Self-serving karmic beliefs: prosociality influences vicarious optimism

Shuying Zheng[1,#], Xinyuan Yan[1,#], Jenifer Z Siegel[2], Vladimir Chituc[2], Shiyi Li[1], Molly J Crockett[2^], Yina Ma[1*^]

[1]State Key Laboratory of Cognitive Neuroscience and Learning, IDG/McGovern Institute for Brain Research, Beijing Key Laboratory of Brain Imaging and Connectomics, Beijing Normal University, Beijing 100875, China

[2]Department of Psychology, Yale University, New Haven, Connecticut, USA

[#] contributed equally to this work.

[^] shared last author

*Correspondence should be addressed to:

Yina Ma, Ph.D.

State Key Laboratory of Cognitive Neuroscience and Learning

Beijing Normal University

19 Xin Jie Kou Wai Da Jie, Beijing, 100875, China

Phone/Fax: 8610-5880-2846

Email: yma@bnu.edu.cn

Abstract

Belief in karma is ubiquitous, appearing early in development and impacting prosocial behavior. Here, we tested the possibility that karmic beliefs are self-serving: are "good" people more likely to believe that good things happen to good people? Study 1 (n=170) showed stronger karmic beliefs in more prosocial individuals. Next, we tested whether self-serving karmic beliefs arose from a motivated deployment of vicarious optimism: prosocial individuals adopt karmic beliefs by prioritizing desirable (the fortunes of good people, the misfortunes of bad people) over undesirable information when predicting the future. Study 2 (n=107) showed that prosocials were more optimistic about the future of morally good (than bad) agents, while individualists were not. This was driven by prosocials' failure to update beliefs from undesirable information about morally good agents. Together, we suggest that karmic beliefs are self-serving, and result from a failure to update beliefs from information that conflicts with a karmic worldview.

Key words: karmic belief, belief updating, social value orientation, prosociality, moral inference

Introduction

"What goes around, comes around." This simple proverb highlights the basic principle behind karma, or the belief that the sum of current and past moral behavior determines future consequences (Krishan, 1997). Although the concept of karma originates in Eastern religious philosophy, a karmic worldview is nevertheless universal and ubiquitous, appearing across cultures (White, 2017) and emerging at a young age, prior to extensive cultural learning (Banerjee & Bloom, 2017).

Research on just-world beliefs has explored this tendency to believe that fundamentally, good things happen to good people and bad things happen to bad people (Lerner 1980), and these beliefs have broad impacts. They are linked to lower levels of negative affect (Lipkusa, Dalbert, & Siegler, 1996), decreased group discontent (Hafer & Olson, 1993), and increased beliefs that personal deprivation was fair (Hafer & Olson, 1989). Furthermore, several studies suggest that there is a direct relationship between belief in a just world and moral behavior. Work in cross-cultural psychology, history, and ethnography suggests that religious beliefs in supernatural punishment and reward, including karmic beliefs, are critical for the development of cooperation in large-scale societies (Norenzayan et al., 2016).

Few studies have explored the cognitive underpinnings of how these karmic beliefs are formed, and most that do rely on self-report scales to measure individuals' beliefs (Dalber, 1999; Kopella, Lehmann, &Farley, 2010). Research on the cognitive mechanisms of belief updating may hold a clue. Recent work on the optimism bias shows that people are more likely to update their beliefs from desirable than undesirable information (Ma et al., 2016; Sharot, Korn, & Dolan, 2011). This optimism bias is evident not only for beliefs about the self, but also extends to others: people show *vicarious optimism* for those they care about (Kappes, Faber, Kahane, Savulescu, & Crockett, 2018). That is, when considering the future of sympathetic (but not unsympathetic) others, people are more likely to update their beliefs from good news than bad news. This hints at a possible mechanism that underpins the development of karmic beliefs: individuals may come to believe that good things will happen to good people by failing to update their beliefs from undesirable information (e.g., the misfortunes of good people, or the fortunes of bad people).

Conceptualizing optimistic belief updating as a *motivated* process introduces the possibility that karmic beliefs may also be motivated: people may be more likely to believe that good things will happen to good people if they themselves are good. Self-serving beliefs have been documented in a variety of contexts (Jenkins, Macrae, & Mitchell, 2008; Taylor & Doria, 1981); here, we investigate whether they can be observed in the domain of karmic beliefs and vicarious optimism. We predicted that individual differences in prosociality would be positively associated with karmic beliefs and vicarious optimism. We tested these predictions in two studies. In our first study, we measured prosociality with social value orientation (SVO; Murphy, Ackermann, & Handgraaf, 2011) via a series of monetary allocations between oneself and an anonymous other person. This allowed us to measure the degree of prosociality of each participant, as well as to categorize each participant as either a "prosocial" or "individualist" type, depending on how selfishly or generously they weigh the wellbeing of a stranger compared to their own. Insofar as someone is a "prosocial" type, they should be motivated to believe in karma, since karma would indirectly benefit them. If someone is an "individualist" type, however, they will be motivated *not* to believe in karma, since karma would indirectly harm them.

In our second study, we investigated the cognitive mechanisms underpinning self-serving karmic beliefs using a task that measures vicarious optimism (Kappes et al., 2018). This task enabled us to quantify beliefs and how they update to both positive and negative information. Participants learned about a morally "good" agent and a morally "bad" agent. Past work shows that people form more vicariously optimistic beliefs about good agents than bad ones (Kappes et al., 2018). Here, we tested whether this effect was moderated by individual differences in prosociality by comparing the extent to which individualist vs. prosocial participants were more vicariously optimistic about good (relative to bad) agents. Because we predicted that belief in karma is motivated, we did not expect prosocial participants to be more vicariously optimistic in general. Rather, we anticipated that their vicarious optimism would extend only to morally good others. Both study protocols were in line with the standards of the Declaration of Helsinki and approved by the research ethics committee at the State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University (Beijing, China).

Study 1.

The goal of Study 1 was to investigate whether individual differences in prosociality are related to karmic beliefs. We expected stronger karmic beliefs in more prosocial individuals.

Method

Participants. We estimated appropriate sample size using G*Power (Faul, Erdfelder, Buchner, & Lang, 2009). To detect a small-to-medium correlation of 0.25 between prosociality (SVO angle, see below) and karmic beliefs (*Belief in a Just World* and *Belief in Karma*, see below) with $\alpha = 0.05$ and 85% power, a sample of 140 participants was needed. A total of 174 participants completed an online survey in exchange for payment. We excluded participants who failed to answer post check questions ($n = 27$). This led to a final sample size of 147 (102 females; age range: 18-50 years, mean age ± SE = 24.57 ± 0.43 years). All participants reported no history of neurological or psychiatric diagnoses. Participants provided informed consent in accordance with the Beijing Normal University Institutional Review Board.

Procedure. We measured participants' social value orientation (SVO (Murphy et al., 2011), belief in a just world (Dalbert, 1999) and belief in karma (Kopalle et al., 2010) with a set of questionnaires.

Social Value Orientation (SVO). Participants completed the 15-item SVO Slider task (Murphy et al., 2011), which measures stable preferences about resource distributions between oneself and others. Each participant made monetary allocation decisions between him/herself and a mutually anonymous partner. For each item, participants were presented with 9 self-partner monetary allocation options over a well-defined continuum of joint payoffs, and were asked to choose the allocation they most preferred. Participants were paid based on their monetary allocation on a randomly chosen item. Based on the inverse tangent of the ratio between mean allocations for the self and the paired partner, the 6 primary items yielded a measure of SVO angle (SVO°) that captures an individual's prosociality and categorized participants into altruist, prosocial, individualist, and competitor. Similar to previous studies (Liu et al., 2019), we referred to both "altruist" and "prosocial" participants as "prosocials" (i.e., SVO° > 22.45°), and both "individualist" and "competitor" as "individualists" (i.e., SVO° < 22.45°). In the current study, 103 participants (72 females;

mean age ± SE = 23.90 ± 0.38 years) were categorized as prosocials and 44 participants (30 females; mean age ± SE = 26.14 ± 1.12 years) were categorized as individualists.

Belief in Karma (BK). Belief in Karma was evaluated with 7 items BK scale (Kopalle et al., 2010); e.g. "Good actions in the present lead to good outcomes in the future either in this life or in the hereafter.") rated on a 7-point scale (1 = completely disagree, 7 = completely agree. Cronbach's $\alpha = 0.73$).

Belief in a Just World (BJW). We assessed participants' belief in a just world with the BJW Scale (Dalbert, 1999), which consisted of 13 items (e.g., "I believe that I usually get what I deserve") rated on a 6-point scale (1 = completely disagree, 6 = completely agree. Cronbach's $\alpha = 0.91$).

