

WHEN A VIRTUE IS NOT A VIRTUE: CONDITIONAL VIRTUES IN MORAL EVALUATION

Jared Piazza
Lancaster University, UK

Geoffrey P. Goodwin, Paul Rozin, and Edward B. Royzman
University of Pennsylvania

Four studies show that people distinguish between two sorts of moral virtues: *core goodness traits* that unconditionally enhance the morality of any agent, and *value commitment traits* that are conditionally good (i.e., that polarize the morality of good and bad agents). Study 1 revealed that commitment traits (e.g., dedicated) amplify the *badness* of a bad agent (terrorist), whereas core goodness traits (e.g., kind) amplify the *goodness* of the bad agent. Study 2 replicated these results while also showing that both commitment and core goodness traits enhance the perceived goodness of neutral and good agents. Studies 2–4 established that commitment traits polarize moral evaluations by signaling agents' commitment to certain values, rather than their agency or effectiveness in pursuing those values. These results extend current understanding of the perceived structure of moral character.

Most people would judge a suicide bomber to be the appropriate target of severe moral condemnation. But, how would they judge him if he also happened to be generally trustworthy and honest? Would they judge him, overall, to be just a little better in this case? Perhaps. But, what about if he also happened to be courageous and dedicated? In this case, it seems much less likely that the suicide bomber would be judged more positively; if anything he is likely to be judged more

This research is an output from a project undertaken as part of The Character Project: New Frontiers in Psychology, Philosophy, and Theology at Wake Forest University, funded by the John Templeton Foundation via a grant to Geoff Goodwin and Paul Rozin. The views expressed are not necessarily those of The Character Project, Wake Forest University, or the John Templeton Foundation.

Correspondence concerning this article should be addressed to Jared Piazza, Fylde College, Department of Psychology, Lancaster University, Lancaster, United Kingdom, LA1 4YF; E-mail: j.piazza@lancaster.ac.uk

negatively because his courage and dedication could amplify his existing badness. These trait terms, *trustworthy*, *honest*, *courageous*, *dedicated*, are all examples of “character traits,” or virtues—they form part of a person’s underlying moral character, and are normally thought to be positive (e.g., Goodwin, Piazza, & Rozin, 2014; Wojciszke, Bazinska, & Jaworski, 1998). But, as the examples just provided suggest, these traits seem to exert strikingly different effects on the moral evaluation of bad actors, pointing toward a fundamental distinction between different kinds of character traits.

The main thesis of the present article is that character traits (or virtues) differ in terms of whether they are evaluated as unconditionally or conditionally good to possess. One sort of virtue, which we call “core goodness traits” are perceived to be absolutely, or unconditionally, good at the trait level, in that possessing this sort of trait enhances the goodness of any agent, good, bad, or neutral. Such traits should include kindness, fairness, and honesty, which encapsulate goals that are, under most circumstances, intrinsically good (e.g., kind acts, greater equality, reliable information), thereby making these traits *unconditionally* good. However, a second sort of virtue, which we refer to as “value commitment traits” (or “commitment traits” for short), should be perceived only to be conditionally good. Such traits, which may include determination, courage, diligence, and dedication, signal an agent’s commitment to a particular set of causes, ends, or values. These traits are “default positive,” in that they enhance the goodness of a good or neutral agent. However, these traits are only conditionally good in that they may also amplify the *badness* of bad agents, in part, by enhancing the perception that such agents are committed to particular immoral values.

In claiming that kindness, honesty, and other core goodness virtues are unconditionally good, we do not mean to imply that kind or honest *acts* are morally preferable across every possible circumstance. For instance, sometimes the dictates of different moral virtues clash, and one must choose which of two core goodness virtues to abide by (e.g., whether to lie to protect someone’s feelings). In cases such as these, it might be seen as morally right to act in contravention of a particular core goodness virtue. Our claim about unconditionality is therefore intended to apply at the level of broad, enduring character traits. In other words, core goodness traits should be seen as unconditionally good for an agent to possess, that is, good regardless of the agent’s other motivations and values (whether moral or immoral).

The distinction between core goodness and commitment virtues captures two important but different facets of the idea that a person can have “good character.” Though the distinction has been seldom observed in moral philosophy (and even less so in psychology), it is not without some precursors. Michael Slote (1983) distinguished between *absolute virtues* that are good without condition or qualification and *dependent virtues* whose virtuosity makes itself known only when accompanied by some pertinent absolute virtue (see also Kupperman, 1991). Drawing on Slote (1983) and Rokeach and Rothman’s (1965) principle of cognitive interaction, Royzman (2000, p. 173) proposed that there are certain commitment-related traits

(e.g., *loyal*) that, though positive in and of themselves, become “corrupted” or “devalued” when added to a target who possesses a negative disposition, rendering the target’s configuration of traits more negative than it otherwise would have been. Finally, Gert (2004) distinguished between “moral virtues,” such as kindness and “personal virtues,” such as fortitude. According to Gert, personal virtues mainly benefit the person in possession of them, which means that personal virtues are essentially amoral in his eyes: Gert notes that having personal virtues “does not rule out acting immorally,” and can “make an immoral person even more dangerous” (p. 78). This position is largely consistent with our theorizing (although we suspect that Gert’s personal values would be seen by many people as comprising part of a person’s moral character, even though such values are not other-regarding).

Thus, the distinction between what we refer to as core goodness and value commitment virtues has been alluded to previously, but it has not yet been subject to empirical test. Accordingly, in the present research, we investigated whether ordinary individuals discriminate between these two classes of moral virtues, by examining their judgments of moral and immoral agents as a function of the traits those agents possess. Before describing our specific studies, we first review existing work within social and moral psychology in order to situate the current inquiry.

PAST RESEARCH ON MORAL CHARACTER

Considerable evidence suggests that moral virtues are a particularly critical aspect of person impressions (Brambilla, Sacchi, Rusconi, Cherubini, & Yzerbyt, 2012; Cottrell, Neuberg, & Li, 2007; Goodwin et al., 2014; Leach, Ellemers, & Barreto, 2007; Peterson & Seligman, 2004; Strohminger & Nichols, 2014; van ’t Wout & Sanfey, 2008; Willis & Todorov, 2006; Wojciszke, 2005; Wojciszke, Bazinska, et al., 1998). People rate traits such as honesty, loyalty, respectfulness, fairness, and kindness as among the most desirable that a person can possess (Goodwin et al., 2014), and such character traits strongly predict global impression formation (see Brambilla, Rusconi, Sacchi, & Cherubini, 2011; Brambilla et al., 2012; Goodwin et al., 2014; Wojciszke & Abele, 2008; Wojciszke, Bazinska, et al., 1998).

But, are there differences among the moral virtues themselves in terms of their role in social evaluation and impression formation? Do all moral virtues have an equal role in this regard? Most existing research has tended to eschew this question, focusing instead on developing taxonomies of virtues (e.g., Linley, Maltby, Wood, et al., 2007; Peterson & Seligman, 2004), identifying different models or trajectories of character development (e.g., Walker & Frimer, 2007), assessing the psychological outcomes of various self-reported virtues (e.g., Park, Peterson, & Seligman, 2004; Toner, Haslam, Robinson, & Williams, 2012), or debating the actual existence of moral character (cf. Doris, 2002; Harman, 1999; Miller, 2003; Vranas, 2005). By contrast, comparatively little is known about the conceptual distinctions

people make with regard to different virtues. One exception to this general trend is a line of research that has applied prototype theory to naive conceptions of moral virtue (e.g., Lapsley & Lasky, 2001; Smith, Smith, & Christopher, 2007; Walker & Hennig, 2004; Walker & Pitts, 1998). This research has revealed that people's concept of "a person with moral character" (or a "good person") is structured in terms of prototypes that consist of both "central" traits, such as *honesty* and *trustworthiness*, that are highly salient and accessible in memory, and "peripheral" traits, such as *cleanliness* and *punctuality*, that are less salient and less accessible.

The current proposal is that people perceive moral virtues to be differentiated along another dimension—conditional versus unconditional—which is different from the central-peripheral dimension uncovered by prototype theory, and that has not previously been explored. We also predict that the underlying reason why some virtues are only conditionally good is that, at least in part, *they signal an agent's commitment to particular values*, which themselves can be either good or bad, moral or immoral. We contend that this feature (value commitment) is the key ingredient in what makes a moral virtue conditional, and in what distinguishes value commitment traits (e.g., *dedicated*, *motivated*) from other agentic traits that are not morally relevant (e.g., *intelligent*).

OVERVIEW OF THE STUDIES

In four studies, we tested the hypothesis that people's moral judgments of others are sensitive to the distinction between core goodness and commitment virtues. We predicted that core goodness traits would enhance the perceived goodness of all agents, good or bad, whereas commitment traits would amplify the perceived goodness of good or neutral agents, but amplify the badness of bad agents. In Study 1, we had participants assess whether adding core goodness or commitment virtues to a bad agent ("a terrorist") would enhance the perceived goodness or badness of the agent. In Study 2, we extended our investigation beyond morally bad agents to include neutral and morally good agents, in order to provide a full test of the hypothesis that commitment virtues are only conditionally good, whereas core goodness virtues are unconditionally good. In this study, we also compared the effects of commitment traits with the effects of intelligence. Intelligence, like commitment, can be seen as enhancing an agent's overall effectiveness, but unlike commitment, intelligence is largely orthogonal to morality (Goodwin et al., 2014; Wojciszke, 2005; Wojciszke, Bazinska, et al., 1998; Wojciszke, Dowhyluk, & Jaworski, 1998), and therefore should not be seen as an indicator of the agent's underlying moral values. Accordingly, intelligence should not have as great a polarizing effect as commitment. In Studies 3–4, we directly compared these two potential mediators of the effects of commitment traits on moral character evaluations: value commitment and perceived effectiveness.

STUDY 1

METHOD

Participants. We collected data from 396 participants, using Amazon's Mechanical Turk platform (www.mturk.com; see Buhrmester, Kwang, & Gosling, 2001) in exchange for payment; 75% of the sample was male, with a mean age of 26.25 years ($SD = 8.08$). Data collection was limited to U.S. residents.

Design, Materials, and Procedures. Participants judged whether 16 different traits made a terrorist (our bad agent) better or worse, compared to a terrorist without any modifying trait. We used a 2 (*phrasing of question*: "who is a better person?" vs. "who is a worse person?") \times 2 (*order of presentation*: a trait-modified terrorist first vs. a non-modified terrorist first) \times 16 (*traits*) mixed-measures design. Phrasing and order were between-subjects variables, and traits was a within-subjects variable. Thus, for example, participants who were assigned to the condition in which the trait-modified terrorist was presented first, and who were presented with a question phrased in terms of "who is a better person?" received the following instructions: "Please indicate who is a better person: a dedicated terrorist, or a terrorist?" Participants selected one of the two options. The dependent variable was, therefore, categorical: which individual the participants chose as the "better" or "worse" person.

