study. In one (Achievement), the primed behavior was bad, while in the other (Gods) the behavior was good. We included 3 versions of each vignette, where the difference between behavior in the control condition and the experimental condition is either high (92% to 21%), moderate (52% to 21%) or low (32% to 21%). After reading the vignette, participants were asked what percentage of participants in the [other] condition cheated or helped. Participants were excluded if they answered incorrectly.

 After reading the vignette and answering the comprehension check, participants responded to items about free will, responsibility, reliability, and awareness just as in Study 2.

3.3.3 Results

Means and standard deviations for dependent variables in Study 3 are described in **Table 2**.

*Table 2. Counts, means, and SDs for dependent variables across condition in Study 3.*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Vignette** | **Probability** | **Count** | **Free** | **Responsible** | **Aware** | **Reliable** |
| *Achieve* | High | 35 | 7.17 (*1.8*) | 6.51 (*2.4*) | 5.00 (*2.6*) | 6.26 (*1.9*) |
| *Achieve* | Moderate | 38 | 7.37 (*1.2*) | 6.39 (*2.1*) | 5.16 (*2.5*) | 6.66 (*1.5*) |
| *Achieve* | Low | 34 | 6.82 (*2.0*) | 6.71 (*1.9*) | 4.91 (*2.4*) | 6.15 (*2.0*) |
| *Gods* | High | 39 | 7.62 (*1.5*) | 6.95 (*1.3*) | 7.21 (*1.7*) | 5.56 (*2.9*) |
| *Gods* | Moderate | 38 | 7.21 (*1.2*) | 6.54 (*1.7*) | 6.85 (*1.6*) | 5.41 (*2.5*) |
| *Gods* | Low | 38 | 7.45 (*1.4*) | 6.47 (*2.0*) | 6.84 (*1.7*) | 5.89 (*2.5*) |

There was no evidence for an effect of probability on judgments of reliability (*F*(2, 220) = 0.08, *p* = .93, h2p < .001).

To test for an effect of probability on attributions of free will or responsibility, separate linear models were fit to predict these judgments from probability, vignette, awareness, and reliability, as well as all possible interactions. There was no evidence for an effect of probability on judgments of free will (*F*(2, 220) = 0.29, *p* = .75, h2p = .003, 90% *CI*[.00, .02]) or responsibility (*F*(2, 220) = 0.14, *p* = .87, h2p = .001, 90% *CI*[.00, .01]) (see **Figure 4**). A Bayesian ANOVA found strong evidence that there was no effect of probability on judgments of free will (BF01 = 13.78) or responsibility (BF01 = 14.93) and no interaction between vignette and probability (BF01 = 139.49 and 715.78, respectively).



*Figure 4.* Judgments of free will (left panel) and responsibility (right panel) by probability and vignette. Error bars represent standard errors.

3.3.4 Discussion

Participants showed no sign of altering judgments of free will or responsibility in response to stronger primes, even when the prime was described as having affected nearly every participant in the experimental condition. Notably, even when primes were described as being uncharacteristically strong, this did not seem to diminish the overall perceived reliability of these studies. This suggests that merely altering the perceived strength of the prime is insufficient to dislodge the commonsense tendency to explain agency and responsibility in terms of inner dispositions (and to exhibit the associated biases that can result from this explanatory framework). One final possibility we wanted to test was whether presenting a series of results, rather than the results of a single experiment, might shift participants away from their dispositionist tendencies and explanatory framework. Thus, in the interest of testing this possibility and further replicating our null findings, we conducted another study.

*3.4 Study 4*

3.4.1 Methods

*Participants*. 808 participants were recruited on Amazon’s Mechanical Turk. Sample size was determined by *a priori* power analysis using G\*Power software. For an ANOVA to have 95% power to detect a minimal effect size of interest (*f* = 0.20), 577 participants were recommended. Based on previous experiments, we anticipated an exclusion rate of 40%. 166 participants were excluded for failing to complete the experiment. 51 participants were excluded for failing a basic attention check (“If Monday is the first day of the week, then what is the third day of the week?”). 85 participants saw a different prompt related to exploratory hypotheses (see below). 173 participants were excluded for failing a rigorous comprehension check. And 65 participants were excluded for completing the survey more quickly than half of the median completion time for participants (median time = 156 seconds) (*N* = 268, *M*age = 37.2, *SD*age = 21.9, 51.3% female).

*Materials and procedure*. The same vignettes as Study 3 were used. Participants saw the same items along with 2 new items:

*Cause (experimental)*: Being subliminally primed to think about [condition-dependent clause] caused John to cheat.

*Cause (control):* John decided to [condition-dependent clause]

Participants saw the experimental cause item if they saw a vignette describing a participant in an experimental condition. Otherwise, they saw the control cause item.

 85 participants viewed a different kind of prompt with different items. Half of the participants were instructed to read the following short excerpt from Doris (2015) about the influence of subconscious priming on decision-making:

The extent to which insignificant and unconscious influences act on us is manifest in behavior. Consider "implicit egotism," where folks prefer things associated with them, no matter how inconsequential the associations. In the Name Letter Effect, people favor letters that appear in their own names (Kitayama and Karasawa 1997; Nuttin 1985, 1987), while in the Birth Date Effect, people favor things associated with their birthdays (Kitayama and Karasawa 1997; cf. Finch and Cialdini: 1989: 224 - 6). Implicit egotism may influence moral behavior: Miller and colleagues (1998) found that people were more likely to cooperate in a prisoner's dilemma game when they believed their opposite number shared their birthday while Burger and colleagues (2004: Study 2) found that women donated twice as much to charity when the solicitor wore a name tag indicating that she shared the prospective donor's first name.