Results

To test the relationship between prosociality and karmic beliefs, we examined the correlations between SVO angle and karmic beliefs. In line with our predictions, SVO angle significantly predicted higher scores on belief in karma ($r(147) = 0.18$, $p = 0.031$), and predicted higher scores on belief in a just world ($r(147) = 0.17$, $p = 0.043$). As a robustness check, we directly compared prosocials and individualists on belief in karma and belief in a just world. We found that prosocials scored higher on belief in karma ($t(145) = 2.43$, $p = 0.016$, mean difference ± SE (prosocials vs. individualists) = 2.62 ± 1.08, 95%CI = [0.49, 4.75], Cohen d' = 0.20) and higher on the belief in a just world ($t(145) = 2.05$, $p = 0.042$, mean difference ± SE (prosocials vs. individualists) = 3.53 ± 1.72, 95%CI = [0.12, 6.93], Cohen d' = 0.17) than individualists. These results lend initial support for the link between prosociality and karmic beliefs, with stronger karmic beliefs for more prosocial individuals.

Study 2

Study 1 suggested that karmic beliefs are self-serving: people are more likely to believe that good things will happen to good people if they themselves are good. In Study 2, we investigated a potential cognitive mechanism that could explain such a relationship between prosociality and karmic beliefs. Here, we predicted that asymmetries in vicarious optimism between strangers with better versus worse moral characters would be stronger for more prosocial individuals. In other words, to the extent that someone has a better moral character,

they should be more optimistic about the future of morally better people, and this optimism should manifest as a stronger integration of good news (relative to bad) into beliefs about the future.

Method

Participants

Sample estimation. Previous research on the prosocial behaviors of different SVO types (Karagonlar & Kuhlman, 2013) reported an average effect size of SVO type (prosocials vs. individualists) of Cohen f = 0.241. Based on such an effect size, the G*Power (Faul et al., 2009) calculation suggested that a sample of 100 participants (50 prosocials and 50 individualists) was needed to reveal a reliable effect of between-factor SVO type, with α = 0.05 and 85% power in repeated measures ANOVA with 4 repeated measurements for 2 (Agent)-by-2 (Feedback). Previous studies have shown that only about 30% to 35% individuals would be classified as individualists (in our Study 1, as well as dataset of Japan and Netherlands in previous studies, Haruno & Frith, 2010; Van Lange, 1999). Therefore, we would need a sample of 166 participants to recruit at least 50 individualists (assuming a rate of 30% individualists) and 50 prosocials.

Participants in pre-screening session. One hundred and sixty-six participants (88 females; age range: 17 -33 years, mean age ± SE = 22.07 ± 0.21 years) were invited to a pre-screening session where they completed the same SVO slider task as in Study 1. In the current sample, we identified 57 individualists (34.33%) and 109 prosocials (65.67%). We invited all the individualists and the same number of prosocials to take part in study 2.

Participants in Study 2. Among the prescreened participants, 107 participants (41 females; age range: 17-33 years, mean age ± SE = 21.76 ± 0.27 years) participated in Study 2, with 54 prosocials (20 females; mean age ± SE = 21.17 ± 0.39 years) and 53 individualists (21 females; mean age ± SE = 22.35 ± 0.34 years).

All participants were right-handed, reported no history of neurological or psychiatric diagnoses, and had normal or corrected-to-normal vision. Participants provided informed

written consent after the experimental procedure had been fully explained, and were instructed of their right to withdraw at any time during the study.

Procedure

General procedure. Upon arrival, participants provided written informed consent and demographic information. Then, participants first completed a hypothetical version of a moral decision task where they had to trade off profit for themselves against pain for another person (Crockett, Kurth-Nelson, Siegel, Dayan, & Dolan, 2014), followed by a moral inference task in which they predicted the moral decisions of a "good" and a "bad" agent, presented in randomized order (Siegel, Mathys, Rutledge, & Crockett, 2018). Next, participants completed a vicarious optimism task (Kappes et al., 2018) in which they indicated and updated their beliefs about the likelihood of adverse life events happening to three targets: themselves (referred to as Self), as well as the good and bad agents from the moral inference task. After the prediction for both agents, participants played a one-shot trust game (Delgado, Frank, & Phelps, 2005) with each agent (i.e., "decider 1" and "decider2") in a randomized order. Finally, participants completed the life orientation test (LOT, Scheier, Carver, & Bridges, 1994), a measure of explicit optimistic beliefs, about themselves, the good agent, and the bad agent. The explicit optimism measurement was always administered after the vicarious optimism task to avoid a potential influence of explicit reports of optimism on optimistic belief updating. We debriefed participants at the end of the study, and participants completed a set of questionnaires online a week later.

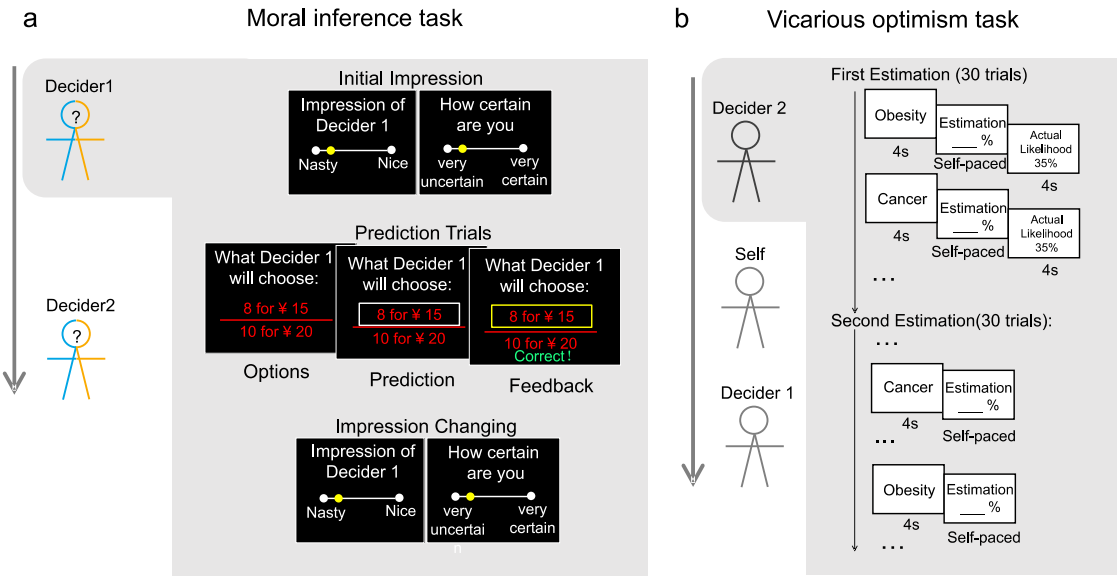

Figure 1. Moral character inference and vicarious optimism tasks. (a) In the moral inference task, participants predicted two agents' (labeled as "decider 1" and "decider 2") binary decisions to profit from inflicting shocks on an anonymous other. Participants predicted 50 trials for each agent and the order was counterbalanced between participants. For each agent, participants first rated an initial impression (from "nasty" to "nice") and the certainty of the impression (from "very uncertain" to "very certain") before prediction. During prediction, the selected option was highlighted by a white box. Then the white box disappeared and the correct option was highlighted by a yellow box, with a feedback of "Correct" or "Wrong" at the bottom. The prediction trials were self-paced. After every 3 trials and at the end of each block, participants rated their impression of the agent's moral character (on a scale from "nasty" to "nice") and the certainty of their impression. (b) In the two-stage vicarious optimism task, participants estimated the likelihood of negative events happening to oneself, to the good agent and to the bad agent (still labeled as "decider 1" and "decider 2"). For each target, participants completed two estimation sessions for one target before beginning with the next target, and the order of targets was counterbalanced across participants. In the first session, participants were presented with 30 different negative life events, and estimated the likelihood of these events (1st estimation, E1), each 1st estimation was followed by the probability of the event occurring to an average person in a similar environment (i.e., feedback). In the second session, participants estimated the likelihood of the same 30 events for the second time (in a random order) without receiving feedback (2nd estimation, E2).