We selected 16 positively valenced traits that in a prior study (Goodwin et al., 2014, Study 1, $N = 1,048$) all scored above $M = 7.00$ on their "usefulness for identifying whether or not a person has admirable character," rated on a 1 (*Not at all useful*) to 9 (*Extremely useful*) scale. We thought this character index was sufficiently broad to capture character traits that are relevant to both core goodness and value commitment. We selected 7.00 as our cut-off point as it was half way between the scale mid-point and its extremity: if a trait's average rating was at or above 7.00, we viewed this as a reasonable indication that it was seen as instantiating the dimension of interest. Within this larger set of 16 traits, we deliberately selected traits that we expected to vary along the dimensions of value commitment and moral goodness. In the prior study just referenced, we did not have an index of the traits' direct relevance to value commitment, but we did have a measure of their relevance for judging the related construct of a person's strength of will (i.e., "usefulness for identifying how strong-willed or weak-willed a person is"). We thought strength of will was a reasonable proxy for commitment, and therefore selected traits for the present study partly on the basis of their relevance for judging strength of will.¹ We

1. Many, if not most, traits that are relevant for judging an agent's value commitment are also highly relevant for judging their strength of will (e.g., *dedicated, committed, motivated*, seem highly relevant to both constructs). The concept of strength of will is perhaps slightly broader than the concept of commitment, in that it includes some traits that are not highly relevant to commitment to a particular cause or value, but that are instead more self-focused (e.g., *self-controlled, disciplined*). Nevertheless, given that the purpose of this first study was to find out whether some moral character traits do indeed have conditional effects on overall judgments of moral character, the fact that this initial sorting of traits into core goodness and commitment bins is imperfect is not of major concern.

did have a direct measure of the trait's usefulness for identifying how good or bad a person is, and we used this as our index of core goodness. Accordingly, we used these indices to select commitment and core goodness traits in the follow ways. Commitment traits scored above $M = 7.00$ on character, below $M = 7.00$ on goodness, and above $M = 7.00$ on strength of will. (Note that there are a number of traits that perceivers rate high on agency, but low on character relevance, e.g., *intelligent*, *confident*, etc.; these low-character, high-agency traits should not be confused with commitment virtues, and will be the subject of comparison in Studies 2–3). This gave us a list of seven putative commitment traits: *committed*, *courageous*, *dedicated*, *determined*, *disciplined*, *hardworking*, and *motivated*² ($M_{\text{character}} = 7.35$, $SD = 0.33$; $M_{\text{goodness}} = 5.91$, $SD = 0.94$; $M_{\text{strength}} = 7.63$, $SD = 0.35$). Core goodness traits scored above $M = 7.00$ on character, above $M = 7.00$ on goodness, but below $M = 7.00$ on strength of will. This gave us a list of eight putative core goodness traits: *charitable*, *honest*, *humble*, *just*, *kind*, *loyal*, *respectful*, and *trustworthy* ($M_{\text{character}} = 7.89$, $SD = 0.32$; $M_{\text{goodness}} = 7.78$, $SD = 0.34$; $M_{\text{strength}} = 5.38$, $SD = 0.73$). The set of core goodness traits scored higher on goodness, $t(13) = 5.22$, $p < .001$, and lower on strength of will, $t(13) = -7.35$, $p < .001$, than did the set of commitment traits, thereby corroborating that the trait sets reliably differed in the anticipated way. The core goodness traits also scored somewhat higher on character, $t(13) = 3.19$, $p < .01$, than did the commitment traits, though the means were well above 7.00 on character for both trait sets, consistent with the assertion that core goodness traits and commitment traits represent two different aspects of moral character. We also included *wise* in our sample of traits, but did not classify it, since it scored below $M = 7.00$ on both goodness and strength of will (5.41 and 5.81, respectively). Nevertheless, "wise" was included in the study for exploratory purposes, given its high character rating ($M = 7.47$).

In the main study, the 16 traits were paired with "terrorist" across 16 trials, in a randomized order without replacement. After participants had completed the 16 trials, they provided demographic information and were debriefed and paid. No other measures were collected.

RESULTS AND DISCUSSION

There was only a negligible effect of presentation order (i.e., whether the trait-modified option was presented first or second) on which option participants selected. Since these order effects occurred for only two out of the 16 traits (*humble*, *courageous*), order was dropped from further analysis. Averaging across the eight core goodness traits, the phrasing of the question (i.e., "who is a better person?" vs. "who is a worse person?") had a marginal effect on selection rates, Mann-Whitney U -test, $z = 1.91$, $p = .06$. However, the predicted effect occurred in both framings. The core goodness modified terrorist was selected as "better" than the unmodified terrorist 82% of the time, Wilcoxon test, $z = 10.60$, $p < .001$ (against a chance value of 50%), and, similarly, was not selected as being "worse" than the unmodified terrorist 80% of the time, Wilcoxon test, $z = 10.94$, $p < .001$. Follow-up sign tests re-

2. Note that *motivated* had a character rating just under 7.00 ($M = 6.96$).

TABLE 1. Frequency of Participants Selecting “an X terrorist” as a Better or Worse Person than “a terrorist” for Each Trait (X) (Study 1)

	An X terrorist is a better person than a terrorist (%)	An X terrorist is a worse person than a terrorist (%)
Core Goodness Traits		
Charitable	90	14
Honest	90	9
Humble	82	17
Just	80	16
Kind	89	12
Loyal	58	55
Respectful	87	15
Trustworthy	81	23
Mean	82	20
Value Commitment Traits		
Committed	40	76
Courageous	52	47
Dedicated	48	77
Determined	45	76
Disciplined	58	54
Hardworking	57	55
Motivated	46	71
Mean	49	65
Unclassified		
Wise	68	53

Note. All frequencies with bold face are significantly different from 0.50 chance levels at $p < .05$, based on a sign test (for the individual traits) or a Wilcoxon sign test comparing the frequency means against 50% (for the aggregated data).

vealed that, in both framings, this predicted pattern was significant for all but one of the core goodness traits, the exception being *loyal* (see Table 1).

Framing had a more substantial effect for the seven commitment virtues (averaged together), Mann-Whitney U -test, $z = 3.46$, $p = .001$. As predicted, the terrorist was judged worse when in possession of a commitment trait 65% of the time, which was greater than chance, Wilcoxon test, $z = 5.79$, $p < .001$ (see Table 1). Follow-up sign tests revealed that the predicted effect was significant in the “worse” framing for four out of the seven traits (*determined*, *dedicated*, *committed*, *motivated*, see Table 1). But, there was no difference for the “better” framing, in which the modified terrorist was selected as better 49% of the time, which did not differ from chance, Wilcoxon test, $z = .25$, $p = .80$. Follow-up sign tests revealed that the predicted effect was significant in the “better” framing for only one of the seven traits (*committed*, see Table 1). The lack of difference with this framing may owe to the semantic mismatch between the question frame (who is better?) and the valence of the moral evaluation being made (both terrorists are likely to be negatively evaluated), which imposes the need to process a negation (i.e., the committed terrorist is *not* better). Framing also had an effect for the trait, *wise*, Mann-Whitney U -test, $z = 4.18$, $p < .001$, such that a wise terrorist was selected as significantly better than an unmodified terrorist in the “better” framing condition, but not in the “worse” framing condition (see Table 1).

Gender exerted no effect on selections of the core goodness traits for either framing (better: Mann-Whitney *U*-test, $z = .74$, $p = .46$; worse: Mann-Whitney *U*-test, $z = .99$, $p = .32$), nor any effect on selections of the commitment traits (better: Mann-Whitney *U*-test, $z = 1.15$, $p = .25$; worse: Mann-Whitney *U*-test, $z = 1.76$, $p = .08$), or *wise* (better: Mann-Whitney *U*-test, $z = .97$, $p = .33$; worse: Mann-Whitney *U*-test, $z = 1.08$, $p = .28$).

To examine the effects of goodness and strength of will in a more fine-grained way, we next ran additional, by-trait analyses. We correlated ratings of each trait's relevance to "goodness" and "strength of will" from our prior study with the rate at which the modified terrorist was selected as a better person than the unmodified terrorist (collapsing across question framing). As predicted, the goodness index positively correlated with choices of the modified terrorist as the better (or not worse) person, $r(14) = .77$, $p < .001$, reflecting the generally enhancing effect of core goodness traits—as traits reflected goodness more clearly, they were more likely to enhance positive evaluations. In contrast, the strength index negatively correlated with choices of the modified terrorist as the better (or not worse) person, $r(14) = -.87$, $p < .001$, showing that as traits were seen as more clearly related to strength of will, the more they amplified the terrorist's badness. Importantly, when we partialled out the positive valence of each trait (mean valence scores were derived from our pre-study; Goodwin et al., 2014), the goodness index was again positively correlated with choices of the modified terrorist as the better (not worse) person, $r(13) = .74$, $p < .001$, whereas the strength index was negatively correlated with choices of the modified terrorist as the better (not worse) person, $r(13) = -.86$, $p < .001$. Indeed, valence on its own was not significantly correlated with choices of the modified terrorist as the better person, $r(14) = .33$, $p = .21$. These analyses therefore rule out valence as a viable alternative mechanism—it is not simply that commitment virtues are less positive which leads to their conditional effects.

Study 1 therefore demonstrated that individuals distinguish, on a continuum, two classes of moral virtues. At one end are core goodness virtues, such as *honesty*, which, with only one exception, enhanced the goodness of the bad agent (a terrorist). On the other end of the continuum are commitment virtues (e.g., *dedicated*), which, to varying degrees, amplified the *badness* of the bad agent. Although the results were somewhat more mixed for the commitment virtues, and dependent on the valence of the question frame, the commitment traits were generally perceived as enhancing the badness rather than the goodness of a bad agent.

However, because Study 1 only examined bad agents, it is not able to provide a full picture of the proposed conditionality of commitment virtues, and the corresponding unconditionality of core goodness virtues. Accordingly, in Study 2, we studied good agents and neutral agents as well as bad agents. The key prediction is that whereas core goodness traits should enhance the goodness of good, neutral, *and* bad agents (i.e., unconditionally), commitment traits should enhance the goodness of good and neutral agents, but should amplify the badness of bad agents (i.e., they should have conditional or polarizing effects).