This is a suggestive run of laboratory studies, but what happens in real life? Pelham and colleagues (2002) discovered evidence for implicit egotism in public information such as Social Security records, census data, and the membership rolls of professional organizations: across ten studies, the results uniformly indicated that implicit egotism influences major life decisions. For example, "women were about 18% more likely to move to states with names resembling their first names than they should have been based on chance"--36% more likely for the perfect matches, Virginia and Georgia (Pelham et al. 2002:474).

The effect also appears in career choices: Pelham's group (2002: 480) found that men named Geoffrey or George were 42% more likely than expected to be geoscientists, compared to the frequency of names used as controls, such as Daniel, Kenneth, and Bennie. It's not just geoscientists: you might think that your dentist is in it for the money or to help others, but would you have thought that the decision to spend his professional life poking about in the fetid maws of an ungrateful public had something to do with the fact that his name is Dennis (Pelham et al., 2002: 479 - 80)? Where you live and work matter: too much, one might think, to let the decision be influenced by the letters of your name.

The other participants saw a short excerpt from Doris (2015) on the importance of empirically-informed approaches to ethics. After reading the prompt, participants answered general questions about agency:

*Free*: In general, people have free will.

*Cause*: What people do is almost entirely a matter of their conscious decisions.

*Responsibility*: In general, people are blameworthy or praiseworthy when they do good or bad things.

*Character*: In general, people act in a way that reflects their character.

*Reliable*: Findings from psychology studies like the ones mentioned above are reliable.

*Aware*: In general, people are aware of why they are doing what they do.

Participants responded using a 9-pt. Likert scale (1 = Strongly disagree, 5 = Unsure, 9 = Strongly agree).

3.4.2 Results

Descriptive statistics are summarized in **Table 3**.

*Table 3*. Means (SD) for all variables in Study 4

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Vignette** | **Condition** | **Count** | **Free** | **Responsible** | **Cause** | **Character** | **Reliable** | **Aware** |
| *Achievement* | *Control* | 29 | 7.00 (*1.3*) | 6.59 (*1.4*) | 7.07 (*1.2*) | 6.14 (*1.9*) | 6.86 (*1.1*) | 6.97 (*1.5*) |
| *Achievement* | *Experiment* | 17 | 6.53 (*1.2*) | 7.00 (*1.1*) | 7.00 (*1.1*) | 6.74 (*1.4*) | 7.11 (*0.9*) | 6.84 (*1.4*) |
| *Gods* | *Control* | 18 | 7.06 (*1.8*) | 7.44 (*1.5*) | 7.50 (*1.2*) | 6.72 (*2.0*) | 7.11 (*1.3*) | 7.28 (*1.5*) |
| *Gods* | *Experiment* | 26 | 7.15 (*1.3*) | 7.12 (*1.3*) | 6.73 (*1.5*) | 7.15 (*1.0*) | 7.08 (*1.4*) | 6.54 (*1.8*) |
| *Jury* | *Experiment* | 39 | 7.18 (*1.4*) | 6.41 (*1.7*) | 6.69 (*1.9*) | 6.74 (*1.6*) | 6.49 (*1.8*) | 6.38 (*2.1*) |
| *Organ* | *Experiment* | 40 | 7.03 (*1.7*) | 6.48 (*1.8*) | 7.13 (*1.4*) | 6.88 (*1.8*) | 7.00 (*1.1*) | 7.43 (*1.0*) |
| *Phone* | *Control* | 30 | 6.83 (*1.3*) | 6.90 (*1.5*) | 6.87 (*1.5*) | 6.60 (*1.6*) | 6.77 (*1.1*) | 6.73 (*1.7*) |
| *Phone* | *Experiment* | 21 | 7.43 (*1.2*) | 6.71 (*1.3*) | 6.67 (*2.1*) | 6.71 (*1.3*) | 6.95 (*1.6*) | 6.19 (*2.2*) |
| *Samaritan* | *Control* | 26 | 6.54 (*1.0*) | 5.88 (*1.8*) | 6.65 (*1.2*) | 6.27 (1.4*)* | 6.69 (*1.1*) | 6.31 (*1.6*) |
| *Samaritan* | *Experiment* | 20 | 6.55 (*1.4*) | 6.20 (*1.6*) | 6.30 (*1.7*) | 6.20 (*1.8*) | 7.00 (*1.1*) | 6.70 (*1.4*) |

An ANOVA found no evidence for an effect of vignette (*p* = .32; BF01 = 42.40) or condition (*p* = .65; BF01 = 5.59) on judgments of free will. There was also no evidence for an interaction (*F*(3,216) = 0.88, *p* = .45; BF01 = 319.88). An ANOVA found no evidence for an effect of condition (*p* = .48; BF01 = 6.84) on judgments of responsibility. There was also no evidence for an interaction (*F*(3,216) = 0.46, *p* = .71; BF01 = 52.90). There was evidence for an effect of vignette (*F*(5, 216) = 3.07, *p* = .01) on responsibility, but this was due to participants who viewed the Gods vignette attributing significantly greater responsibility compared to participants who viewed the Samaritan vignette (*t*(216) = 3.53, *p* = .007). An ANOVA also found no evidence for an effect of vignette (*p* = .30; BF01 = 4.49) or condition (*p* = .06; BF01 = 4.01) on judgments of causality. There was also no evidence for an interaction (*p* = .92; BF01 = 184.20; see **Figure 5**).



*Figure 5.* Judgments of free will (left panel), responsibility (middle panel), and character (right panel) by condition and vignette. Error bars represent standard errors.

 To explore how judgments of free will changed as a function of causal judgments across conditions, we fitted a model to predict judgments of free will from causal judgments, condition, and their interaction. We found evidence for an effect of cause (B = 0.52, *p* < .001), with judgments of free will increasing as judgments of causality increased. There was also evidence for an interaction with condition (B = -0.33, *p* = .02). Judgments of free will exhibited greater sensitivity to causal judgments in the Control conditions compared to the Experimental conditions (see **Figure 6**).