Moral Inference task. In the moral inference task (Siegel et al., 2018), participants predicted a series of 50 choices for two "agents" undergoing the moral decision task they just completed (Figure 1a). Participants predicted all 50 choices for one agent before beginning with the next, and the order of agents was randomized across participants. At the beginning of each

block, before making any predictions, participants indicated their initial impression of the moral character of each agent on a scale from "nasty" to "nice" (agents were labeled as "decider1" and "decider2") and their uncertainty about the initial impression (from "very uncertain" to "very certain"). In each trial, participants were presented with the two options the agent faced and predicted which option the agent would choose. After making their prediction, participants received feedback indicating whether it was correct. After every three trials and at the end of each block, participants reported their current impression of the agent's moral character (from "nasty" to "nice") and how certain they were about their current impression (from "very uncertain" to "very certain") on continuous scales.

To manipulate moral character of the different agents, their choices were simulated with different preferences towards profiting from pain (see Supplementary Method). This was operationalized as their exchange rate between money for themselves and shocks for the other, describes as their *harm aversion*. Agents' harm aversion was parameterized as κ: when $\kappa = 0$, agents are minimally harm averse and will accept any number of shocks to the other to increase their profits; as κ approaches 1, agents become maximally harm averse and will pay increasing amounts of money to avoid a single shock to the other. The "good" agent was simulated with $\kappa = 0.7$, and the "bad" agent was simulated with $\kappa = 0.3$. Effectively, this meant that the bad agent was less averse to harming others and would therefore require less money to inflict pain than the good agent. On every trial, the agents faced the same two options, but because the agents had different preferences towards harm, they often chose differently.

Trust game: To motivate accurate predictions, participants were instructed to pay attention and learn about the agents' behaviors because they would later decide whether to trust the agents in a one-shot trust game that could earn them additional money. This game was conducted after participants predicted all the decisions of each agent. Participants were given 100 points which they could entrust with each agent. Any amount that they entrusted would be tripled and the agent could choose how much of the tripled amount to return to the participant. The trust game primarily served as a manipulation check, as we expected participants to entrust more points with the good agent, relative to the bad agent.

Vicarious optimism task. In the vicarious optimism task (Kappes et al., 2018), we asked participants to estimate the likelihood of adverse future events happening to three targets:

the self, the good agent and the bad agent from the moral inference task (Figure 1b). Within the vicarious optimism task, participants were presented with "decider1" or "decider2" just as in the moral inference task. The order of target was counterbalanced across participants, and all events within each condition were randomly presented.

In the first session, participants were presented with 30 different adverse life events and estimated the likelihood of each one happening to the target (1st estimation, referred to as E1). Participants were then presented with the probability of the event occurring to an average person in a similar environment (referred to as Feedback). In the second session, participants estimated the likelihood of the same 30 events happening to the target person again (in a random order) without receiving feedback (2nd estimation, E2). We systematically manipulated the feedback to control for the frequency and distributions of desirable and undesirable trials, though this was not made aware to participants. To make the feedback more authentic and in accordance with previous work (Sharot et al., 2011), we set Feedback between 3 to 77 ($3 < Feedback < 77$). To generate the desirable and undesirable feedback, we subtracted or added varying values from E1 ($Feedback_{desirable} = E1 - e$; $Feedback_{undesirable} = E1 + e$, $0 \leq e \leq 25$). By doing so, we balanced the proportion of desirable feedback (the probability of the negative life events was better than participants thought, i.e. estimation error = Feedback - E1 < 0) and undesirable feedback (the probability of the negative life events was worse than expected, i.e. estimation error = Feedback - E1 > 0) for each participant. This also diminished the difference in estimation errors, even if first estimations might vary across conditions.

Explicit optimism measure. We modified the life orientation test (LOT, Scheier et al., 1994) scale to assess participants' explicit optimism for the three targets (self, Good agent, and Bad agent). Participants first completed the LOT scale for the self, which was exactly the same as the original LOT scale. Then, participants were asked to complete the LOT scales for the two agents, in which we replaced the pronouns "I" "me" in each item with either "decider 1" or "decider 2". For example, the original item "If something can go wrong for me, it will." was modified as "If something can go wrong for decider1 (decider 2), it will". Decider 1 (2) referred to the first (second) decider participants predicted in the moral inference task. The order of LOT scales for the two agents was

counterbalanced across participants. A higher score in LOT scales indicated stronger optimism towards the target.

Questionnaire measurement. A week after the behavioral session, participants completed a questionnaire set online. We measured participants' empathy trait, emotion regulation, anxiety and psychological well-being with a set of questionnaires (Table S1) and found no difference between prosocial and individualistic participant groups in any of these scales (see Table S1).

Data analysis

Quality control. Following data exclusion criteria from our previous studies (Kappes et al., 2018), participants had to provide at least four valid updates after good news and four valid updates after bad news, which ensured the reliability of our measures. Three participants failed to reach these numbers and we therefore excluded them from further data analysis. Thus, our data analysis was conducted on a final sample of 53 prosocials (20 females; age range: 17–33 years, mean age ± SE = 21.17 ± 0.40 years) and 51 individualists (21 females; age range: 18–28 years, mean age ± SE = 22.33 ± 0.36 years).

Moral inference task

Impression ratings (from "nasty" to "nice") and certainty ratings (from "very uncertain" to "very certain") were transformed into values ranging from 0 to 1. We calculated summary statistics from these ratings and compared them between agents.

Computational modeling: We modelled participants' predictions for each agent separately using a hierarchical Bayesian learning model for learning hidden states (Mathys, Daunizeau, Friston, & Stephan, 2011). This produced a trial-wise sequence of belief estimates about the agent's moral character (μ) as well as a global estimate of belief volatility (ω) that captures inter-individual variability in the evolution rate of beliefs over time. As ω approaches ∞, beliefs become increasingly volatile and beliefs are updated more readily from new information. As ω approaches $-\infty$, beliefs become increasingly stable, so beliefs rely more heavily on prior beliefs (For full details about the model, Siegel et al., 2018). Our model fit participant's prediction well, explaining behavior with 81% accuracy, on average.

To investigate whether explicit ratings and model estimates differed between good and bad agents, or between prosocial and individualistic participants, we conducted mixed model 2 x 2 ANOVAs with Agent (good vs. bad) as a within-subject factor and SVO type (prosocial vs. individualist) as a between-subjects factor. We also employed linear regression models to estimate the effects of Agent, SVO type, and their interaction on explicit character and uncertainty ratings, controlling for time. Where main effects were qualified by a significant interaction, we conducted simple effects analyses.

Vicarious optimism task

Summary statistics about the first estimation (E1) were calculated and compared between target conditions (self *vs.* good agent *vs.* bad agent) and SVO type (prosocial vs. individualist).

For each event, we calculated a belief update as the difference between the participant's second and the first estimation: belief-update = 2nd Estimation – 1st Estimation. As the feedback was set between 3% and 77%, trials with E1 > 77% would always lead to desirable feedback and trials with E1 < 3% would always lead to undesirable feedback. Thus, we excluded these trials from our belief-update analyses (see Table S2 for results of all trials, including the excluded trials) to minimize feedback error difference between desirable and undesirable feedback conditions. We also considered participants' belief update upon desirable and undesirable feedback, respectively: desirable belief update = 1st Estimation – 2nd Estimation; undesirable belief update = 2nd Estimation – 1st Estimation. The difference between desirable and undesirable belief-update was used to test a bias toward optimism, i.e., optimistic biased belief updating: a greater difference indicated more optimism. To examine the effect of Target and SVO type on optimistic biased belief updating, we conducted an ANOVA with Target (self *vs.* good agent *vs.* bad agent) as a within-subject factor and SVO type (prosocial vs. individualist) as a between-subject factor. To confirm that any condition differences in belief updating did not reflect condition differences in initial estimates, we also conducted a similar analysis but controlled for differences in estimation errors in the analysis. Since the two SVO types differed in age (individualists vs. prosocials: $t(102) = 2.17$, $p = 0.032$; see *Participants*), we also controlled for age and gender in SVO type-related data analyses.

Results

Inferring moral character from moral decisions.