STUDY 2

In addition to including neutral and morally good agents alongside morally bad agents, we also sought to investigate whether the conditional morality of commitment virtues applies in a similar way to “intelligence”—a trait seen as being highly related to agency, or the effectiveness of a person when pursuing their goals ($M = 7.43$), but only marginally related to moral goodness ($M = 4.30$; means derived from Goodwin et al., 2014). An “intelligent terrorist,” just as a “committed terrorist,” ought to be perceived as being more effective or agentic than simply “a terrorist,” that is, as being more effective in pursuing their terroristic goals, and as therefore being a worse person (see esp. Abele, Uchrowski, Suitner, & Wojciszke, 2008). And indeed, some prior research has suggested that intelligence (or competence) does amplify the overall negativity of immoral agents (Wojciszke, Bazinska, et al., 1998, Study 4). Yet, intelligence is perceived to be largely orthogonal to morality (Goodwin et al., 2014; Reeder, 1993; Wojciszke, 2005; Wojciszke, Bazinska, et al., 1998; Wojciszke, Dowhyluk, et al., 1998), and, in comparison with moral character traits, it should have fewer implications regarding a person’s underlying value system (i.e., how committed they are to a particular goal or cause). We therefore wanted to investigate how commitment traits and intelligence function as modifiers of good and bad agents. If the reason why commitment traits enhanced the immorality of bad agents in Study 1 stems entirely from their enhancing the perceived effectiveness or agency of such agents, then commitment traits and intelligence should have roughly similar amplifying effects. On the other hand, if the reason why commitment traits amplified the immorality of bad agents stems at least partly from what they signal about the agents’ value systems, beyond what they convey about effectiveness, then commitment traits should have greater amplifying effects on the badness of bad agents than does intelligence. Such a pattern of findings would rule out the possibility that commitment virtues obtain their conditional effects on moral evaluation simply because of what they signal about a person’s agency.

All participants considered four versions of a good, a bad, and a neutral agent. Each agent was paired with a core goodness trait, a commitment trait, intelligence, or no trait, to create the four versions, which participants rank ordered from “best to worst person.” As in Study 1, we predicted that an agent possessing a core goodness trait would be ranked best overall, regardless of the target. In contrast, we predicted that an agent possessing a commitment trait (e.g., “highly motivated”), or intelligence, would be ranked as better than an unmodified agent in the case of good or neutral agents, but as worse than an unmodified agent in the case of bad agents. Moreover, we predicted that the polarizing effects of commitment virtues would be larger than the polarizing effects of intelligence, given that commitment traits signal the agent’s commitment to morally praiseworthy or condemnable values. That is, we predicted that agents possessing commitment virtues would be ranked more positively than those possessing intelligence in the case of good agents, but more negatively in the case of bad agents. However, this difference between intelligence and commitment virtues should vanish for the neutral agent, whose moral values are unknown.

Thus, in sum, the predictions were as follows. For bad agents, in order of increasing goodness, we expected the rankings to be: committed < intelligent < unmodified < good. Whereas, for good agents, in order of increasing goodness, we expected the rankings to be: unmodified < intelligent < committed < good. Finally, for the neutral agent, we predicted: unmodified < intelligent = committed < good.

METHOD

Participants. We collected data from 162 new participants (61 female; $M_{\text{age}} = 32.32$ years, $SD = 12.05$) via the same web service as before, in exchange for payment, excluding those who participated in the previous studies, and limiting participation to U.S. residents.

Pre-Test: Good and Bad Agents. "Person" was used as our neutral agent. To obtain good and bad agents to use in the present study, we conducted a pre-test. Participants ($N = 100$) rated 30 different agents in terms of their moral valence, using a scale ranging from 1 (*Extremely bad*) to 9 (*Extremely good*). We selected two good agents matched on valence: *volunteer aid worker* ($M_{\text{valence}} = 7.85$, $SD = 0.99$) and *human rights activist* ($M_{\text{valence}} = 7.40$, $SD = 1.29$); and two bad agents matched on valence: *Nazi* ($M_{\text{valence}} = 1.65$, $SD = 1.35$) and *terrorist* ($M_{\text{valence}} = 1.40$, $SD = 0.97$). The badness ratings for bad agents tended to be more extreme than the goodness ratings for good agents, perhaps reflecting an instance of negativity dominance (see Rozin & Royzman, 2001).

Design, Materials, and Procedures. Participants were informed that they would be "ranking individuals—from among a list of four individuals—from best to worst," and that they would do this for ten sets of four individuals (which yielded 10 trials in total). Each of these 10 trials concentrated on a single agent, modified in four different ways: by a core goodness trait, a commitment trait, intelligence, and no trait, which participants rank ordered from 4 ("the worst person") to 1 ("the best person"). For example, a participant might have been asked to rank order a "kind Nazi," a "dedicated Nazi," an "intelligent Nazi," and a "Nazi."

The 10 trials resulted from crossing five different possible agents (two good: volunteer aid worker, human rights activist; two bad: Nazi, terrorist; one neutral) with two different trait lists. The first trait list included the terms: *kind* (core goodness), *dedicated* (commitment), *intelligent*, and a final case in which no trait term was used; the second list included the terms: *honest* (core goodness), *highly motivated* (commitment), *intelligent*, and a final case in which no trait term was used. Importantly, the valence scores for the commitment traits (*dedicated* = 7.83, *highly motivated* = 7.93) and *intelligent* (7.94) were equivalent, if not slightly higher for intelligence; thus, any relative difference in the conditional influence of these two trait dimensions cannot be attributed to differences in valence.

Within each trial, the four targets were presented in a randomized starting order, and participants were instructed to rearrange them into their preferred rank order. After completing all ten trials in a randomized order (the design was fully within-subjects), and reporting demographic information, participants were debriefed and paid. No other measures were collected.

RESULTS

Preliminary Analysis of Trait Exemplars and Agent Exemplars. Preliminary analyses were conducted to determine whether we were justified in collapsing across the two trait sets, and across the two exemplars of goodness and badness, respectively. These analyses examined the overall concordance of the ranking patterns across the trait sets, and across the exemplars, and revealed that we were justified in collapsing across both of these variables (see Appendix A).

Main Analysis: Hypothesis Testing. The main analysis tested the ranking patterns across the four traits for each sort of agent. As predicted, bad agents modified by core goodness traits were ranked most highly (1.27), followed by bad agents with no trait (2.44), bad agents modified by intelligence (2.63), and lastly bad agents modified by commitment traits (3.55; see Figure 1). The predicted declining trend was highly robust, according to Page's non-parametric trend test, Page's $L = 4675.00$, $z = 17.01$, $p < .001$. More specific contrasts revealed that core goodness traits improved rankings of the bad agent compared with the unmodified bad agent, Wilcoxon test, $z = 10.20$, $p < .001$, whereas both commitment traits and intelligence decreased rankings of the bad agent compared with the unmodified bad agent, $z = 9.11$, $p < .001$, and $z = 2.62$, $p = .009$, respectively. Moreover, as predicted, commitment traits decreased rankings of the bad agents more so than did intelligence, $z = 9.89$, $p < .001$.³

For good agents, as predicted, agents modified by core goodness traits were ranked most highly (1.76), followed by those modified by commitment traits (1.97), then those modified by intelligence (2.54), and finally, the unmodified agents (3.72), thus also corroborating the predicted trend, Page's $L = 4608.00$, $z = 15.19$, $p < .001$ (see Figure 1). More specific contrasts revealed that good agents with core goodness traits, with commitment traits, and with intelligence, were all ranked higher than unmodified good agents, Wilcoxon test, $z = 10.86$, $p < .001$, $z = 10.46$, $p < .001$, and $z = 9.43$, $p < .001$, respectively. Good agents modified by core goodness traits were ranked significantly higher than good agents modified by commitment traits, $z = 2.42$, $p < .02$. More importantly, and corroborating the prediction, good agents with commitment traits were ranked higher than those with intelligence, $z = 5.89$, $p < .001$.⁴

Finally, the rankings for the neutral agent were also consistent with our predictions, with agents modified by core goodness traits ranked highest (1.39), then

3. Women tended to exemplify this trend more clearly than did men, Mann-Whitney U -test, $z = 3.90$, $p < .001$, though both genders exhibited it (we tested this by using the contrast weighting procedure described in Appendix A). Follow-up Mann-Whitney U -tests showed that men tended to rank bad agents with commitment traits or with intelligence relatively higher than did women (commitment: $z = 3.05$, $p = .002$; intelligence: $z = 2.26$, $p = .024$), whereas women ranked unmodified bad agents relatively higher than did men, $z = 3.53$, $p < .001$. There was no gender difference for bad agents modified by core goodness traits: $z = 0.20$, $p = .85$.

4. Again, women tended to exemplify the predicted trend more clearly than did men, Mann-Whitney U -test, $z = 2.74$, $p = .006$, though both genders exhibited it. Follow up Mann-Whitney U -tests showed that women ranked good agents with core goodness traits relatively higher than did men, $z = 3.58$, $p < .001$, while men ranked the intelligent good agents relatively higher than did women, $z = 1.94$, $p = .053$ (for the other two comparisons, z s < 1.26 , p s $> .20$).

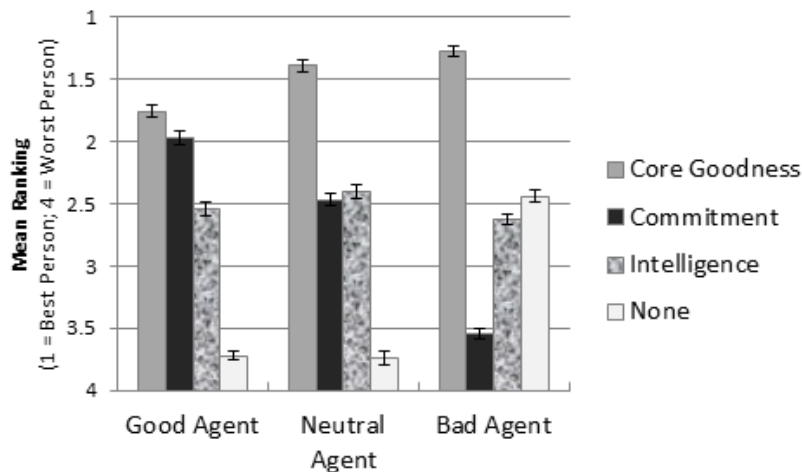


FIGURE 1. Study 2 aggregate rankings of the bad agent, the neutral agent, and the good agent when possessing core goodness traits, commitment traits, intelligence, and no traits. Increasing scores in the figure represent a “better person” (higher goodness ranking). Error bars ± 1 SE (range .04–.058).

agents modified by intelligence (2.40) or commitment traits (2.47) ranked about equally, with unmodified agents ranked lowest (3.74). A trend test revealed that the observed rankings fit the expected pattern, Page’s $L = 2215.00$, $z = 15.06$, $p < .001$ (see Figure 1). More specific contrasts revealed that neutral agents possessing core goodness traits, commitment traits, and intelligence were all ranked higher than unmodified agents, $z = 11.05$, $p < .001$, $z = 9.15$, $p < .001$, and $z = 9.15$, $p < .001$, respectively. Intelligent neutral agents were ranked no differently than neutral agents with commitment traits, $z = 1.14$, $p = .25$.⁵

Given that there are 24 possible ways to rank-order four items, the predicted pattern of results will emerge simply due to chance 4.2% of the time (or 8.3% for the neutral agent, for which we did not predict a difference between intelligence and value commitment). For all three agents the proportion of responses that exhibited the predicted pattern⁶ was well-above chance levels: bad agent (44.4%), binomial test, $p < .001$; good agent (35.2%), binomial test, $p < .001$; neutral agent (57.4%), binomial test, $p < .001$, further corroborating the robustness of the predicted patterns.