*Figure 6.* Relationship between causal judgments and judgments of free will across condition. Error bars represent standard errors.

Finally, when participants read a prompt with more references to the automaticity literature and answered general questions about human agency, there was no evidence for an effect of condition on these general responses (all *p* > .38; BF01 for free will = 3.83; BF01 for responsibility = 4.21; BF01 for character = 4.34; see **Figure 7**).



*Figure 7*. Judgments of free will (left panel), responsibility (middle panel), and character (right panel) based on reading about the results of multiple automaticity studies (*general condition*) or reading about the importance of empirical approaches to ethics (*general no prime* *condition*). Error bars represent standard errors.

3.4.3 Discussion

We again found no evidence for a difference in attributions of free will or responsibility to agents who acted on the basis of situational primes compared to those who acted on conscious decisions. This aligns with the null findings reported in the studies above. We also found no evidence that people view behaviors caused by situational primes as any less indicative of underlying character compared to behaviors caused by conscious decisions.

 The exploratory analysis of the relationship between causal judgments and free will might explain why this is the case. When evaluating behaviors that stem from conscious decisions, participants attributed more free will to an agent the more they saw the relevant behavior as caused by the decision. This pattern did not hold when participants evaluated behavior that stemmed from situational primes. That is, participants who saw the situational prime as causing the relevant behavior judged that behavior to be marginally freer than participants who denied that the situational prime caused the relevant behavior. This suggests that the causality of situational factors is not relevant for attributions of free will, which explains why the results of the experiments are not threatening to folk conceptualizations of free agency.

 Moreover, we examined whether these findings would be more effective when presented as a package of findings rather than one-off results. However, when reading a passage about numerous automaticity findings, participants still did not make different judgments relative to baseline tendencies to attribute freedom and responsibility. Even in the aggregate, people do not consider the results of these studies to be relevant to attributions of free will and responsibility.

1. **General Discussion**

Empirical findings on situationism and automaticity have been at the center of a heated debate at the crossroads of philosophy and psychology about the nature of human agency. While the findings have led some researchers to draw deflationary or even skeptical conclusions, others have defended the metaphysical ramparts from what has famously been called the “unbearable automaticity of being” (Bargh & Chartrand, 1999). We were curious which side of this dispute holds sway in the court of public opinion. So, we had people read descriptions of studies about the impact that environmental stimuli can have on our behavior without our awareness, and we explored whether this influenced attributions of agency and responsibility.

Our preliminary results suggest that these beliefs are more robust than skeptics (including one of the present authors) may have assumed. Using a wide variety of vignettes based on real studies from the relevant empirical literature, we consistently found that judgments of free will, responsibility, and character are insensitive to descriptions of behavior as caused by situational factors of which agents are unaware. While the key evidence for this is a series of null results, we replicated these null findings across multiple experiments. These results suggest that traditional views about human agency can accommodate evidence that our moral behavior is often influenced in surprising ways by trivial features of our environment. Of course, even if people may not be as alarmed by these findings as their more skeptically minded counterparts in philosophy and psychology, it remains an open metaphysical question whether they should be more concerned about the footing of agency than they are!

The proponent of skepticism based on situationism might suggest that these results say nothing about whether situationist experiments threaten (what they take to be) the underlying commitments of commonsense morality such as the importance of character in moral agency or the importance of conscious deliberation in practical reasoning. Our participants might misunderstand the experiments and if they understood them correctly, they would relinquish their commitments. But there are two issues with this response. First, there are substantive disagreements about how to interpret the results of situationist experiments, and not all interpretations align with the skeptical conclusions proposed by some proponents of situationism (see Miller, 2013; Mudrik et al., 2022). Some have argued that the results of situationist experiments could be interpreted as showing that we are *less* free and responsible than we ordinarily suppose (Vargas, 2014; Mudrik et al., 2022). The results would thus challenge ordinary beliefs about the *degrees* of freedom and responsibility people have rather than *whether* people are free and responsible (see Hall & Vierkant, 2022). The results of Study 2, however, provide evidence against this interpretation. When the response scales were adjusted to capture changes in the degree to which people attributed agency and responsibility, there was no evidence that people attributed less freedom and responsibility to agents acting under the influence of a situational cue rather than some internal factor. Thus, while one might maintain that the results of situationist experiments *should* challenge the degree to which we think of ourselves as free and responsible, people do not seem to accept this implication of the results.

This brings us back to the point about participants misunderstanding the experiments. One could insist that participants *should* react to these results by adjusting their beliefs and their failure to do so indicates a misunderstanding of the force of the evidence. Even if it is correct to say that people mistakenly interpret the results, the response implies that the *experimental results* do not threaten the folk conception of autonomy and freedom, but rather that the *results under a particular interpretation* do. But it might be the case that the interpretation makes assumptions about human psychology and decision-making that guides how to understand the results of such experiments. One could then plausibly argue that the assumptions (rather than the experimental results) are what really threaten the folk conception of autonomy and freedom independently of any empirical work (for more on this second reply, see Berker, 2009). In terms of the results reported above, the results of situationist experiments would threaten attributions of free will and responsibility on the assumption that awareness of (some of) the causal factors underlying one’s decisions is essential to free and responsible agency. Given that people do not seem to alter their attributions of agency and responsibility in response to such results, it might be the case that participants fail to understand the force of the evidence. But another plausible interpretation is that participants simply reject whatever assumptions about awareness that make the experimental results threatening. The results of Study 2 align with the alternative interpretation, where participants perceived lower awareness of the agent when the causality of the situational cue was made explicit, but attributed free will and responsibility to similar degrees compared to participants who made judgments about someone acting on the basis of internal factors.