To ensure that participants were equally adept at predicting the good and bad agents' decisions by the end of the moral inference task and to examine whether the SVO type influenced the performance, we conducted a mixed model ANOVA with Agent as within-subject factor and SVO type as between-subject factor on the learning accuracy. Given that we aimed to reveal whether participants have learned the agents to a similar extent before making estimation about these agents' future, our analysis focused on the accuracy of second-half trials. This analysis showed a significant main effect of SVO type ($F(1,101)$[1] $= 5.64$, $p = 0.019$, $\eta_p^2 = 0.05$), but neither the main effect of Agent ($F(1,101) = 0.34$, $p = 0.559$, $\eta_p^2 < 0.01$) nor its interaction with SVO type ($F(1,101) = 0.53$, $p = 0.470$, $\eta_p^2 < 0.01$) was significant. This suggested that prosocials ($M \pm SE = 0.79 \pm 0.01$) were better at learning agents' moral characters, in general, than individualists were ($M \pm SE = 0.76 \pm 0.01$), and participants learned good ($M \pm SE = 0.78 \pm 0.01$) and bad ($M \pm SE = 0.77 \pm 0.01$) agents' moral character equally well (Figure 2a). This pattern remains after controlling for age and gender (main effect of SVO type: $F(1,99)=4.78$, $p=0.031$, $\eta_p^2 = 0.05$; main effect of Agent: $F(1,99)=0.40$, $p=0.529$, $\eta_p^2 < 0.01$; SVO type × Agent interaction: $F(1,99)=0.45$, $p=0.506$, $\eta_p^2 < 0.01$).

Stronger trust in good than bad agent.

To test how effectively participants differentiated between agents after learning and potential SVO type modulation, we conducted an Agent-by-SVO type ANOVA on the number of points participants entrusted with each agent in the trust game. There was a significant main effect of Agent ($F(1,101) = 47.39$, $p<0.001$, $\eta_p^2 = 0.32$), as predicted, participants entrusted significantly more points with the good agent ($M \pm SE = 56.23 \pm 2.84$, 95% CI = [50.60, 61.87]) than the bad agent ($M \pm SE = 30.78 \pm 2.82$, 95% CI = [25.18, 36.38]).This suggested that participants successfully differentiated the two agents and attributed better moral character to the good agent. We found no main effect of SVO type in the amount ($F(1,101) = 0.96$, $p = 0.330$, $\eta_p^2 <0.01$) nor an interaction between SVO type and Agent ($F(1,101) = 0.005$, $p = 0.942$, $\eta_p^2 < 0.01$). All results remained unchanged after controlling age and gender

[1] The data of one participant was not recorded due to program crash in the moral inference task, thus data of 103 participants was analyzed moral inference task.

(main effect of Agent: F(1,99)=4.82, p=0.030, η_p^2 = 0.05; main effect of SVO type, p=0.477, and SVO type × Agent interaction, p = 0.722).

Impression ratings.

Initial impressions of the two agents' character did not differ before observing their choices (Agent x SVO-type ANOVAs on the initial impression rating, Agent: F(1,101) = 0.57, p = 0.451, η_p^2 < 0.01; SVO-type: F(1,101) = 2.65, p = 0.107, η_p^2 = 0.03; Agent x SVO-type: F(1,101) = 0.24, p = 0.629, η_p^2 < 0.01, Figure 2b). We analyzed the final impression of the two agents by conducting Agent-by-SVO type ANOVA. This analysis revealed a significant Agent x SVO-type interaction (F(1,101) = 5.01, p = 0.027, η_p^2 = 0.05; Figure 2c) because prosocials (t(52)=7.914, p<0.001, mean difference ± SE (good vs. bad agent) = 0.36 ± 0.05, 95% CI = [0.27, 0.46], Cohen d' = 1.09) differentiated the impression of the good and bad agents to a greater degree than individualists did (t(49)=4.60, p<0.001, mean difference ± SE (Good vs. Bad agent) = 0.22 ± 0.33, 95%CI = [0.12, 0.31], Cohen d' = 0.65). Here despite that we observed a pattern of both better impression of good agent and worse impression of bad agent in prosocials (relative to individualists), post-hoc t-test results did not reach significant (good agent: prosocial vs. individualists: t(101) = 1.47, p = 0.145, 95%CI = [-0.02, 0.14], Cohen d' = 0.15; bad agent: prosocials vs. individualists: t(101) = -1.79, p = 0.076, 95%CI = [-0.19, 0.01], Cohen d' = -0.18).

Results of uncertainty ratings, as well as belief estimates (κ), volatility estimate (ω) in computational modeling estimates are provided in the Supplementary Results (Section1 & 2; Figure S1).

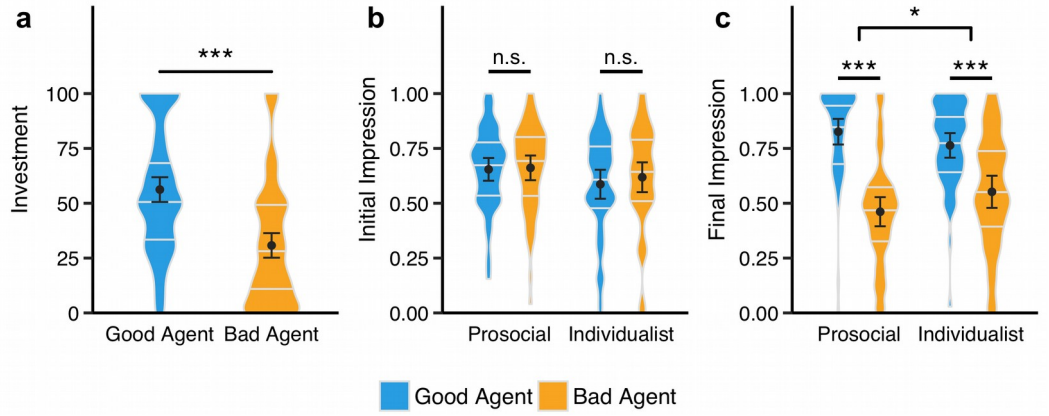

Figure 2. Participants learned the moral characters of the two agents in moral inference task. (a) Participants invested more money to good agent than bad agent after moral inference task; (b) Initial and

(c) final impression ratings of the good and bad agents. Participants rated their impression on a continuous scale from "nasty" to "nice" (rating scores z-transformed ranging from 0 to 1). Initial impression as well as the final impression at the end, presented separately for individualists and prosocials. All participants rated the good agent significantly nicer than the bad agent. Prosocials reported more distinguished impressions between the good and bad agents than individualists. The violin plots indicate kernel probability density estimation. Solid lines inside represent median and quartiles respectively. Black points represent mean value with error bars representing 95% confidential interval. $p<0.05$ *, $p < 0.01$ **, $p < 0.001$ ***; n.s., not significant.

Vicarious optimism is amplified for morally good agents.

To examine participants' optimistic beliefs for different targets (i.e., themselves, the good agent, and the bad agent), we first compared the initial estimates (1st estimation, E1) of the likelihood of encountering adverse life events in the future. A one-way ANOVA on E1 revealed a significant main effect of Target (self *vs.* good agent *vs.* bad agent), $F(2, 206) = 17.67$, $p < 0.001$, $\eta_p^2 = 0.15$, Figure 3a), as participants believed that negative life events were less likely to happen to themselves ($M \pm SE = 22.64 \pm 1.34$, 95%CI = [-9.39,-4.22]) and the good agent ($M \pm SE = 22.43 \pm 1.24$) than the bad agent ($M \pm SE = 29.45 \pm 1.43$) (self vs. bad agent: $t(103) = -5.23$, $p < 0.001$, 95%CI = [-9.39,-4.22], Cohen d' = -0.51; good vs. bad agents: $t(103) = -4.87$, $p < 0.001$, 95%CI = [-9.87,-4.16], Cohen d' = -0.48; self vs. good agent: $t(103) = 0.16$, $p = 0.872$, 95%CI = [-2.33,2.74], Cohen d' = 0.02, Figure 3a). Similarly, we observed the same pattern of results examining optimism via LOT score. The one-way ANOVA on LOT scores revealed significant main effect of Target ($F(2, 206) = 83.03$, $p < .001$, $\eta_p^2 = 0.45$). Participants were explicitly more optimistic for themselves ($M \pm SE = 21.63 \pm 0.32$) and the good agent ($M \pm SE = 21.91 \pm 0.32$) than for the bad agent ($M \pm SE = 16.92 \pm 0.28$) (self vs. bad agent, $t(103) = 11.12$, $p < 0.001$, 95%CI = [3.86, 5.54], Cohen d' = 1.09; good agent vs. bad agent, $t(103) = 10.75$, $p < 0.001$, 95%CI = [4.07, 5.91], Cohen d' = 1.05; self vs. good agent, $t(103) = -0.69$, $p = 0.490$, 95%CI = [-1.11, 0.54], Cohen d' = -0.07).