DISCUSSION

Study 2 more clearly demonstrated the unconditional nature of core goodness virtues, and the conditional nature of commitment virtues. Core goodness traits,

5. Women also exemplified the predicted trend more clearly than did men, Mann-Whitney U -test, $z = 1.95$, $p = .052$, though both genders exhibited it. Follow up Mann Whitney U -tests showed that women ranked the neutral person with core goodness traits relatively higher than men did, $z = 2.98$, $p = .003$, and men ranked the intelligent neutral person relatively higher than women did, $z = 2.58$, $p = .01$ (for the other two comparisons, z s < 1.09 , $ps > .27$).

6. Unpredicted ties between targets were treated as misses, rather than hits, to provide a conservative estimate.

such as *kind* and *honest*, amplified the perceived goodness of good, bad, and neutral agents. In contrast, commitment traits (e.g., *dedicated*, *motivated*) amplified the goodness of good and neutral agents (thus showing that they are “default positive”), but amplified the badness (and not the goodness) of bad agents. Intelligence showed a similar conditional pattern as did the commitment traits. However, for both good and bad agents, commitment traits had a greater polarizing effect than did intelligence, and this was true despite the equivalent valence scores shared by these traits. Since both commitment virtues and intelligence should both convey positive information about an agent’s effectiveness, the observed difference between them suggests that the polarizing effect of commitment traits is carried by something beyond what they convey about the overall effectiveness of the agent. Our proposed account of this is that commitment virtues convey important information about an agent’s value system, which we test directly via mediation analysis in Study 3.

STUDY 3

The main objective of Study 3 was to isolate more precisely the mechanism underlying the impact of commitment virtues on moral and immoral agents. Since Studies 1 and 2 consistently demonstrated that core goodness virtues, such as *kindness* and *honesty*, are perceived to be unconditionally beneficial, we did not include core goodness traits in the current investigation, but focused exclusively on commitment traits. The account we have proposed is that commitment virtues enhance perceptions of an agent’s relative goodness or badness because they generate inferences about the agent’s underlying value system (i.e., the nature of the agent’s moral convictions, how strongly the agent endorses particular values, etc.). When applied to a moral or immoral agent, commitment virtues may therefore enhance the perception that the agent strongly endorses certain kinds of acts, and it is this perception of deep conviction, not just information about the agent’s overall effectiveness, which mediates judgments of the agent’s moral character.

However, an alternative possibility is that commitment traits are perceived to enhance agents’ goodness or badness via inferences about their overall “agency” (i.e., their effectiveness at achieving their goals). Commitment virtues might increase the perception that an individual is effective, which in turn mediates perceptions of the individual’s overall goodness or badness. One main limitation with this position is that it cannot account for the relative differences that occurred between the commitment traits and intelligence in Study 2. If agency or effectiveness is the correct explanation, and assuming that commitment traits and intelligence convey roughly equivalent information about an agent’s effectiveness (which we can only surmise, without knowing directly from Study 2), then commitment virtues and intelligence should have had roughly equivalent overall effects on judgments of the agents’ morality.

In Study 3, we tested these competing mediation models more directly by presenting participants a good agent (“human rights activist”) and a bad agent (“Nazi”), each possessing a commitment virtue (“dedicated”), intelligence, or no

trait, and having them rate each agent in terms of their overall virtue (activist) or evilness (Nazi), their effectiveness as an agent,⁷ and how strongly they endorse Nazi/human rights values (value commitment). In line with our previous findings, we predicted that a dedicated Nazi would be rated as worse (more evil) than a Nazi or an intelligent Nazi, while a dedicated human rights activist would be rated as better (more virtuous) than a human rights activist or an intelligent human rights activist. More importantly, we predicted that ratings of the agent's endorsement of Nazi/human rights values would mediate the differential effects of commitment traits and intelligence on perceptions of the agent's virtuosity (or evilness), whereas we did not expect ratings of the agent's effectiveness to mediate these effects.

To address the possible criticism that the directly comparative nature of the method used in Studies 1–2 may be susceptible to demand effects, in Study 3, we opted for a non-comparative design, manipulating the trait category between subjects, thus, eliminating any demand.

METHOD

Participants. Participants were 260 adults (95 female) recruited from the same web service, in exchange for payment, excluding previous participants and limited to U.S. residents. The mean age of the sample was 28.85 years ($SD = 10.96$).

Design. We used a 3 (trait: *dedicated* vs. *intelligent* vs. *no trait*) \times 2 (agent: *Nazi* vs. *human rights activist*) mixed-measures design. Trait was a between-subjects variable, while agent was a within-subjects variable. Rather than having participants compare or rank the agents, as we had done in Studies 1–2, in this study we had participants rate each agent's moral character on a continuous scale.

Materials and Procedures. Participants were each asked to rate two different agents, a "Nazi" (bad agent), and a "human rights activist" (good agent), in a randomized order. For each participant, the two agents were each paired with the same trait, either the commitment trait ("dedicated"), "intelligent," or no trait. So, for example, if the participant was asked to imagine a "dedicated Nazi" in the first block, they were also asked to imagine a "dedicated human rights activist" in the second block. Once again, "intelligent" and "dedicated" were equated on valence (7.94 vs. 7.83, respectively), which pre-emptively ruled out valence as an alternative explanation. Participants were instructed to imagine what the person would be like, and then responded to three questions for each agent. The first question was a measure of virtuosity/evilness (virtuosity was assessed for the human rights activist, and evilness for the Nazi), designed to assess the perceived morality/immorality of the agent: "How virtuous [evil] do you think this person would be?" (1 = *Not at all virtuous [evil]*; 6 = *As virtuous [evil] as humanly possible*). The next two measures assessed the potential mediating variables; first was a measure of how strongly the agent was perceived to endorse the relevant morally good or bad values. For the Nazi, participants read: "How strongly do you think this person

7. Our assessment of agency in terms of the perceived "effectiveness of an agent at accomplishing his or her goals" is consistent with past operationalizations of agency (e.g., Abele & Wojciszke, 2007).

endorses Nazi values (e.g., strong anti-Semitism stemming from the conviction that the Aryan master race is superior to all races)?” For the human rights activist, the question read: “How strongly do you think this person endorses human rights values (e.g., strong conviction that there are inalienable fundamental rights to which every person is entitled to simply because he or she is a human being)?” These questions were answered on 1 (*Only weakly*) to 9 (*As strongly as humanly possible*) scales. The final item measured perceptions of the agent’s effectiveness as a Nazi or human rights activist: “How effective do you think this person would be at being a Nazi [human rights activist] (i.e., how effective would they be at accomplishing the goals and plans of a Nazi [human rights activist])?” assessed on 1 (*Not at all effective*) to 9 (*Extremely effective*) scales. Having completed these measures, and reported demographic information, all participants were then debriefed and paid. No other measures were collected.

RESULTS AND DISCUSSION

Preliminary 2 (order) \times 3 (trait) ANOVAs on ratings of the bad agent’s evilness and separately the good agent’s virtuosity, revealed no influence of order on the perceived immorality of the bad agent, and no interaction of order and trait, $F_s < 1.60$, $p_s > .20$. Likewise, order had no influence on the perceived morality of the good agent, and there was no interaction of order and trait, $F_s < 1.22$, $p_s > .29$. Thus, order was dropped from further analysis.

For the good agent, the main effect of trait on morality ratings did not reach statistical levels of significance, $F(2, 257) = 1.04$, $p = .35$, $\eta_p^2 = .008$, though the means were largely in the predicted direction ($M_{\text{commitment}} = 4.24$, $SD = 1.04$; $M_{\text{no trait}} = 4.08$, $SD = 1.10$; $M_{\text{intelligent}} = 4.00$, $SD = 1.20$). Given this nonsignificant main effect for the good agent, we did not pursue further analysis of the potential mediating variables.

For the bad agent, trait had the predicted main effect on ratings of evilness, $F(2, 257) = 4.54$, $p = .01$, $\eta_p^2 = .03$. Planned contrasts of the three trait conditions revealed that, as predicted, a “dedicated Nazi” was perceived as more evil ($M = 4.56$, $SD = 1.21$) than an “intelligent Nazi” ($M = 3.96$, $SD = 1.33$), $t(257) = 3.01$, $p < .01$, and (marginally) more evil than a “Nazi” ($M = 4.24$, $SD = 1.36$), $t(257) = 1.63$, $p = .10$, while an “intelligent Nazi” and a “Nazi” were perceived to be about equally evil, $t(257) = 1.39$, $p = .16$.⁸

Next we analyzed the effect of trait condition on the two potential mediating variables, with regard to the bad agent. For these analyses, we aggregated the intelligent and unmodified Nazi conditions, which together were rated as less evil

8. We examined the effects of gender by conducting a 2 (gender) \times 2 (agent) \times 3 (trait) ANOVA on ratings of the bad agent’s virtuosity/evilness, which revealed a three-way interaction, $F(2, 254) = 3.09$, $p = .047$, $\eta_p^2 = .024$. Subsequent two-way interactions revealed that there was no interaction between gender and trait for virtuosity ratings, $F(2, 254) = .12$, $p = .89$, $\eta_p^2 = .001$. However, gender did interact with trait condition for evilness ratings, $F(2, 254) = 3.93$, $p = .021$, $\eta_p^2 = .030$. The interaction was largely explained by higher baseline evilness ratings in the no trait condition made by women ($M = 4.76$), compared to men ($M = 3.91$). Ratings in the other trait conditions were virtually identical for men and women, respectively (intelligent: $M = 3.96$ vs. 3.97 ; dedicated: $M = 4.63$ vs. 4.41).

than the dedicated Nazi, $F(1, 258) = 7.13, p < .01, \eta^2_p = .03$. The dedicated Nazi was also perceived to endorse Nazi values ($M = 7.64, SD = 1.38$) to a significantly greater degree than the intelligent/unmodified Nazi ($M = 6.06, SD = 2.47$), $F(1, 258) = 30.61, p < .001, \eta^2_p = .11$, but the dedicated Nazi was perceived to be only marginally more effective ($M_{\text{commitment}} = 5.89$ vs. $M_{\text{intelligent/none}} = 5.31$), $F(1, 258) = 2.85, p = .092, \eta^2_p = .01$.