Admittedly, the skeptic can always respond that better materials might prompt shifts in judgment that align with skeptical predictions (Murray, Dykhuis, & Nadelhoffer, 2024). The format in which we presented these results might not be correct for getting people to integrate the information with their underlying beliefs about agency and responsibility. Perhaps audiovisual stimuli or more narrative presentations of empirical results would constitute a greater threat. This response is related to a concern that our discussion depends on interpreting null results, which is a conceptually fraught issue (Harms & Lakens, 2018). We note two things in response to this. First, the skeptic might still find our results informative insofar as they indicate that people do not properly appreciate—from the skeptic’s perspective—the force of the empirical evidence. But the onus is then on the skeptic to determine *how* results should be presented such that people tend to interpret them correctly. Second, the results of supplemental Bayesian analyses suggest that many of our null results support the null hypothesis. That is, in many instances we found moderate-to-strong support for the claim that our manipulations had no effect on judgments of free will, responsibility, or character. Significant effects often failed to replicate in subsequent studies. Thus, the consistent underlying findings are null results. We think this provides good evidence that there is no effect rather than a mere absence of evidence for an effect.

The one exception is the effect of presenting multiple studies in more narrative format to participants. The skeptic might reply that the real evidential force of situationist experiments lies in the aggregate (e.g., Doris, 2015). We tested whether presenting a summary of results across various domains of behavior might affect judgments of free will and responsibility. While we found no evidence for an effect, it is possible that a different mode of presentation might elicit different responses. More work should be done to understand how judgments might change when the mode of presentation is modified to a greater or lesser extent.

Two further issues are worth addressing in future work. In Study 3, we manipulated the strength of the situational prime by altering the number of people who acted congruently with the expected effect of the prime. This assumes that strength corresponds to consistent effectiveness. A different way to manipulate strength is to change the corresponding actions (e.g., a stronger prime might induce people to act *more* charitably or *more* selfishly relative to a weaker prime). People might alter attributions of agency and responsibility in response to this kind of change in prime strength. Additionally, the responsibility item we used in Study 4 for participants reading about a series of studies measured whether people are blameworthy or praiseworthy when they do good or bad things. People might remain committed to this even as their view about appropriate blame or praise changes. For example, people might think that certain kinds of blame (e.g., retributive blame) might not be appropriate in a situation where conscious deliberation plays a limited role in action production. Thus, more sensitive items might be able to capture subtle changes in how people think about agency and responsibility.[[15]](#footnote-15)

Our goal wasn’t to adjudicate the metaphysical debate with empirical evidence about folk intuitions. We wanted to investigate how people think about the range of surprising findings from psychology that some researchers take to challenge or undermine our commonsense understanding of virtue and moral agency. Despite our best efforts to stress test people’s beliefs about agency and responsibility using findings on situationism and automaticity, these beliefs proved to be resilient at least in part because they are bolstered by dispositionist biases that make it hard for people to update their metaphysical and moral beliefs in the face of the findings. Even when we overly exaggerated the strength of the causal influence of environmental stimuli on moral behavior to make sure participants appreciated the role played by these stimuli, they tended to stand firm on their agentic commitments come what may. This suggests that laypersons are less likely to adopt skepticism than their counterparts in philosophy and psychology. It also suggests that laypersons might not share some of the commitments imputed to them by skeptics, thereby making their commonsense views of agency and responsibility less susceptible to threat by experimental findings on situationism and automaticity. As the title of our paper suggests, perhaps the automaticity of being is bearable. But more research needs to be done in this area if we are to have a better understanding of where people draw the line when it comes to the boundaries of agency and responsibility.

**Bibliography**

Adams, R. M. (2006). *A Theory of Virtue: Excellence in Being for the Good*. Oxford University Press.

Alfano, M. (2013). *Character as Moral Fiction*. Cambridge University Press.

Amaya, S., & Doris, J.M. (2015). No excuses: Performance mistakes in morality. In J. Clausen & N. Levy (eds.) Handbook of Neuroethics (Springer), 253 – 272.

Anderson, J. R. (1992). Automaticity and the ACT\* theory. *American Journal of Psychology*, 105, 165-180.

Annas, J. (1999). *Platonic Ethics, Old and New*. Cornell University Press.

Annas, J. (2011). *Intelligent Virtue*, New York: Oxford University Press.

Annas, J. (2005). Comments on John Doris’ Lack of Character? *Philosophy and Phenomenological Research* 73: 636-47.

Anscombe, G.E.M. (1958). Modern Moral Philosophy. *Philosophy*, 33: 1–19.

Appiah, K. A. (2008). *Experiments in Ethics*. Harvard University Press.

Aristotle. (1994). *The complete works of Aristotle, Oxford Revised Translation*. Ed. J. Barnes. Princeton: Princeton University Press.

Arpaly, N. (2005). Comments on Lack of Character by John Doris. *Philosophy and Phenomenological Research* 73: 643-7.

Athanassoulis, N. (2000). A response to Harman: Virtue ethics and character traits. *Proceedings of the Aristotelian Society*, 100, 215–21.

Bargh, J. A. (1982). Attention and automaticity in the processing of self-relevant information. *Journal of Personality and Social Psychology*, 43, 425-436.

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (2nd ed., pp. 1-40). Hillsdale, NJ: Lawrence Erlbaum Associates.

Bargh, J. A. (1996). Principles of automaticity. In E.T. Higgins &A. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 169-183). New York: Guilford.

Bargh, J. A. (1997). The automaticity of everyday life. In R. S. Wyer, Jr. (Ed.), *The automaticity of everyday life: Advances in social cognition*, Vol. 10, pp. 1–61). Lawrence Erlbaum Associates Publishers.

Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. American Psychologist, 54(7), 462–479. [https://doi.org/10.1037/0003-066X.54.7.462](https://psycnet.apa.org/doi/10.1037/0003-066X.54.7.462)

Bargh, J. A., Gollwitzer, P. M., Lee-Chai, A., Barndollar, K., & Trötschel, R. (2001). The automated will: nonconscious activation and pursuit of behavioral goals. *Journal of personality and social psychology*, 81(6), 1014.

Bateson, Melissa; Callow, Luke; Holmes, Jessica R.; Roche, Maximilian L. Redmond; Nettle, Daniel (2013). Do Images of 'Watching Eyes' Induce Behaviour That Is More Pro-Social or More Normative? A Field Experiment on Littering. *PLOS ONE*. 8 (12): e82055.

Bering, J. M., McLeod, K., & Shackelford, T. K. (2005). Reasoning about dead agents reveals possible adaptive trends. *Human Nature*, 16, 360–81.

Berker, S. (2009). The normative insignificance of neuroscience. *Philosophy & Public Affairs*, 37, 293–329.

Badhwar, N. K. (2009). The Milgram Experiments, Learned Helplessness, and Character Traits. *Journal of Ethics* 13: 257-89.

Besser-Jones, L. (2008). Psychology, Moral Character, and Moral Fallibility. *Philosophy and Phenomenological Research* 76: 310-32.

Beyfus, T. A., Dawson, N. L., Danner, C. H., Rawal, B., Gruber, P. E., & Petrou, S. P. (2016). The use of passive visual stimuli to enhance compliance with handwashing in a perioperative setting. *American Journal of Infection Control*, 44(5), 496–499.

Brink, D. O. (2013). Situationism, responsibility, and fair opportunity. *Social Philosophy and Policy*, 30(1-2), 121-149.

Brody, N. (1988). Personality: In search of individuality. *Psychological Monographs*, 52, 39-55.

Burnham, T. C., & Hare, B. (2007). Engineering human cooperation: Does involuntary neural activation increase public goods contributions? *Human Nature*, 18(2), 88–108.

Dear, K., Dutton, K., & Fox, E. (2019). Do ‘watching eyes’ influence antisocial behavior? A systematic review & meta-analysis. Evolution and Human Behavior, 40(3), 269-280.

Dennett, D. 1984. Elbow Room: The Varieties of Free Will Worth Wanting, Cambridge, Mass: MIT Press.

Doris, J. M. 1998. “Persons, Situations, and Virtue Ethics.” *Nous* 32: 504-30.

Doris, J. M. 2002. *Lack of Character: Personality and Moral Behavior*. New York: Cambridge University Press.

Doris, J. M. 2005. “Precis” and “Replies: Evidence and Sensibility.” *Philosophy and Phenomenological Research* 72: 632-5, 656-77.

Doris, J. M. 2006. “Out of Character: On the Psychology of Excuses in the Criminal Law.” In Hugh La Follette (ed.), *Ethics in Practice*. Oxford: Blackwell Publishing, pp. 474-84.

Doris, J. M. 2009. “Skepticism about Persons.” *Philosophical Issues* 19, Metaethics: 57-91.

Doris, J. M. 2010. “Heated Agreement: Lack of Character as Being for the Good.” *Philosophical Studies* 148: 135-46.

Doris, J.M. (2015). *Talking to Our Selves: Reflection, ignorance, and agency*. Oxford: Oxford University Press.

Doris, J.M. and Stich, S. P. 2005. "As a Matter of Pact: Empirical Perspectives on Ethics." In F. Jackson and M. Smith (eds), *The Oxford Handbook of Contemporary Philosophy*. Oxford: Oxford University Press, pp. 114-52.

Doris, J. M. and Stich, S. P. 2006. "Moral Psychology: Empirical Approaches." Edward N. Zalta (ed.), The Stanford Encyclopedia of Philosophy Winter 2003, <http://plato. stanford.edu/entries/moral-psych-emp/>.

Driver, J. (2001). *Uneasy Virtue*, New York: Cambridge University Press.

Ernest-Jones, M., Nettle, D., & Bateson, M. (2011). Effects of eye images on everyday cooperative behaviour: A field experiment. *Evolution and Human Behaviour*, 32(3), 172–178.

Fathi, M., Bateson, M., & Nettle, D. (2014). Effects of watching eyes and norm cues on charitable giving in a surreptitious behavioral experiment. *Evolutionary Psychology*, 12(5), 147470491401200502.

Haley, K. J., & Fessler, D. M. (2005). Nobody's watching?: Subtle cues affect generosity in an anonymous economic game. *Evolution and Human Behavior*, 26(3), 245–256.

Fischer, J.M. 1987. “Responsiveness and Moral Responsibility,” in Schoeman (ed.) 1987, 81–106.

Fischer, J.M. 1994. *The Metaphysics of Free Will*, Oxford: Blackwell Publishers

Fischer, J.M., & Ravizza, M. (1998). *Responsibility and control: A theory of moral responsibility*. Cambridge: Cambridge University Press.

Flanagan, O. (1991). *Varieties of moral personality: Ethics and psychological realism*. Cambridge MA: Harvard University Press.

Flanagan, O. (2009). Moral science? Still metaphysical after all these years. In D. Narvarez & D. Lapsey (Eds.), *Personality, identity, and character: Explorations in moral psychology* (pp. 52–78). Cambridge: Cambridge University Press.

Foot, P. (1978). *Virtues and Vices*, Oxford: Blackwell.

Giraud, T., & Cova, F. (2024). Educated folk intuitions about free will and determinism: a case study in experimental public philosophy. *Philosophical Psychology*, 1-30.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, 102, 4-27.