Next, we investigated the effect of Target (i.e., self, good agent, and bad agent) on belief updating. We conducted an ANOVA on belief updating upon desirable and undesirable feedback with Target (self, good agent, and bad agent) and Feedback (desirable vs. undesirable) as within-subject factors. We found a significant interaction of Target and Feedback ($F(2, 206) = 4.10$, $p = 0.018$, $\eta_p^2 = 0.04$, Figure 3b; remained reliable after controlling estimation error: $F(2, 200) = 5.18$, $p = 0.006$, $\eta_p^2 = 0.05$). This interaction arose from the fact that participants showed optimistic biased belief updating (significantly higher

belief-update upon desirable than undesirable feedback) for themselves (mean difference ± SE = 4.72 ± 0.94, t(103) = 5.04, p <0.001, 95%CI = [2.86, 6.58], Cohen d' = 0.49) and for the good agent (mean difference ± SE = 2.39 ± 1.04, t(103) = 2.30, p = 0.023, 95%CI = [0.33, 4.46], Cohen d' = 0.23), but not for the bad agent (mean difference ± SE = 0.85 ± 1.24, t(103) = 0.68, p = 0.497, 95%CI = [-1.62, 3.31], Cohen d' = 0.07). Optimistic biased belief updating for oneself and the good agent (but not the bad agent) was specifically driven by reduced updating from undesirable feedback. That is, there was a significant main effect of Target on undesirable belief-update ($F(2, 206) = 6.57$, $p = 0.002$, $\eta_p^2 = 0.06$), but no effect of Target on updating from desirable feedback ($F(2, 206) = 1.00$, $p = 0.368$, $\eta_p^2 = 0.01$).

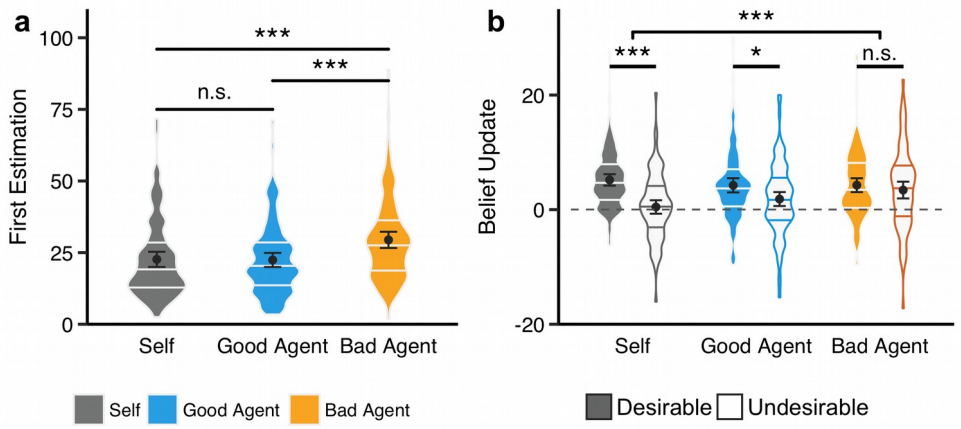

Figure 3. (a) Participants hold more optimistic prior beliefs about themselves and the good agent than the bad agent when estimating adverse future life events (b) Optimistic belief updating was modulated by the target: participants optimistically updated beliefs about the future for themselves and the good agent, but not for the bad agent. (The violin plots indicate kernel probability density estimation. Solid lines inside represent median and quartiles respectively. Black points represent mean value with error bars representing 95% confidential interval. $p < 0.05$ *, $p < 0.01$ **, $p < 0.001$ ***; n.s., not significant).

Social value orientation modulates vicarious optimism.

A mixed-model ANOVA with Target (self, Good agent, and Bad agent) and SVO type (prosocial vs. individualist) on the 1st estimation showed no effect of SVO type ($F(1, 102) = 1.95$, $p = 0.166$, $\eta_p^2 = 0.02$) nor its interaction with Target ($F(2, 204) = 1.80$, $p = 0.168$, $\eta_p^2 = 0.02$), indicating that SVO type did not affect participants' first estimate across Targets. In other words, SVO type did not influence participants' explicit vicarious optimism, and the same conclusion was confirmed by the LOT questionnaire scores (see Supplementary Results Section4 & Figure S2).

We then asked whether SVO-type affected the belief updating processes for each of the target conditions. We conducted a mixed model ANOVA on the optimistic biased belief updating, with Target (self *vs.* good agent *vs.* bad agent) as a within-subject variable and SVO-type as a between-subject variable. This analysis revealed a significant interaction of Target × SVO-type (linear test of within subjects contrast, $F(1,102) = 5.19$, $p = 0.025$, $\eta_p^2 = 0.05$), suggesting the modulation of SVO-type on optimistic biased belief updating was strongest for the good agent and smallest for the self (SVO-type effect: good agent > bad agent > self).

Next, we examined the modulation of SVO-type on optimism for self and others. For optimism for self, a one-way ANOVA with SVO type as a between-subject variable showed no difference in optimistic bias for the self between prosocials and individualists ($F(1,102) = 0.02$, $p = 0.888$, $\eta_p^2 < 0.01$; Figure 4a). For vicarious optimism for the two agents, we conducted repeated-ANOVA with Target (good *vs.* bad agent) as a within-subject variable and SVO-type as a between-subject variable. Results showed a significant interaction of Target × SVO-type ($F(1,102) = 5.19$, $p = 0.025$, $\eta_p^2 = 0.05$; Figure 4b), which remained reliable even after controlling for age, gender and estimation errors ($F(1,98) = 4.02$, $p = 0.048$, $\eta_p^2 = 0.04$). Specifically, prosocials showed stronger vicarious optimism for the good agent than the bad agent (mean difference ± SE (Good agent vs. Bad agent) = 4.51 ± 1.95, $t(52) = 2.31$, $p = 0.025$, 95%CI = [0.60, 8.42], Cohen d' = 0.32). Moreover, prosocials only showed optimistic biased belief updating for the good agent (M ± SE = 4.88 ± 1.46, $t(52) = 3.33$, $p = 0.002$, 95%CI = [1.94,7.81], Cohen d' = 0.46) but not for the bad agent (M ± SE = 0.37 ± 1.88, $t(52) = 0.20$, $p = 0.844$, 95%CI = [-3.34,4.07], Cohen d' = 0.03). To further understand the effect that prosocials hold stronger vicarious optimism for good than bad agents, we conducted Target-by-Feedback ANOVA on the belief-update upon desirable and undesirable feedback. This analysis revealed that such effect was driven by higher belief-update upon undesirable feedback for bad than good agents (bad vs. good, M±SD=0.27±6.18; $t(52) = 2.64$, $p = 0.011$, 95%CI = [0.73,5.30], Cohen d' = 0.36) but comparable belief-update upon desirable feedback for the two agents (bad vs. good, $t(52) = -1.50$, $p = 0.194$, 95%CI = [-3.77,0.78], Cohen d'=-0.18). Individualists, however, did not show a significant main effect of Target (mean difference ± SE (good agent vs. gad agent) = -1.53 ± 1.79, $t(50) = -0.86$, $p = 0.396$, 95%CI = [-5.13, 2.06], Cohen d' = -0.12). Furthermore, individualists showed no vicarious optimism for either the good agent (M ± SE = -0.18 ± 1.41, $t(50) = -0.13$, $p = 0.897$, 95%CI = [-3.01, 2.64], Cohen d' = 0.02) nor the bad agent (M ± SE = 1.35±

1.64, t(50) = 0.81, p = 0.424, 95%CI = [-2.01,4.71], Cohen d' = 0.11). These results indicated that, while individualists did not show vicarious optimism for the good or the bad agent, prosocials showed vicarious optimism for only the good agent.

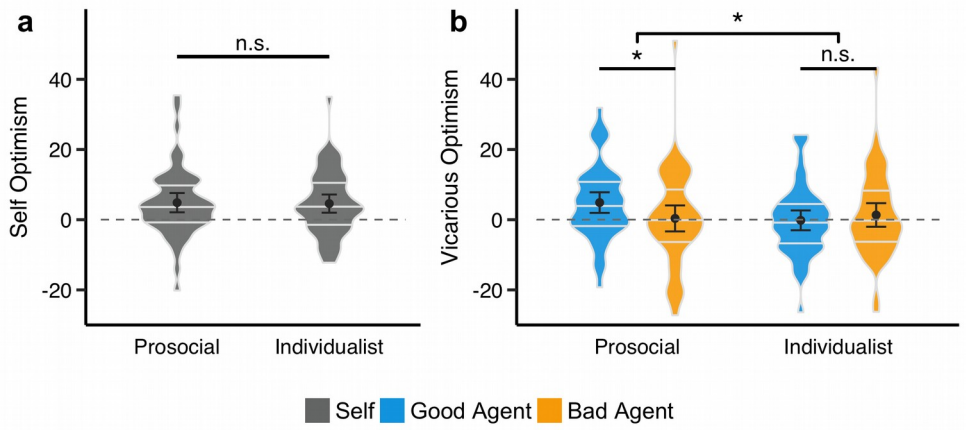

Figure 4. Self optimism and vicarious optimism in prosocial and individualistic participants. (a) Social value orientation did not affect participants' optimism bias for themselves. (b) Social value orientation modulated participants' optimism bias for the good agent and bad agent. Prosocial individuals hold more optimism for good agent than bad agent, but individualists did not have a distinct optimism bias between good and bad people. (The violin plots indicate kernel probability density estimation. Solid lines inside represent median and quartiles respectively. Black points represent mean value with error bars representing 95% confidential interval. $p<0.05$ *, $p < 0.01$ **, $p < 0.001$ ***; n.s., not significant.