Test of Mediation. To test our mediational hypothesis for the bad agent (Nazi), we conducted a bootstrapping procedure using multiple mediators (Preacher & Hayes, 2008). In the analysis, trait condition (0 = intelligent/no trait, 1 = dedicated) was entered as the independent variable and evilness ratings as the dependent variable, with the two potential mediating variables, value endorsement and effectiveness, as mediators (entered simultaneously in the same model). The results of the bootstrapping procedure revealed that perceptions of the agent's endorsement of Nazi values mediated the effect of trait condition on evilness ratings, while effectiveness ratings had no mediating effect. The direct effect of trait condition on the dependent variable (perceptions of evilness) was not significant with the mediators in the model, $t < 1, p = .68$, while only the indirect effect of the value endorsement ratings on perceptions of evil was significant, $t(256) = 8.77, p < .001$, with 95% bias-corrected and accelerated confidence intervals (CIs) ranging from .314 to .687 (non-overlapping with zero). The indirect effect of effectiveness was not significant, $t(259) = 1.39, p = .17$ (95% CI [-.005, .098]; overlapping zero). Thus, these analyses were consistent with our predictions, and with the mediating role of value commitment: how strongly the Nazi was perceived to endorse Nazi values mediated the effects of the Nazi's dedication on ratings of his evilness, whereas perceptions of his effectiveness did not similarly mediate these effects.⁹ These results thus provide the most direct support for the hypothesis that commitment virtues amplify ratings of bad agents' immorality through what they convey about the agent's value system, and not merely because of what they convey about the agent's effectiveness. They thus bolster the initial evidence provided in support of this hypothesis by Study 2. Nevertheless, the conditional effect only emerged at significant levels for the bad agent. Given this limitation, we conducted one final test of our mediation hypothesis, this time using a richer, vignette methodology and a full between-subjects design.

9. A critic might perhaps be concerned that our "value commitment" mediating variable is closer to a manipulation check than it is to a true mediating variable. That is, it may seem that a Nazi's being committed to Nazi values follows almost deductively from the description of this agent as "dedicated," whereas this is not the case for the corresponding inference about the agent's effectiveness. Indeed, this inference from dedication to value commitment seems all the more watertight when considering an agent with strong ideological commitments, such as a Nazi. We are sympathetic to this argument. Nonetheless, we still thought the mediation analyses reported here clarified the interpretation of the main result. While the inference from dedication to value commitment may be obvious (or deductive), it was an entirely open question whether this inference of value commitment, or alternatively, the inference of effectiveness, best explained why participants subsequently regarded the dedicated Nazi as more evil. And, as the results show, it was value commitment and not effectiveness that carried this effect. A similar point applies to the mediation analyses in Study 4.

STUDY 4

In Studies 1–3, we manipulated virtues exclusively at the level of abstract traits. To demonstrate that our findings generalize to other contexts, in Study 4 we used a vignette methodology that manipulated commitment traits within a more naturalistic description of a target individual. We dropped intelligence traits and focused exclusively on the possession or non-possession of commitment virtues. Furthermore, to improve upon the design used in Study 3, we sought to control for extraneous differences between the good and bad agent. Rather than using good and bad agents who support very different causes, we focused participants' attention upon a single, timely social issue—gun control—and manipulated whether the target supported or was opposed to it. We assumed that gun control would generally be regarded as a good cause.¹⁰ However, because participants likely differed in their opinions of this issue, we first measured their general support for gun control in order to group them according to their prior beliefs. For participants who favored increased gun control, we predicted that the character of the pro-gun-control target would be rated more positively when he displayed commitment virtues, as opposed to when he did not display them, while we expected the inverse to be true of the anti-gun-control target. By contrast, we predicted that participants who favored less gun control would not exhibit this pattern of responses, and might even show the reverse pattern. In short, we predicted a three-way interaction of cause, commitment virtue, and participants' own stance toward gun control. Furthermore, we predicted that perceptions of value commitment, not the agent's effectiveness, would largely mediate the conditional effects of commitment virtues.

METHOD

Participants. We recruited a new sample of 484 adult participants (310 male, 174 female; $M_{\text{age}} = 30.96$ years, $SD = 10.06$) via the same website as before (www.mturk.com), in exchange for payment. The sample was limited to residents of the United States.

Design. We used a 2 (Cause: *Anti-* vs. *Pro-gun-control*) \times 2 (Virtue: *Commitment* vs. *None*) between-subjects design. Participants were randomly assigned to one of these four conditions.

Materials and Procedures. Participants first answered two preliminary questions. The first assessed their position on gun control. Participants were asked, "Are you generally in favor of more or less stringent gun control in America?" and selected between two options: "Generally, I favor more stringent gun control"; "Generally, I favor less stringent gun control." The majority of participants favored more

10. Study 4 was conducted only months after the shooting massacre at Sandy Hook Elementary School in Newtown, Connecticut, and a day after the measure to expand background checks for firearm purchases failed in the Senate by a 54-46 vote on April 17, 2013. According to a Gallup poll conducted soon after, the majority of Americans (65%) voiced that they were in favor of the measure and disappointed with the Senate's ruling against it (Newport, 2013).

stringent gun control ($n = 319$), while a sizable minority favored less gun control ($n = 165$). The second question asked participants whether their stance on gun control changed as a result of the recent shooting massacre at Sandy Hook elementary school in Newtown, Connecticut (“more in favor,” “less in favor,” or “did not change”). This question was included simply to raise the salience of the general issue of gun control rather than for any analytic purpose.

We developed two versions of the vignette in which we manipulated the two, crossed independent variables. Both vignettes described a man named John, who was a gun owner. In the anti-gun-control version, John was described as endorsing a strong anti-gun-control stance, whereas, in the pro-gun-control version, he was described as endorsing a strong pro-gun-control stance. The pro-gun-control vignette paralleled the anti-gun-control vignette in form, while the content was modified to describe a gun owner who was pro-gun-control. In the pro-gun-control version, John was described as aligning his values with the National Gun Victims Action Council, whereas in the anti-gun-control version, John was described as aligning his values with the Gun Owners of America. Subsequently, in a second paragraph we manipulated the commitment virtues variable: John was either described as very determined and dedicated to serving the cause of anti-gun-control (“gun rights”) or pro-gun-control (“gun control”) and did so by volunteering his time at the organization he aligned his values with, or there was no mention of John’s determination or dedication, with the passage simply describing one of his hobbies (see Appendix B for vignettes).

Immediately following the vignette, participants answered two value commitment probes: “How committed do you think John is to serving the cause of gun control [gun rights]?”; “How much do you think John values gun control [gun rights]?” ($\alpha = .77$); one effectiveness probe: “How effective do you think John is at serving the cause of gun control [gun rights]?”; and a manipulation check assessing the extent to which participants viewed John as possessing the following commitment traits: *dedicated*, *committed*, *persistent*, *motivated*, and *determined* (randomly presented). All items were assessed on 1–9 scales (e.g., 1 = *Not at all*; 9 = *Extremely*). These items were followed by two questions assessing perceptions of John’s moral character: “What is your overall impression of how good John is as a person?”; “What is your overall impression of how moral John is as a person?” ($\alpha = .95$), rated on a 0–100 scale (e.g., 0 = *Not at all a good person*; 100 = *An extremely good person*). Finally, one item assessed participants’ overall impression of John: “How positive or negative is your overall impression of John?” on a -100 (*Extremely negative*) to 100 (*Extremely positive*) scale. Afterward, participants reported demographic information, and were debriefed and paid. No other measures were collected.

RESULTS

Manipulation Check. Ratings of the five commitment virtues were internally reliable (Cronbach’s $\alpha = .97$) and were therefore aggregated to form a single index. The manipulation of commitment virtues was significant for both the anti-gun-control target ($M_s = 7.45$ vs. 5.25), $F(1, 239) = 133.51$, $p < .001$, $\eta_p^2 = .358$, and the pro-gun-control target ($M_s = 8.18$ vs. 6.77), $F(1, 241) = 61.50$, $p < .001$, $\eta_p^2 = .203$; that

is, participants rated John as possessing more commitment traits in the commitment virtue condition than in the no virtue condition.

Moral Character Ratings. We conducted 2 (personal stance on gun control) \times 2 (cause) \times 2 (virtue) ANOVAs for each dependent measure. Adding gender to the analysis did not alter the results, and there were no main or interactive effects of gender, $F_s < 2.87$, $p_s > .09$, so we dropped gender from further consideration.

There was a main effect of cause, $F(1, 476) = 55.66$, $p < .001$, $\eta_p^2 = .11$, such that John's character (two-item measure of goodness and character) was rated more positively when he was pro-gun-control ($M = 74.77$, $SD = 16.38$) than when he was anti-gun-control ($M = 56.86$, $SD = 21.95$). There was also a main effect of personal stance, $F(1, 476) = 23.63$, $p < .001$, $\eta_p^2 = .05$, such that, collapsing across participants' own personal stance, those who were themselves in favor of less gun control rated John's character more positively overall ($M = 71.52$, $SD = 17.73$) compared to those who favored more gun control ($M = 62.81$, $SD = 22.43$). A predictable two-way interaction between cause and personal stance emerged, $F(1, 476) = 88.57$, $p < .001$, $\eta_p^2 = .16$, reflecting the fact that pro-gun-control participants rated John's character more positively when he supported gun control than when he did not, whereas the reverse was true (directionally, though not significantly) for anti-gun-control participants. Importantly, these effects were qualified by a significant three-way interaction of personal stance, cause, and virtue, $F(1, 476) = 8.37$, $p = .004$, $\eta_p^2 = .017$. Given the three-way interaction, we separated participants according to their personal stance on gun control, and analyzed each group separately.