Hall, J., & Vierkant, T. (2022). Does free will come in degrees?. In *Free Will: Philosophers and Neuroscientists in Conversation*. Oxford University Press.

Harman, G. (1999a). Moral philosophy and linguistics. In K. Brinkmann (Ed.), *Proceedings of the 20th World Conference of Philosophy: Vol. 1: Ethics* (pp. 107–15). Bowling Green, OH: Philosophy Documentation Center.

Harman, G. (1999b). Moral philosophy meets social psychology: Virtue ethics and the fundamental attribution error. *Proceedings of the Aristotelian Society*, 99, 315–31.

Harman, G. (2000). The nonexistence of character traits. *Proceedings of the Aristotelian Society*, 100, 223–6.

Harman, G. (2001). Virtue ethics without character traits. In A. Byrne, R. Stalnaker, & R. Wedgwood (Eds.), *Fact and value* (pp. 117–127). Cambridge: MIT Press.

Harman, G. (2003). No character or personality. *Business Ethics Quarterly*, 13, 87–94.

Harman, G. (2009). Skepticism about character traits. *The Journal of Ethics*, 13, 235–242.

Harms, C. & Lakens, D. (2018). Making 'null effects' informative: statistical techniques and inferential frameworks. *Journal of clinical and translational research*, *3*(Suppl 2), 382–393.

Hursthouse, R. (1999). *On virtue ethics*. New York: Oxford University Press.

Hursthouse, R., 1996, 'Normative virtue ethics', in *How Should One Live? Essays on the Virtues*, ed. R. Crisp (Oxford: Oxford University Press), pp. 19-36.

Kamtekar, R. (2004). Situationism and virtue ethics on the content of our character. *Ethics*, 114, 458–91.

Kelsey, C., Grossmann, T., & Vaish, A. (2018). Early reputation management: Three-year-old children are more generous following exposure to eyes. *Frontiers in Psychology*, *9*, 698.

Kelsey, C., Vaish, A., & Grossmann, T. (2018). "Eyes, More Than Other Facial Features, Enhance Real-World Donation Behavior". *Human Nature*. 29 (4): 390–401.

Knobe, J. and Leiter, B. 2007. "The Case for Nietzschean Moral Psychology.” In B. Leiter and N. Sinhababu (eds), *Nietzsche and Morality*. Oxford: Oxford University Press, pp. 83-109.

Kozuch, B., & Nichols, S. (2011). Awareness of Unawareness Folk Psychology and Introspective Transparency. *Journal of Consciousness Studies*, *18*(11-12), 135-160.

Kupperman, J. (2001). The indispensability of character. *Philosophy*, 76, 239–250.

Latané, B., & Darley, J. (1970). *The unresponsive bystander: Why doesn’t he help?* New York: Appleton-Century Crofts.

Lin, P. K., & Suárez, L. (2020). The effect of supernatural priming on cheating behaviour. *Religions*, 11(6), 315.

Machery, E. 2010. “The Bleak Implications of Moral Psychology.” *Neuroethics* 3: 223-31.

MacIntyre, A. (1957). Determinism. *Mind*, 66: 28–41.

McKenna, M. 2013. “Reasons-Responsiveness, Agents, and Mechanisms.” In D. Shoemaker (ed). *Oxford Studies in Agency and Responsibility*, vol. 1. Oxford University Press: 151–84.

McKenna, M., & Warmke, B. (2017). Does situationism threaten free will and moral responsibility? *Journal of Moral Philosophy*, 14, 698-733.

Mele, A. (2014). *Free: Why Science Hasn’t Disproved Free Will*. New York: Oxford University Press.

Mele, A. & Shepherd, J. (2013). Situation and Agency. *Journal of Practical Ethics* 1: 62–83.

Merritt, M. (2000). Virtue ethics and situationist personality psychology. *Ethical Theory and Moral Practice*, 3, 365–83.

Merritt, M., Doris, J., & Harman, G. (2010). Character. In J. Doris (Ed.), *The moral psychology handbook* (pp. 355–401). Oxford: Oxford University Press.

Merritt, M. 2009. “Aristotelean Virtue and the Interpersonal Aspect of Ethical Character.” *Journal of Moral Philosophy* 6: 23-49.

Milgram, S. (1974). *Obedience to authority*. New York: Harper and Row.

Miller, C. (2003). Social psychology and virtue ethics. *The Journal of Ethics*, 7, 365–92.

Miller, C. (2008). Motivational internalism. *Philosophical Studies*, 139, 233–55.

Miller, C. (2009). Social psychology, mood, and helping: Mixed results for virtue ethics. *The Journal of Ethics*, 13, 145–73.

Miller, C. (2013). *Moral character: An empirical theory*. Oxford: Oxford University Press.

Miller, C. (2014). *Character and moral psychology*. Oxford: Oxford University Press.

Miller, C. (2017). *The character gap: How good are we?* New York: Oxford University Press.

Montmarquet, J, (2003). Moral character and social science research. *Philosophy*, 78, 355–368.

Morey, R. D., & Rouder, J. N. (2018). *BayesFactor: Computation of Bayes Factors for Common Designs*. Retrieved from https://CRAN.R-project.org/package=BayesFactor

Morse, S. J. (2004). Reason, results, and criminal responsibility. *U. Ill. L. Rev*., 363.

Morse, S. J. (2008). Psychopathy and criminal responsibility. *Neuroethics*, 1(3), 205-212.