Discussion

Karma denotes the belief that good things will happen to people who have done good deeds, while misfortunes will befall bad people in the future. In the current studies, combining the moral character learning and vicarious belief update tasks, we are able to quantify the beliefs about the future of people who have done objectively good or bad deeds. We show that individuals hold optimistic beliefs about the future of good people and discount undesirable feedback when predicting their futures. In contrast, individuals similarly incorporate desirable and undesirable feedback into their beliefs about bad people's futures. These results suggest that vicarious optimism is one possible cognitive mechanism that gives rise to karmic beliefs. Furthermore, we show that prosocial individuals (relative to individualists) hold stronger karmic beliefs and stronger vicarious optimism for good relative to bad people, suggesting that karmic beliefs are self-serving: good people more likely to believe that good things happen to good people.

We provide evidence for a correlation between prosociality and karmic beliefs. However, the causal direction of this relationship remains open to discussion, and is likely bidirectional. Previous studies showed that priming of karmic beliefs increased generosity and prosocial behavior (White, Kelly, Shariff, & Norenzayan, 2019), suggesting that karmic beliefs may be a precursor to prosocial behavior. However, studies developmental work suggests that prosocial behavior may emerge earlier than karmic beliefs; preverbal infants (6-10 months) show disapproval of antisocial behavior (Hamlin, Wynn, & Bloom, 2007) and infants between 12 and 24 months exhibit prosocial behaviors (Brownell, 2013), whereas karmic beliefs have only been demonstrated in 4-6-year-old children (Banerjee & Bloom, 2013, 2017). Thus, it may also be the case that prosociality promotes the development of karmic beliefs. Prosocial behavior is often costly (Crocker, Canevello, & Brown, 2017). Karmic beliefs that morally good behavior will be rewarded could provide one type of justification for these costs and serve as a psychological compensation (Bäckman & Dixon, 1992). In addition, helping others also brings positive "side effects" (Carlson & Zaki, 2018), such as positive feelings (Aknin, Van de Vondervoort, & Hamlin, 2018) and social praise (Eisenberg, Wolchik, Goldberg, & Engel, 1992). Thus over time, the beliefs that performing good deeds increases the chance of future desirable outcomes may be reinforced into a karmic worldview.

People hold karmic beliefs in both first-party and third-party contexts (Hafer & Olson, 1989). If prosocials and individualists hold karmic beliefs to a similar extent, we might expect strong optimistic belief updating for the self in prosocials, and pessimistic belief updating for the self in individualists. However, we observed that prosocials and individualists were similarly optimistic about their own futures. One potential explanation is that the wishful thinking for oneself outweighs karmic believes when there are any conflicts (Mata & Simão, 2019). Alternatively, individualists may not identify themselves as "bad people" given vast evidence that most people tend to view themselves in a positive light (Sanitioso & Wlodarski, 2004). Thus, it is possible that individualists believe in karma but view themselves as good people who deserve an optimistic future. Given that we found weaker karmic beliefs in individualists, the finding of similar optimistic beliefs for the self in individualists and prosocials would lend further support to our hypothesis that karmic beliefs are self-serving, so that strong karmic beliefs motivates prosocial individuals to believe in a bright future

(possibly caused by the good deeds they did). Taken together, this suggests that the self-serving nature of karmic beliefs applies to both the self and other people.

In current study, individualists not only failed to show asymmetric vicarious optimism towards good and bad agents; they also did not show vicarious optimism at all. Consistent with previous findings that individualists maximize the differences between the self and others in allocating monetary reward (Haruno & Frith, 2010; Liu et al., 2019) or responding to painful stimuli (Singer et al., 2008), individualists also differentiate optimistic future beliefs toward the self and others (only showing optimism towards self, but not to others: $t(50) = 2.35$, $p = 0.023$, 95% CI = [0.58, 7.43], Cohen d' = 0.33). Taken together, this suggests individualists prefer to maximize self-other differences not only in material outcomes (i.e., monetary allocation, physical pain) but also in immaterial beliefs about the future.

One limitation of our second study is that we only provide evidence for 'half' of the karmic worldview, i.e., that good things will happen to good people; we did not observe evidence for beliefs that bad things will happen to bad people. Even prosocials who showed stronger karmic beliefs did not express pessimistic beliefs about bad agents. This might be due to that, in prosocials' karmic belief system, good things not happening is already a type of punishment for the bad people, given that prosocial generally care about others and prefer not to do harm to others (Penner., Dovidio., Piliavin., & Schroeder., 2005), thus prosocials do not predict bad consequence for morally bad people. Indeed, the current sample, we found evidence that prosocials showed stronger harm aversion in the moral decision task where they trade off profit for themselves against pain for another person (harm aversion: prosocials vs. individualists, $t(101) = 4.39$, $p < 0.001$, 95% CI = [0.11, 0.29], Cohen d' = 0.43).

In conclusion, the current study provides a novel framework to decipher the cognitive processes that give rise to karmic beliefs, and further proposes that karmic beliefs may be subject to self-serving motivations. Our findings suggest that karmic beliefs – a feature of many religious traditions – may be a key component of a positive feedback loop between beliefs and behavior that together contribute to large-scale cooperation.

Author Contributions

Y. Ma. conceived the project. S. Zheng., S. Li., and Y. Ma. designed research. S. Zheng., X. Yan., and S. Li. performed research. S. Zheng., X. Yan., and J. Siegel. analyzed and interpreted data under the supervision of M. Crockett and Y. Ma. All authors contributed to paper-writing.

Open Practices Statement

Neither of the experiments reported in this article was formally preregistered. The data have not been available on a permanent third-party archive. Experiment materials are available at https://osf.io/znw89/. Further requests for the data or materials can be sent via email to the lead author at yma@bnu.edu.cn.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (Projects 31722026; 31771204 to YM), the Fundamental Research Funds for the Central Universities (2018EYT04 to YM); the Open Research Fund of the State Key Laboratory of Cognitive Neuroscience (to YM and MC).

Reference

Aknin, L. B., Van de Vondervoort, J. W., & Hamlin, J. K. (2018). Positive feelings reward and promote prosocial behavior. Current Opinion in Psychology, 20, 55–59. https://doi.org/10.1016/j.copsyc.2017.08.017

Bäckman, L., & Dixon, R. A. (1992). Psychological compensation: A theoretical framework. Psychological Bulletin, 112(2), 259–283. https://doi.org/10.1037/0033-2909.112.2.259

Banerjee, K., & Bloom, P. (2013). Would Tarzan believe in God? Conditions for the emergence of religious belief. Trends in Cognitive Sciences, 17(1), 7–8. https://doi.org/10.1016/j.tics.2012.11.005

Banerjee, K., & Bloom, P. (2017). You get what you give: Children's karmic bargaining. Developmental Science, 20(5), e12442. https://doi.org/10.1111/desc.12442

Brownell, C. A. (2013). Early Development of Prosocial Behavior: Current Perspectives. Infancy, 18(1), 1–9. https://doi.org/10.1111/infa.12004

Carlson, R. W., & Zaki, J. (2018). Good deeds gone bad: Lay theories of altruism and selfishness. Journal of Experimental Social Psychology, 75, 36–40. https://doi.org/10.1016/j.jesp.2017.11.005

Crocker, J., Canevello, A., & Brown, A. A. (2017). Social Motivation: Costs and Benefits of Selfishness and Otherishness. Annual Review of Psychology, 68(1), 299–325. https://doi.org/10.1146/annurev-psych-010416-044145

Crockett, M. J., Kurth-Nelson, Z., Siegel, J. Z., Dayan, P., & Dolan, R. J. (2014). Harm to others outweighs harm to self in moral decision making. Proceedings of the National Academy of Sciences, 111(48), 17320–17325. https://doi.org/10.1073/pnas.1408988111

Dalbert, C. (1999). The world is more just for me than generally: About the personal belief in a just world scale's validity. Social Justice Research, 12(2), 79–98.