Participants who were pro-gun-control rated John as a better person when he was portrayed as pro-gun-control ($M = 77.62$, $SD = 14.17$) than when he was anti-gun-control ($M = 49.51$, $SD = 20.04$), $F(1, 315) = 216.32$, $p < .001$, $\eta_p^2 = .41$; however, simply possessing commitment virtues did not lead to overall better character, $F(1, 315) = 1.92$, $p = .17$, $\eta_p^2 = .006$. Importantly, and as predicted, the main effect of cause was qualified by a significant interaction of virtues and cause, $F(1, 315) = 16.90$, $p < .001$, $\eta_p^2 = .05$ (see Figure 2 for means and standard errors). Follow-up simple-effects tests revealed that, as predicted, possessing commitment virtues increased the perceived moral goodness of the gun-control advocate, $F(1, 149) = 5.24$, $p = .023$, but was detrimental to the perceived moral goodness of the anti-gun-control advocate, $F(1, 166) = 12.34$, $p = .001$.

Surprisingly, participants who were anti-gun-control rated John no better when he was described as anti-gun-control ($M = 73.34$, $SD = 16.48$) than when he was described as pro-gun-control ($M = 70.01$, $SD = 18.66$), $F(1, 161) = 1.36$, $p = .25$, $\eta_p^2 = .008$. Furthermore, there was no main effect of virtue, $F = .06$, $p = .80$, $\eta_p^2 = .00$, nor any interaction of cause and virtue for these participants, $F < 1$, $p = .52$, $\eta_p^2 = .003$. Thus, participants who favored less gun control rated the goodness of the two targets equally, and they failed to exhibit the conditional effect of commitment virtues. The fact that only those participants who favored gun control disparaged the target who disagreed with their position, suggests that participants who did not favor gun control may have held more ambivalent views toward the issue, thus making them a less suitable group for testing our main hypothesis.

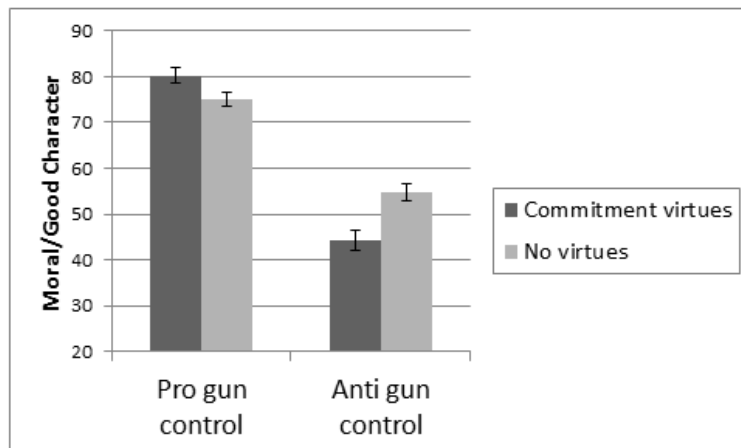


FIGURE 2. Ratings of the moral goodness of the target (John) as a function of the cause he supported, and whether or not he exemplified commitment virtues (Study 4). Analysis includes only participants who are pro-gun-control ($n = 319$). All simple effects are significant ($ps < .03$).

Because the overall impression ratings were rated on a different scale than the moral goodness ratings, they were analyzed separately. However, since the results were identical, regardless of which dependent measure was used, for brevity's sake we report in the text the results only for the moral character ratings.¹¹

Mediation Analysis. In the mediation analyses we focused exclusively on the pro-gun-control participants. We hypothesized that perceptions of value commitment, not effectiveness, should mediate the conditional effects of commitment virtues observed within this group. To test this hypothesis, we entered both mediators simultaneously into a bootstrapping procedure with multiple mediators (Preacher & Hayes, 2008), with virtue as the independent variable and moral goodness ratings as the outcome variable. We conducted these analyses separately for the pro-gun-control and anti-gun-control targets. The results were consistent with our hypothesis. For the anti-gun-control target (i.e., the "bad" agent), as predicted, value commitment emerged as a significant *negative* mediator of the effect of commitment virtues on moral goodness (see Figure 3, top panel). As the agent's value commitment increased, he was regarded as a worse person. In contrast, effectiveness emerged as a significant *positive* mediator (see Figure 3, top panel)—as the

11. Using overall impression scores, the same main effects of cause, $F(1, 476) = 56.44, p < .001, \eta_p^2 = .10$, and personal stance, $F(1, 476) = 26.73, p < .001, \eta_p^2 = .05$, emerged, as did the two-way interaction of cause and personal stance, $F(1, 476) = 160.47, p < .001, \eta_p^2 = .25$. Most importantly, the same significant three-way interaction also emerged, $F(1, 476) = 7.45, p = .007, \eta_p^2 = .15$. As with the character ratings, for pro-gun-control participants, the interaction of cause and virtue condition was significant, $F(1, 315) = 14.04, p < .001, \eta_p^2 = .043$; and global impression means were in the same direction as for moral goodness ratings (see Figure 2)—the pro-gun-control target was rated more positively when he was committed, whereas the anti-gun-control target was rated more negatively when he was committed. For anti-gun-control participants, the interaction of cause and virtue was not significant, $F(1, 315) = 0.53, p = .47, \eta_p^2 = .003$.

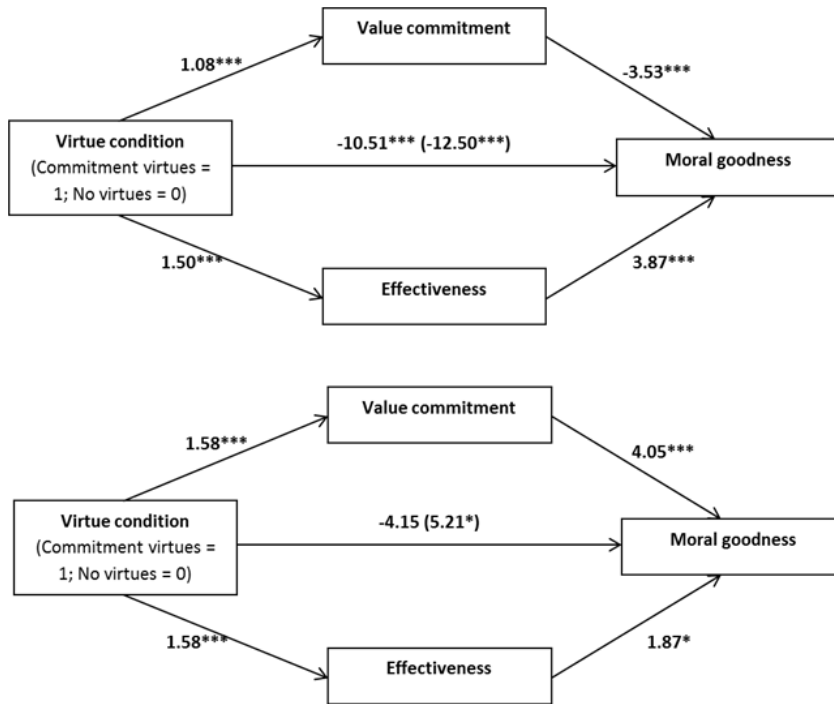


FIGURE 3. Multiple mediation analyses of the effects in Study 4: Analyses are separated by agent and include only participants who are pro-gun-control ($n = 319$). Top Panel: Anti-gun-control target (i.e., "bad" agent). Bottom Panel: Pro-gun-control target (i.e., "good" agent). * $p < .05$; *** $p < .001$.

agent was seen as more effective, he was viewed more *positively*. Perceptions of effectiveness therefore opposed (and partially suppressed) the overall effect of perceptions of value commitment. Confidence intervals for the estimated indirect effect of value commitment did not overlap with zero (95% CI [-7.44, -1.30]). This was also the case for the indirect effect of effectiveness (95% CI [2.92, 10.02]). However, the direct effect of the commitment manipulation remained significant even after accounting for these mediators. In short, these results demonstrate clearly that perceptions of value commitment, and not perceptions of effectiveness, best explain the negative effect of the commitment virtues in this context.

For the pro-gun-control target (i.e., the "good" agent), value commitment again emerged as a significant mediator of the effect commitment virtues had on moral goodness, but this time it had a *positive* indirect effect on moral goodness (95% CI [2.22, 10.40]; see Figure 3, bottom panel). As with the bad agent, effectiveness also had a significant positive indirect effect as well (95% CI [.27, 6.41]). In this case, the direct effect of virtue on moral goodness was reduced to nonsignificance with both mediators in the analysis (see Figure 3).

In sum, only appraisals of value commitment mediated the conditional effect of commitment virtues on moral goodness ratings. Consistent with the overall ef-

fect of the commitment virtues, value commitment was a positive mediator when John supported the good cause (i.e., was pro-gun-control), and a negative mediator when John supported a bad cause (i.e., was anti-gun-control). By contrast, appraisals of John's effectiveness were positively associated with moral goodness ratings regardless of which cause he supported.

GENERAL DISCUSSION

SUMMARY OF FINDINGS

Inspired by philosophical observations made by Kupperman (1991), Slote (1983), Gert (2004), and others, we tested the hypothesis that people discriminate between what we referred to as "core goodness" virtues, and "commitment" virtues. Core goodness virtues, such as honesty, respectfulness, and kindness, are perceived to be intrinsically valuable in social-moral enterprises, and therefore are perceived as unconditionally good. In contrast, commitment virtues, such as determination, diligence, and dedication, are perceived to be only conditionally good, that is, as improving moral goodness when applied to neutral or good agents, but as amplifying badness (or immorality) when applied to bad agents or engaged in the service of bad ends.

Study 1 provided initial support for this hypothesis by showing that participants perceived core goodness traits as enhancing the moral goodness of a bad agent (a terrorist), whereas they perceived commitment traits as enhancing the moral *badness* of a bad agent. Study 1 also ruled out valence as the critical factor underlying the differential effects these sets of traits had on the terrorist's perceived moral goodness or badness. Study 2 expanded upon these findings by including good and neutral agents, and by directly contrasting the effects of commitment traits with the effects of intelligence. If commitment traits amplified the moral badness of bad agents in Study 1 simply because they enhanced perceptions of the agent's agency or effectiveness, then a comparable effect should also have occurred for intelligence, since intelligence is also highly relevant to effectiveness. If, on the other hand, commitment traits enhanced moral badness because they signal information about the agent's corrupt character (i.e., the agent's endorsement of evil values), then the amplifying effect of commitment traits should be larger than that for intelligence. In Study 2, as predicted, core goodness traits were found to enhance the goodness of all agents, good, bad, or neutral, thus fully demonstrating their unconditionality. In contrast, and also as predicted, commitment traits had a conditional effect—they enhanced the goodness of good and neutral agents, but they amplified the *badness* of bad agents. Furthermore, commitment traits were found to have a greater amplifying effect on the evaluation of bad and good (though not neutral) agents than intelligence. The lack of difference between commitment traits and intelligence for neutral agents likely owes to the fact that this agent's values were unspecified.