Mudrik, L., Arie, I. G., Amir, Y., Shir, Y., Hieronymi, P., Maoz, U., O'Connor, T., Schurger, A., Vargas, M., Vierkant, T., Sinnott-Armstrong, W., & Roskies, A. (2022). Free will without consciousness? *Trends in Cognitive Sciences, 26*(7), 555 - 566. <https://doi.org/10.1016/j.tics.2022.03.005>

Murray, S., Dykhuis, E. & Nadelhoffer, T. (2024) Do People Understand Determinism? The Tracking Problem for Measuring Free Will Beliefs. In J. Knobe and S. Nichols (eds), Oxford Studies in Experimental Philosophy, vol. 5. Oxford (pp. 120 – 151). <https://doi.org/10.1093/9780198918905.003.0007>

Murray, S., & Nadelhoffer, T. (2023). Not What I Expected! Feeling of Surprise Differentially Mediates Effect of Personal Control on Attributions of Free will and Responsibility. *Review of Philosophy and Psychology*, 1-25.

Nahmias, E. 2007. “Autonomous Agency and Social Psychology.” In Maraffa, De Caro, and Ferriti, Eds., *Cartographies of the Mind*, Vol. 4, Dordrecht: Springer: 169–85.

Neely, W., 1974. “Freedom and Desire,” *Philosophical Review*, 83: 32–54.

Nelkin, D.K. (2005). Freedom, Responsibility, and the Challenge of Situationism. *Midwest Studies in Philosophy* 29: 181–206.

Nelkin, D.K. (2011). *Making Sense of Freedom and Moral Responsibility*. Oxford University Press.

Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.

Nisbett, R. & Ross, L. (1991). *The person and the situation: Perspectives of social psychology*. New York: McGraw-Hill.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84, 231-259.

Northover, S. B., Pedersen, W. C., Cohen, A. B., & Andrews, P. W. (2017). Artificial surveillance cues do not increase generosity: Two meta-analyses. *Evolution and Human Behavior*, *38*(1), 144-153.

Pfattheicher, S., Strauch, C., Diefenbacher, S., & Schnuerch, R. (2018). A field study on watching eyes and hand hygiene compliance in a public restroom. *Journal of Applied Social Psychology*, 48(4), 188-194.

Plato. (1997). *Plato, Complete Works*. Ed. J. Cooper. Indianapolis: Hackett Publishing Co.

Pink, T. (2009). Reason, voluntariness, and moral responsibility. In L. O'Brien & M. Soteriou, *Mental Actions* (Oxford: Oxford University Press), 95 – 120.

Ross, L. (1977). The intuitive psychologist and his shortcomings. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10). New York: Academic.

Ross, L. and Nisbett, R. (1991) *The Person and the Situation* (Philadelphia: Temple University Press).

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, 16(2), 225–237. https://doi.org/10.3758/PBR.16.2.225.

Russell, D. C. (2009). *Practical intelligence and the virtues*. New York: Oxford University Press.

Sabini, J., & Silver, M. (2005). Lack of character? Situationism critiqued. *Ethics*, 115, 535–62.

Samuels, S. and Casebeer, W. D. 2005. “A Social Psychological View of Morality: Why Knowledge of Situational Influences on Behaviour Can Improve Character Development Practices.” *Journal of Moral Education* 34: 73-87.

Sarkissian, H. (2010). Minor tweaks, major payoffs: The problem and promise of situationism in moral philosophy. *Philosophers’ Imprint*, 10, 1–15.

Skinner, B.F. (1971) *Beyond freedom and dignity*. Middlesex, UK: Penguin Books.

Slote, M. (1992). *From Morality to Virtue*. Oxford: Oxford University Press.

Slote. M. (2001). *Morals from Motives*, Oxford: Oxford University Press.

Snow, N. E. (2010). *Virtue as social intelligence: An empirically grounded theory*. New York: Routledge.

Solomon, R. (2003). Victims of circumstances? A defense of virtue ethics in business. *Business Ethics Quarterly*, 13, 43–62.

Solomon, R. 2005. "What’s Character Got to Do with It?” *Philosophy and Phenomenological Research* 73: 648-55.

Sreenivasan, G. (2002). Errors about errors: Virtue theory and trait attribution. *Mind,* 111, 47–68.

Sreenivasan, G. (2020). *Emotion and virtue*. Princeton, NJ: Princeton.

Swanton, C. (2003). *Virtue ethics: A pluralistic view*. Oxford University Press.

Taylor, R., (1991) *Virtue Ethics*. Linden Books.

Upton, C. 2009. *Situational Traits of Character: Dispositional Foundations and Implications for Moral Psychology and Friendship*. Lanham, MD: Rowman & Littlefield.

Vargas, M. (2013) “Situationism and Moral Responsibility: Free Will in Fragments,” in A. Clark, J. Kiverstein, and T. Vierkant (eds), *Decomposing the Will*. Oxford: Oxford University Press, pp. 325–49.

Vargas, M. (2014). Reconsidering scientific threats to free will. In W. Sinnott-Armstrong (ed.), *Moral Psychology*, vol. 4. Cambridge, MA: MIT Press.

Vekony, R., Mele, A., & Rose, D. (2021). Intentional action without knowledge. *Synthese*, *199*, 1231-1243.

Velleman, D. (1980). *Practical Reflection*, Princeton University Press.

Vranas, P. (2005). The indeterminacy paradox: Character evaluations and human psychology. *Noûs*, 39, 1–42.

Vrij, A., & Firmin, H. R. (2001). Beautiful thus innocent? The impact of defendants' and victims' physical attractiveness and participants' rape beliefs on impression formation in alleged rape cases. *International Review of Victimology*, 8(3), 245-255.

Wallace, R.J., (1994). *Responsibility and the Moral Sentiments*. Harvard University Press.

Webber, J. (2006a). Virtue, character and situation. *Journal of Moral Philosophy*, 3, 193–213.

Webber, J. (2006b). Character, Consistency, and Classification. *Mind* 115: 651-8.