Delgado, M. R., Frank, R. H., & Phelps, E. A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. Nature Neuroscience, 8(11), 1611–1618. https://doi.org/10.1038/nn1575

Eisenberg, N., Wolchik, S. A., Goldberg, L., & Engel, I. (1992). Parental Values, Reinforcement, and Young Children's Prosocial Behavior: A Longitudinal Study. The Journal of Genetic Psychology, 153(1), 19–36. https://doi.org/10.1080/00221325.1992.10753699

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. Behavior Research Methods, 41(4), 1149–1160.

Hafer, C. L., & Olson, J. M. (1989). Beliefs in a just world and reactions to personal deprivation. Journal of Personality, 57(4), 799–823.

Hafer, C. L., & Olson, J. M. (1993). Beliefs in a just world, discontent, and assertive actions by working women. Personality and Social Psychology Bulletin, 19(1), 30–38.

Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. Nature, 450(7169), 557–559. https://doi.org/10.1038/nature06288

Haruno, M., & Frith, C. D. (2010). Activity in the amygdala elicited by unfair divisions predicts social value orientation. Nature Neuroscience, 13(2), 160–161. https://doi.org/10.1038/nn.2468

Jenkins, A. C., Macrae, C. N., & Mitchell, J. P. (2008). Repetition suppression of ventromedial prefrontal activity during judgments of self and others. Proceedings of the National Academy of Sciences, 105(11), 4507–4512. https://doi.org/10.1073/pnas.0708785105

Kappes, A., Faber, N. S., Kahane, G., Savulescu, J., & Crockett, M. J. (2018). Concern for Others Leads to Vicarious Optimism. Psychological Science, 29(3), 379–389. https://doi.org/10.1177/0956797617737129

Karagonlar, G., & Kuhlman, D. M. (2013). The role of Social Value Orientation in response to an unfair offer in the Ultimatum Game. Organizational Behavior and Human Decision Processes, 120(2), 228–239. https://doi.org/10.1016/j.obhdp.2012.07.006

Kopalle, P. K., Lehmann, D. R., & Farley, J. U. (2010). Consumer expectations and culture: The effect of belief in karma in India. Journal of Consumer Research, 37(2), 251–263.

Krishan, Y. (1997). The Doctrine of Karma: Its Origin and Development in Brāhmaṇical, Buddhist, and Jaina Traditions. Bharatiya Vidya Bhavan.

Kulow, K., & Kramer, T. (2016). In Pursuit of Good Karma: When Charitable Appeals to Do Right Go Wrong. Journal of Consumer Research, 43(2), 334–353. https://doi.org/10.1093/jcr/ucw018

Lerner, M. J. (1980). The belief in a just world. In The Belief in a just World (pp. 9–30). Springer.

Lipkusa, I. M., Dalbert, C., & Siegler, I. C. (1996). The Importance of Distinguishing the Belief in a Just World for Self Versus for Others: Implications for Psychological Well-Being. Personality and Social Psychology Bulletin, 22(7), 666–677. https://doi.org/10.1177/0146167296227002

Ma, Y., Li, S., Wang, C., Liu, Y., Li, W., Yan, X., … Han, S. (2016). Distinct oxytocin effects on belief updating in response to desirable and undesirable feedback. Proceedings of the National Academy of Sciences, 113(33), 9256–9261. https://doi.org/10.1073/pnas.1604285113

Mata, A., Simão, C. (2019). Karmic forecasts: The role of justice in forecasts about self and others. Motivation Science. Advance online publication. https://doi.org/10.1037/mot0000162

Mathys, C., Daunizeau, J., Friston, K. J., & Stephan, K. E. (2011). A Bayesian foundation for individual learning under uncertainty. Frontiers in Human Neuroscience, 5, 39.

Murphy, R. O., Ackermann, K. A., & Handgraaf, M. (2011). Measuring Social Value Orientation. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1804189

Norenzayan, A., Shariff, A. F., Gervais, W. M., Willard, A. K., McNamara, R. A., Slingerland, E., & Henrich, J. (2016). The cultural evolution of prosocial religions. Behavioral and Brain Sciences, 39. https://doi.org/10.1017/S0140525X14001356

Penner., L. A., Dovidio., J. F., Piliavin., J. A., & Schroeder., D. A. (2005). Prosocial Behavior: Multilevel Perspectives. Annual Review of Psychology, 56(1), 365–392. https://doi.org/10.1146/annurev.psych.56.091103.070141

Sanitioso, R. B., & Wlodarski, R. (2004). In Search of Information That Confirms a Desired Self-Perception: Motivated Processing of Social Feedback and Choice of Social Interactions. Personality and Social Psychology Bulletin, 30(4), 412–422. https://doi.org/10.1177/0146167203261882

Scheier, M. F., Carver, C. S., & Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A reevaluation of the Life Orientation Test. Journal of Personality and Social Psychology, 67(6), 1063–1078. https://doi.org/10.1037/0022-3514.67.6.1063

Sharot, T., Korn, C. W., & Dolan, R. J. (2011). How unrealistic optimism is maintained in the face of reality. Nature Neuroscience, 14(11), 1475–1479. https://doi.org/10.1038/nn.2949

Siegel, J. Z., Mathys, C., Rutledge, R. B., & Crockett, M. J. (2018). Beliefs about bad people are volatile. Nature Human Behaviour, 2(10), 750–756. https://doi.org/10.1038/s41562-018-0425-1

Singer, T., Snozzi, R., Bird, G., Petrovic, P., Silani, G., Heinrichs, M., & Dolan, R. J. (2008). Effects of oxytocin and prosocial behavior on brain responses to direct and vicariously experienced pain. Emotion, 8(6), 781–791. https://doi.org/10.1037/a0014195

Taylor, D. M., & Doria, J. R. (1981). Self-Serving and Group-Serving Bias in Attribution. The Journal of Social Psychology, 113(2), 201–211. https://doi.org/10.1080/00224545.1981.9924371

Van Lange, P. A. M. (1999). The pursuit of joint outcomes and equality in outcomes: An integrative model of social value orientation. Journal of Personality and Social Psychology, 77(2), 337–349. https://doi.org/10.1037/0022-3514.77.2.337

White, Cindel Jennifer Melina. (2017). Belief in karma: The content and correlates of supernatural justice beliefs across cultures. University of British Columbia.

White, Cindel J.M., Kelly, J. M., Shariff, A. F., & Norenzayan, A. (2019). Supernatural norm enforcement: Thinking about karma and God reduces selfishness among believers. Journal of Experimental Social Psychology, 84, 103797. https://doi.org/10.1016/j.jesp.2019.03.008

Supporting Material

Supporting Method

Moral Inference Task

Trials were created using the same methods described in Siegel et al. 2018. Each trial contained a pair of choices that matched the indifference point of a specific κ value. We first created a set of 25 trials with values of κ that were normally distributed around the Good agent's indifferent point (mean = 0.7, standard deviation = 0.15). Next, we generated a corresponding set of 25 trials around the Bad agent's indifference point by subtracting each κ value from 1. Given a set of κ values, we then generated shock and money options for each κ value by generating 10,000 random pairs of positive shock movements Δs ($1 < \Delta s < 20$), and positive money movements Δm ($0.10 < \Delta m < 19.90$), and selected the pair closest to the indifference point of that κ value [Δs, Δm]. Next, these pairs were transformed into choices containing fewer amounts of shocks and money (s- and m-) and greater amounts of shocks and money (s+ and m+) as follows: s- was a positive integer between 0 and 20, randomly drawn from a uniform discrete distribution with the constraint that $0 < s- + \Delta s < 20$. Similarly, m- was a positive number between 0 and 20, randomly drawn from a uniform discrete distribution, rounded to the nearest 10th and constrained such that $0 < m \pm \Delta m < 20$. s+ and m+ were then set by adding Δs and Δm to s- and m-, respectively.

We simulated the agents' decisions by computing the utility for choosing the more harmful option over the less harmful option as a function of the agent's κ ($\kappa_{Bad} = 0.3$, $\kappa_{Good} = 0.7$). This model has been validated in previous studies and best predicts human choices in the moral decision task (Crockett, Kurth-Nelson, Siegel, Dayan, & Dolan, 2014):

$$\Delta V = (1 - \kappa_i)\,\Delta m + \kappa_i \Delta s$$

Where κ_i is the κ for agent i. The trial-by-trial value differences can be transformed into the choice probabilities using a softmax function $P = \frac{1}{1+e^{-\Delta V}}$. The final decisions in every trial were further simulated depending on the probabilities.