Finally, in Studies 3–4, using a non-comparative between-subject design, which is less susceptible to demand effects, we replicated the conditional effect of commitment virtues, while also uncovering more direct evidence for the mechanism by which commitment virtues polarize the prevailing morality of the agent in question. Mediation evidence showed that commitment traits communicate that the agent strongly endorses moral or immoral values, and that this specific feature, rather than what such commitment traits communicate about an agent's effectiveness, best explains their influence on moral evaluations.

In sum, the main contribution of the present findings is to reveal a novel distinction between different character traits or virtues that guides ordinary individuals' assessments of moral character. When a person is known to support a moral or immoral cause, and is perceived to possess qualities that indicate value commitment (e.g., dedication, determination), this person is perceived to have deeper convictions than an otherwise equivalent person without these qualities. As a result, the committed person is praised or condemned even further. Thus, our findings show that inferences about moral values, and a person's commitment to these values, play a prominent role in person evaluations. Critically, these findings cannot be attributed to baseline differences in valence between the virtues—all the virtues we examined were positively valenced by default, and yet only some of these virtues enhanced the moral *badness* of an agent. Nor can these findings be explained simply by the usefulness of commitment traits for signaling agency—since highly agentic, non-moral traits (e.g., intelligence) did not exert the same effects as did the commitment traits. In other words, commitment virtues appear to be a special sub-class of agentic traits (or “strength of will” traits) that have important implications for moral evaluation.

RELATION TO EXISTING RESEARCH

Although both commitment traits and intelligence are perceived to enhance the overall effectiveness of an agent, only commitment traits were perceived consistently to enhance the overall immorality of bad agents (e.g., intelligence only had a polarizing effect in the ranking procedure of Study 2, but not in the rating procedure of Study 3, whereas commitment traits amplified immorality in both studies and to a larger degree than did intelligence). A related prior study by Wojciszke, Bazinska, et al. (1998, Study 4) showed that intelligence not only enhanced the overall impression of a morally good person, but just as with commitment traits in the present research (in Study 2), it also (slightly) amplified the negative overall impression of morally bad individuals. While this finding is similar to the present findings, it demonstrates only the modifying role of intelligence, and not that of value commitment traits. The present results go beyond these prior findings by: (a) demonstrating the conditional effect of commitment virtues within moral evaluations, (b) clearly differentiating between the effects of intelligence and value commitment traits, and (c) revealing different mechanisms for their effects. Com-

mitment traits polarize evaluations because of what they convey about a person's commitment to specific values, and not because of what they convey about a person's overall effectiveness (or competence). This mechanism is therefore quite distinct from the mechanism observed in Wojciszke et al.'s prior study, which appears to depend on inferences of effectiveness (from a person's intelligence). Perhaps as a result of this difference in mechanism, the amplifying effects of commitment traits in this study were found to be larger and more consistent than the effect of intelligence.¹²

As a consequence, the effect of commitment traits is not reducible to what they signal about agency, which primarily pertains to agents' effectiveness in carrying out their goals (see Abele et al., 2008; Abele & Wojciszke, 2007; Fiske, Cuddy, & Glick, 2007). A different source of evidence further corroborates this point. In a prior study (reported in Goodwin et al., 2014), we found that while many commitment virtues were rated high on usefulness for judging agency, that is, "how effectively a person can attain his or her goals and exert influence over events in his or her life" (e.g., *dedicated*, $M = 7.92$; *determined*, $M = 7.97$), traits related to intelligence and competence (e.g., *intelligent*, *competent*) were rated almost equally high on agency ($M_s = 7.72, 7.43$, respectively). However, only the commitment virtues were rated high on their usefulness for judging moral *character* ($M = 7.54, M = 7.10$) and strength of will ($M = 7.24, M = 8.17$), whereas, the competence traits were rated appreciably lower on these indices: character ($M_s = 6.31, 6.37$) and strength of will ($M_s = 5.39, 5.66$). It would seem then that the key difference between commitment virtues and traits that relate to intelligence comes from their relation to moral character, and not from their relation to agency. Thus, while commitment virtues are undoubtedly agentic, the set of commitment virtues is narrower than the set of agentic traits, since commitment virtues refer only to those agentic traits that are also deemed relevant for moral character whereas the set of agentic traits includes additional non-moral traits related to intelligence, competence, ingenuity, power, and dominance.

As we alluded to earlier, prior research has shown that character traits differ in their prototypicality for defining a person with "moral character" (i.e., how central or accessible a trait is in a person's memory; see e.g., Lapsley & Lasky, 2001; Walker & Hennig, 2004; Walker & Pitts, 1998). In this respect, it is worth noting that many of the peripheral character traits uncovered by Lapsley and Lasky (2001) overlap (though not completely) with the commitment traits identified in the present studies. This points to the possibility that the conditionality of certain virtues contributes in some way to their non-prototypicality as character traits, which would be an interesting task for future research to explore.

Finally, it is also important to note that the conceptual distinction we have identified does not simply reduce to the distinction made by Immanuel Kant between perfect and imperfect duties (see Trafimow & Trafimow, 1999). Within

12. Also, unlike Wojciszke et al., we did not examine only global evaluations, but focused more directly on evaluations of agents' goodness or their moral character.

Kant's (1797/1991) formulation, perfect duties arise from normative rules, and are always obligatory—they include, for example, keeping one's promises, and telling the truth. Imperfect duties, by contrast, are supererogative, that is, while praiseworthy, actors have discretion as to when to perform them—for example, being generous or charitable. Honesty and trustworthiness—virtues that support the performance of perfect duties—are clearly core goodness traits, as our studies show. However, as we have also shown, many traits that Kant would have classified as relevant to the enactment of imperfect duties (or as "supererogative" traits) also appear to be core goodness traits, in that they operate unconditionally, for example, traits related to charity, kindness, respectfulness, and humility (see Studies 1–3). In other words, even though supererogative traits, such as kindness, are likely to be perceived as partially discretionary (i.e., as relevant to the performance of imperfect duties), most of our participants perceived these traits to be unconditional virtues—that is, as improving the moral worthiness of all agents. Thus, the distinction between conditional and unconditional virtues does not appear to be reducible to a Kantian framework.¹³

An interesting task for future research would be empirically to investigate the relation between the unconditionality of various virtues and their perceived obligatoriness (i.e., whether they are perceived as perfect or imperfect duties).

LIMITATIONS AND UNEXPECTED FINDINGS

One potential limitation of the current studies is that, although we utilized gender-neutral agents (Studies 1–3), or male targets only (Study 4), we cannot rule out the possibility that agents' gender may affect how traits modify evaluations of them. Research by Rudman and her colleagues on gender-stereotype violations suggests that women are often penalized for possessing highly agentic traits (e.g., *dominating*), while men are penalized for lacking these traits (Moss-Racusin, Phelan, & Rudman, 2010; Rudman & Glick, 2001). While it remains to be seen whether commitment traits would influence the morality of male and female agents differently, we have our doubts. For one thing, as we have discussed above, commitment virtues cannot be reduced to agency, and many of the agentic traits that are seen as negative for women (e.g., *self-centered*, *arrogant*) fall outside the domain of commitment traits. Second, because of their direct relevance for moral character and value commitment, most, if not all, of the commitment traits tested here have broad social appeal, and therefore may be equally desirable in men and women. Nevertheless, future research should consider the potential moderating effects of target gender more closely.

13. By extension, our findings are not reducible to Reeder and Brewer's (1979) distinction between partially restrictive and hierarchically restrictive traits, since Trafimow and Trafimow (1999) have shown that Kant's notions of perfect and imperfect duties map onto Reeder and Brewer's trait dimensions, respectively.

One unexpected result from Study 3 was that the differential effects of commitment traits and intelligence were larger for the bad agent than for the good agent. Indeed, for the good agent, we were unable to conduct a mediation analysis, owing to the fact that the modifying effect of the commitment trait on overall moral evaluations was nonsignificant (though directionally consistent with the effects for the bad agent). And, although we obtained a significant effect of commitment virtues for both the good and bad agent in Study 4 (for participants who endorsed gun control as a good cause), the size of the effect was still smaller for the good agent. This smaller effect on the morality of the good agent may have to do with the fact that commitment traits (e.g., *dedication*) are more likely to be inferred by default for good than for bad agents (or when used in the service of good vs. bad ends). As a consequence, adding such default-positive commitment traits to bad agents (for which such traits are not assumed) would have a greater impact on evaluations of those agents than would adding such traits to good agents (for which such traits are already assumed). Future research would be needed to establish this with greater certainty.

CONCLUSION

The benefits of commitment traits as one part of a person's broader character are obvious. A determined and dedicated person can be relied upon to perform relevant duties to a greater extent than a lazy or uncommitted person. But, as we have seen, people are more likely to condemn individuals who possess virtues such as dedication when such virtues are used to achieve morally bad ends. By contrast, virtues such as honesty, justice, respectfulness, humility, and kindness are favored unconditionally, regardless of the agent in question.

Core goodness virtues such as kindness and honesty may be problematic when practiced in the extreme (an overly kind person may be taken advantage of, and an overly honest person may hurt the people she loves; see Aristotle, trans. 2011). Moreover, as we discussed in the introduction, sometimes core goodness virtues come into conflict (e.g., fairness with loyalty, and honesty with kindness). There may therefore be circumstances in which it is morally preferable not to behave in accordance with a particular core goodness trait. Thus, we do not want to claim that discrete *acts* that stem from such core goodness traits will be seen as good unconditionally, across all possible instances. Rather, our point about the unconditionality of core goodness traits is intended to apply at a broader, dispositional level, regarding the possession of such *traits*. And at this level, our research has uncovered a distinction between moral virtues that divides along this dimension of conditional and unconditional goodness.

Our findings do not imply that the commitment virtues are not genuine moral virtues. Indeed, in our prior work, which we relied upon here, commitment traits were rated as highly relevant to moral character (Goodwin et al., 2014). However,

our results do suggest that the way such commitment virtues contribute to having good moral character is more complex than the way core goodness virtues contribute to having good character, thereby revealing an important distinction underlying person-centered moral evaluations.

APPENDIX A.

PRELIMINARY ANALYSES FROM STUDY 2

Our initial analyses tested how consistent the data were across different trait sets, and different exemplars, for the purposes of aggregation. We first tested whether the two sets of core goodness and commitment traits produced roughly the same ranking patterns for all five agents. To do this, for each of the five agents, we examined the extent to which the two different trait sets produced concordant rankings. We first computed, for each participant, an index of the extent to which their observed rankings matched the predicted pattern for each of the ten different trials. For the bad agents, each participant's ranking scores (1 "best person" to 4 "worst person") were multiplied by values of 3 (value commitment), 1 (intelligent), -1 (no trait), or -3 (core goodness), and then summed. Larger resulting numbers thus reflect a more clear-cut corroboration of the predicted trend. Similarly, for the good agent, the ranking scores were multiplied by 3 (no trait), 1 (intelligent), -1 (value commitment), or -3 (core goodness), and, for the neutral agent by 4 (no trait), -1 (intelligent), -1 (value commitment), or -2 (core goodness), then summed, which again means that larger numbers on this index better reflect the predicted trend. The weights for the neutral agent, though unevenly spaced, were used to avoid applying zero to the middle values (i.e., [3, 0, 0, -3]), and thus eliminating the influence of these middle values in the procedure.