Webber, J. (2007a). Character, Common-sense, and Expertise. *Ethical Theory and Moral Practice* 10: 89-104.

Webber, J. (2007b). Character, Global and Local. *Utilitas* 19: 430-4.

Wolf, S. (1990). *Freedom within Reason*. Oxford: Oxford University Press.

1. \* SM and TN were involved in conceptualizing and designing the study. SM collected and analyzed data. SM and TN were equally involved in drafting, editing, and revising the manuscript. Materials, data, additional analyses, and code can be found at the OSF repository for the project: https://osf.io/uatnw/. [↑](#footnote-ref-1)
2. For meta-analyses and discussion of some failed replications of the Watching Eye Effect, see Dear, Dutton, and Fox (2019) and Northover et al. (2017). Dear et al. found a large effect of the presence of eyes on antisocial behavior and no evidence for publication bias. [↑](#footnote-ref-2)
3. “If practical reasoning is an inferential process, can it be unconscious? There is an important ambiguity here. The question could be whether practical reasoning can occur without *manifestations in consciousness*, or whether it can occur without *S’s consciousness of it* as reasoning…[T]he inferential process conception [of practical reasoning] rules out unconsciousness in the first, *awareness sense*. For *S* must in some way both attent to at least one premise and draw a conclusion from it’ and one must surely be aware of these events” (emphasis original). Audi, however, leaves open the possibility that practical reasoning does not require consciousness in the *recognitional* sense. [↑](#footnote-ref-3)
4. At least some, but likely not all. There might be some neural event that provides a crucial link between action and volition, but nobody would contend that people ever exhibit awareness of the occurrence of specific neural events. Of course, it is possible that someone might reason from the requirement that people be aware of the causes of their actions to the denial that neural events are causes of actions *because* people cannot be aware of their causal influence. This, however, represents a highly idiosyncratic conception of causation in action. [↑](#footnote-ref-4)
5. A related claim that is not relevant for our purposes here is that the *contents* of conscious deliberation are typically transparent to introspection (Kozuch & Nichols, 2011). [↑](#footnote-ref-5)
6. The automaticity challenge exposes an important ambiguity about the role of consciousness in practical reasoning. Audi (2006) distinguishes between the manifestations of the contents of reasoning in consciousness (the awareness sense) and recognizing some reasoning process *as* reasoning (the recognition sense). An automatic process might bypass consciousness in the recognition sense without necessarily bypassing consciousness in the awareness sense (where the contents of reasoning are manifested in consciousness, but the process is not seen *as* a process of reasoning). [↑](#footnote-ref-6)
7. “…people do expect that we *typically* have access to certain mental processes…[I]t turns out that *decision-formation* is treated as especially available to introspection, at least as compared to urges and associations. The last experiment suggests further that people also expect that the decision to act is available to consciousness before the initiation of behavior” (Kozuch & Nichols, 2011: p. 154). In one experiment, Kozuch & Nichols asked about the Watching Eye Effect and whether people know that they gave more money because eyes were present. Based on our analysis of data provided in the paper, a one-sample *t*-test provided no evidence for a statistically significant difference between the theoretical midpoint (*M* = 3) and participant responses (*M* = 2.71, *SD* = 1.64, *n* = 49, *p*= .22, *d* = -0.18). [↑](#footnote-ref-7)
8. The disavowal prediction stems from the idea that people tend to think that conscious deliberation, rather than situational influence, causes behavior. This is likely exaggerated in the case of praiseworthy behavior, where people are particularly motivated to see *good* behavior as a manifestation of underlying character traits rather than the product of situations. [↑](#footnote-ref-8)
9. Based on Bargh, Gollwitzer, Lee-Chai, Barndollar, and Trötschel (2001). Other vignettes were based on studies reported in Bering, McLeod, & Shackelford (2005); Lin, & Suárez (2020). [↑](#footnote-ref-9)
10. We asked about reliability to assess whether people trusted the results of the study (see Murray & Nadelhoffer, 2023). A different concern is whether these results generalize to real-world situations. We address this issue in Study 4 by using more naturalistic vignettes. [↑](#footnote-ref-10)
11. This might seem to miss the point. The situationist might claim that how people understand the experiments is confused. If they understood them *correctly*, however, they would provide responses that support the original hypotheses. We discuss this objection in the General Discussion. [↑](#footnote-ref-11)
12. A similar study was conducted using a sample collected on Amazon’s Mechanical Turk (see *Supplementary materials* §2). We collected a sample on Academic Prolific to address concerns about data quality in MTurk samples (Giraud & Cova, 2024). [↑](#footnote-ref-12)
13. This means 43 participants were excluded from the analyses on judgments of free will, responsibility, and character inference. Similar to results we report below, when looking at the full data set there is still no evidence for an effect of motive attribution on judgments of free will (*F*(1, 617) = .02, *p* = .88, h2p < .001, 90% *CI*[.00, .00]) or character (*F*(1, 617) = 2.72, *p* = .10, h2p < .001, 90% *CI*[.00, .01]). However, we found a small effect of motive attribution on judgments of responsibility (*F*(1, 617) = 10.68, *p* = 001, h2p = .002, 90% *CI*[.00, .01]). This makes sense, as participants who attribute a higher degree of awareness are more likely to attribute greater responsibility without necessarily attributing greater free will or seeing greater diagnosticity of behavior (see Murray & Nadelhoffer, 2023 for similar findings). [↑](#footnote-ref-13)
14. Despite concerns about data quality, the results of Study 2 suggest that responses were similar from participants recruited from Prolific and Mechanical Turk. However, the exclusion rate for Mechanical Turk samples was much higher. [↑](#footnote-ref-14)
15. Thanks to an anonymous reviewer for these suggestions. [↑](#footnote-ref-15)