Then, for each of the five different agents, we correlated the resulting scores across the two different trait sets. The ranking patterns were strongly correlated for each agent, reflecting a similar pattern of results for the different trait sets for each of the five agents: Nazi, $\rho(162) = .54, p < .001$, terrorist, $\rho(162) = .63, p < .001$, human rights activist, $\rho(162) = .60, p < .001$, volunteer aid worker, $\rho(162) = .50, p < .001$, and person, $\rho(162) = .70, p < .001$. Because of these concordant results, we aggregated across the two different trait sets in the main analyses.

We next examined the consistency of the data across agents. To test whether the two different good agents (human rights activist, aid worker), and separately, the two different bad agents (Nazi, terrorist) produced the same general pattern of rankings, we correlated the indices of the extent to which participants' rankings matched the predicted trend (as described above) within each agent category (good, bad). Ranking scores for the bad-agent exemplars (Nazi and terrorist) were highly correlated, $\rho(162) = .61, p < .001$, as were the ranking scores for the good-agent exemplars (activist and aid worker), $\rho(162) = .59, p < .001$. Given these high concordance ratings, we aggregated ranking scores for the good-agent exemplars and the bad-agent exemplars as well.

APPENDIX B.

VIGNETTES USED IN STUDY 4

Anti-Gun-Control Vignette

All Participants Assigned to the Anti-Gun-Control Version First Read:

John is a gun enthusiast. He loves the thrill of hunting, and handling firearms. He has a large collection of handguns and rifles in his basement. Because of his desire to own guns, John is anti-gun control. He is opposed to any gun-control measures that would restrict his freedom to purchase or sell firearms. For example, he was opposed to the recent proposal in the Senate aimed at expanding federal background checks for firearms purchases. John aligns his values very much with those of the Gun Owners of America (an extremist anti-gun-control organization).

Participants Then Read One of the Following Depending on Which Virtues Condition They Were Assigned to:

Conditional Virtues. John is very determined and committed to serving the cause of gun rights. When he retired from his job last year, he immediately volunteered as a behind-the-scenes fundraiser for the Gun Owners of America to help fight against gun restrictions. He now spends a great deal of his time working tirelessly to raise funds for this extremist organization.

No Virtues. John retired from his job last year, and he now spends a great deal of time fishing.

Pro-Gun-Control Vignette

All Participants Assigned to the Pro-Gun-Control Version First Read:

John is a gun owner. He keeps a single handgun locked securely in his bedroom for his family's protection. However, John is also pro-gun control. He is in favor of more gun-control measures that would restrict the purchasing or selling of firearms. For example, he supported the recent proposal in the Senate aimed at expanding federal background checks for firearms purchases. John aligns his values very much with those of the National Gun Victims Action Council (an organization which advocates for increased gun control in the U.S.).

Participants Then Read One of the Following Depending on Which Virtues Condition They Were Assigned to:

Conditional Virtues. John is very determined and committed to serving the cause of gun control. When he retired from his job last year, he immediately volunteered as a behind-

the-scenes fundraiser for the National Gun Victims Action Council to help support gun control. He now spends a great deal of his time working tirelessly to raise funds for this organization.

No Virtues. John retired from his job last year, and he now spends a great deal of time fishing.

REFERENCES

- Abele, A. E., Uchrowski, M., Suitner, C., & Wojciszke, B. (2008). Towards an operationalization of the fundamental dimensions of agency and communion: Trait content ratings in five countries considering valence and frequency of word occurrence. *European Journal of Social Psychology, 38*, 1202-1217.
- Abele, A. E., & Wojciszke, B. (2007). Agency and communion from the perspective of self versus others. *Journal of Personality and Social Psychology, 93*, 751-763.
- Aristotle. (trans. 2011). *Nicomachean ethics*. (R. C. Bartlett & S. D. Collins, Trans.). Chicago: The University of Chicago Press.
- Brambilla, M., Rusconi, P., Sacchi, S., & Cherubini, P. (2011). Looking for honesty: The primary role of morality (vs. sociability and competence) in information gathering. *European Journal of Social Psychology, 41*, 135-143.
- Brambilla, M., Sacchi, S., Rusconi, P., Cherubini, P., & Yzerbyt, V. Y. (2012). You want to give a good impression? Be honest! Moral traits dominate group impression formation. *British Journal of Social Psychology, 51*, 149-166.
- Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6*, 3-5.
- Cottrell, C. A., Neuberg, S., & Li, N. P. (2007). What do people desire in others? A sociofunctional perspective on the importance of different valued characteristics. *Journal of Personality and Social Psychology, 92*, 208-231.
- Doris, J. M. (2002). *Lack of character: Personality and moral behavior*. New York: Cambridge University Press.
- Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Science, 11*, 77-83.
- Gert, B. (2004). *Common morality: Deciding what to do*. New York: Oxford University Press.
- Goodwin, G., Piazza, J., & Rozin, P. (2014). Moral character information predominates in person perception and evaluation. *Journal of Personality and Social Psychology, 106*, 148-168.
- Harman, G. (1999). Moral philosophy meets social psychology: Virtue ethics and the fundamental attribution error. *Proceedings of the Aristotelian Society, 99*, 315-332.
- Kant, I. (1991). *The metaphysics of morals* (M. Gregor, Trans.). Cambridge, UK: Cambridge University Press. (Original work published 1797)
- Kupperman, J. (1991). *Character*. New York: Oxford University Press.
- Lapsley, D. K., & Lasky, B. (2001). Prototypic moral character. *Identity, 1*, 345-363.
- Leach, C. W., Ellemers, N., & Barreto, M. (2007). Group virtue: The importance of morality (vs. competence and sociability) in the positive evaluation of in-groups. *Journal of Personality and Social Psychology, 93*, 234-249.

- Linley, P. A., Maltby, J., Wood, A. M., Joseph, S., Harrington, S., Peterson, C., Park, N., & Seligman, M. E. P. (2007). Character strengths in the United Kingdom: The VIA Inventory of Strengths (VIA-IS). *Personality and Individual Differences, 43*, 341-351.
- Miller, C. (2003). Social psychology and virtue ethics. *The Journal of Ethics, 7*, 365-392.
- Moss-Racusin, C. A., Phelan, J. E., & Rudman, L. A. (2010). When men break the gender rules: Status incongruity and backlash against modest men. *Psychology of Men and Masculinity, 11*, 140-151.
- Newport, F. (2013, April 29). *Americans wanted gun background checks to pass Senate* [Web log post]. Retrieved from <http://www.gallup.com/poll/162083/americans-wanted-gun-background-checks-pass-senate.aspx>.
- Park, N., Peterson, C., & Seligman, M. E. P. (2004). Strengths of character and wellbeing. *Journal of Social and Clinical Psychology, 23*, 603-619.
- Peterson, C., & Seligman, M. E. P. (2004). *Character strengths and virtues: A handbook and classification*. Washington, DC: APA Press and Oxford University Press.
- Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40*, 879-891.
- Reeder, G. D. (1993). Trait-behavior relations and dispositional inference. *Personality and Social Psychology Bulletin, 19*, 586-593. doi: 10.1177/0146167293195010
- Reeder, G. D., & Brewer, M. B. (1979). A schematic model of dispositional attribution in interpersonal perception. *Psychological Review, 86*, 61-79.
- Rokeach, M., & Rothman, G. (1965). The principle of belief congruence and the congruity principle as models of cognitive interaction. *Psychological Review, 72*, 128-142.
- Royzman, E. B. (2000). Are experiments possible? The limitations of a posteriori control in experimental behavior analysis. *Theory and Psychology, 10*, 171-196.
- Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review, 5*, 296-320.
- Rudman, L. A., & Glick, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. *Journal of Social Issues, 57*, 732-762.
- Slote, M. A. (1983). *Goods and virtues*. Oxford: Clarendon.
- Smith, K. D., Smith, S. T., & Christopher, J. C. (2007). What defines the good person? Cross-cultural comparisons of experts' models with lay prototypes. *Journal of Cross-Cultural Psychology, 38*, 333-360.
- Strohmingner, N., & Nichols, S. (2014). The essential moral self. *Cognition, 131*, 159-171.
- Toner, E., Haslam, N., Robinson, J., & Williams, P. (2012). Character strengths and wellbeing in adolescence: Structure and correlates of the Values in Action Inventory of Strengths (VIA-IS) for children. *Personality and Individual Differences, 52*, 637-642.
- Trafimow, D., & Trafimow, S. (1999). Mapping perfect and imperfect duties onto hierarchically and partially restrictive trait dimensions. *Personality and Social Psychology Bulletin, 25*, 687-697.
- van 't Wout, M., & Sanfey, A. G. (2008). Friend or foe: The effect of implicit trustworthiness judgments in social decision-making. *Cognition, 108*, 796-803.
- Vranas, P. B. M. (2005). The indeterminacy paradox: Character evaluations and human psychology. *Nous, 39*, 1-42.
- Walker, L. J., & Frimer, J. A. (2007). Moral personality of brave and caring exemplars. *Journal of Personality and Social Psychology, 93*, 845-860.
- Walker, L. J., & Hennig, K. H. (2004). Differing conceptions of moral exemplarity: Just, brave, and caring. *Journal of Personality and Social Psychology, 86*, 629-647.
- Walker, L. J., & Pitts, R. C. (1998). Naturalistic conceptions of moral maturity. *Developmental Psychology, 34*, 403-419.
- Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science, 17*, 592-598.

- Wojciszke, B. (2005). Morality and competence in person and self-perception. *European Review of Social Psychology, 16*, 155-188.
- Wojciszke, B., & Abele, A. E. (2008). The primacy of communion over agency and its reversals in evaluations. *European Journal of Social Psychology, 38*, 1139-1147.
- Wojciszke, B., Bazinska, R., & Jaworski, M. (1998). On the dominance of moral categories in impression formation. *Personality and Social Psychology Bulletin, 24*, 1245-1257.
- Wojciszke, B., Dowhyluk, M., & Jaworski, M. (1998). Moral and competence-related traits: How do they differ? *Polish Psychological Bulletin, 29*, 283-294.