All the trials were presented in a fixed order to eliminate order effects and participants were randomized to complete one of two sets of trial sequences.

Supporting Results for study2

Section 1. Uncertainty ratings in moral inference task

Similar with impression ratings analyses, we conducted Agent-by-SVO type ANOVA on final uncertainty ratings. We found significant main effect of Agent ($F(1,101) = 10.905$, $p = 0.001$, $\eta_p^2 = 0.097$), and no significant interaction effect was found ($F(1,101) = 0.064$, $p = 0.801$, $\eta_p^2 = 0.001$). These results suggested that both prosocials and individualists showed less uncertainty for Good agent than Bad agent ($t(102) = -3.312$, $p = 0.001$, mean difference ± SE (Good vs. Bad agent) = -7.557 ± 2.282, 95% CI = [-12.083, -3.031], Cohen's $d = 0.326$).

Section 2. Belief estimate(κ) and volatility estimate(ω) in moral inference task

To examine whether individual differences in SVO type modulate estimates derived from the model, we conducted separate mixed-model ANOVAs with Agent (good vs. bad) and SVO type (prosocial vs. individualist) on final belief about the agents' κ. As expected, there was a significant main effect of Agent on final beliefs ($F(1,101) = 3251.042$, $p < 0.001$, $\eta_p^2 = 0.970$), indicating higher beliefs about the Good agent's κ than the Bad agent's. We also found a significant main effect of SVO type on final beliefs ($F(1,101) = 10.505$, $p = 0.002$, $\eta_p^2 = 0.094$), where prosocials had lower beliefs about the agents' κ than individualists. The interaction between Agent and SVO type was not significant ($F(1,101) = 0.071$, $p = 0.790$, $\eta_p^2 = 0.001$; Figure S1A).

For volatility estimate, we found a significant effect of Agent on ω ($F(1,101) = 7.995$, $p = 0.006$, $\eta_p^2 = 0.073$), which was qualified by a significant interaction between Agent and SVO type ($F(1,101) = 6.819$, $p = 0.010$, $\eta_p^2 = 0.063$; Figure S1B), no significant main effect of SVO type was found ($F(1,101) = 0.786$, $p = 0.378$, $\eta_p^2 = 0.008$). Simple effects analysis indicated that beliefs about the Bad agent were more volatile for individualists than prosocials ($t(101) = 2.475$, $p = 0.015$, mean difference ± SE (prosocials vs. individualists) = 0.325 ± 0.131, 95% CI = [0.064, 0.586], Cohen's d = 0.245), but beliefs about the Good agent were similarly volatile between groups ($t(101) = -0.979$, $p = 0.330$, mean difference ± SE (prosocials vs. individualists) = -0.140 ± 0.143, 95% CI = [-0.423,0.143], Cohen's d = -0.096).

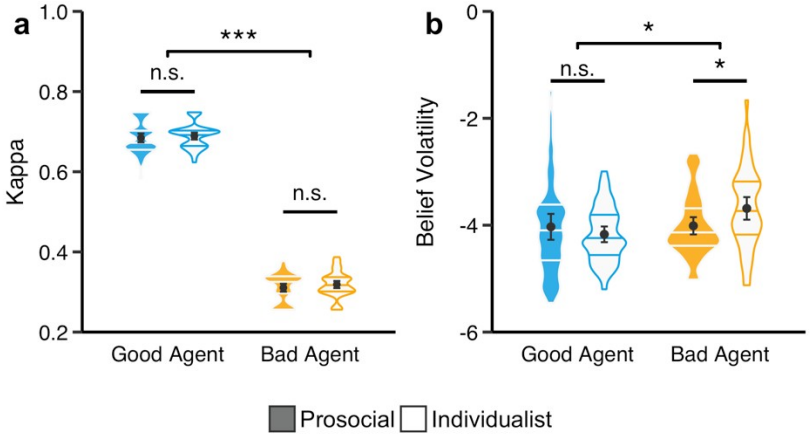

Figure S1. Results of moral character inference task for prosocials and invidualists. (a) Participants would have higher estimates for Good agent's moral character than Bad agent, SVO type did not modulate this effect; (b)Volatility estimate in modeling as a function of Target and SVO type; prosocials had less belief volatility for estimating Bad agent's moral

character than individualists, both prosocials and individualists had similar volatility when estimated Good agent. (Lines depicted inside the violin plots indicate the median and quartiles; plots and error bars represent mean values and 95% confidential intervals respectively; $p<0.05$ *, $p < 0.01$ **, $p < 0.001$ ***, n.s. not significant.)

Section 3. Moral inference shapes explicit vicarious optimism

We first examined participants' explicit optimistic beliefs about themselves, the Good agent and the Bad agent, which was measured by the modified life orientation test (LOT, see Method). The one-way ANOVA on LOT scores with Target (self *vs.* Good agent *vs.* Bad agent) as within-subject variable revealed significant main effect of Agent ($F(2, 206) = 83.028$, $p < .001$, $\eta_p^2 = 0.446$). *Post hoc* analysis showed that participants showed less optimism for Bad agent (M±SE = 16.923±0.284) than for themselves (M±SE =21.625±0.321; self vs. Bad agent, $t(103) = 11.118$, $p < 0.001$, mean difference ± SE (self vs. Bad) = 4.701 ± 0.422, 95%CI = [3.863, 5.540], Cohen d' = 1.090;) and for Good agent (M±SE = 21.913±0.322; Good agent vs. Bad agent, $t(103) = 10.753$, $p < 0.001$, mean difference ± SE (self vs. Bad) = 4.990 ± 0.464, 95%CI = [4.070, 5.910], Cohen d' = 1.054), indicating less optimism about the Bad agent. Moreover, participants showed comparable optimistic beliefs about themselves and the Good agent ($t(103) = -0.693$, $p = 0.490$, mean difference ± SE (self vs. Bad) = -0.288 ± 0.415, 95%CI = [-1.113, 0.536], Cohen d' = -0.067). These results indicated that people hold explicit optimistic beliefs towards others, only for those about whom they infer a good moral character.

Section 4. Social value orientation did not modulate explicit vicarious optimism

A mixed-model ANOVAs with Target (self vs. Good agent vs. Bad agent) and SVO type (prosocial vs. individualist) on LOT scores showed no differences between prosocials and individualists in the magnitude of optimism towards the self, Good and Bad agents (main effect of SVO type: $F(1,102) = 0.261$, $p = 0.611$, $\eta_p^2 = 0.003$; the interaction between SVO type and Target: $F(2,204) = 0.566$, $p = 0.569$, $\eta_p^2 = 0.006$. Figure S2). For each target, no difference was found between prosocials and individualists (Self: $p=0.994$; Good Agent: $p=0.793$; Bad Agent: $p = 0.270$), furtherly confirming that explicit optimism was not modulated by social value orientation.

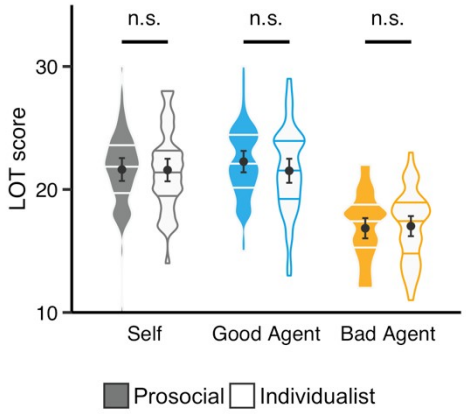

Figure S2. Results of life orientation test for prosocials and invidualists. Higher LOT score indicated more optimism. Social value orientation (SVO type) did not show moderation effects on LOT scores. (Lines depicted inside the violin plots indicate the median and quartiles; plots and error bars represent mean values and 95% confidential intervals respectively; n.s., not significant).

Table S1. Questionnaires comparison between prosocials and individualists in Study 2

	Prosocials	Individualists	t value	p value	95% CI
IRI	93.962	93.392	-0.298	0.766	(-4.366, 3.225)
emotional ERQ	31.415	29.725	-1.629	0.106	(-3.746, 0.367)
cognitive ERQ	16.603	15.902	-0.724	0.471	(-2.625, 1.221)
TA	45.698	44.588	-0.646	0.520	(-4.518, 2.298)
PWB	78.452	81.039	1.174	0.212	(-1.781, 6.954)
LS	18.226	17.509	-0.607	0.545	(-3.056, 1.623)

IRI: Interpersonal Reactivity Index
ERQ: Emotion Regulation Questionnaire
TA: Trait anxiety
PWB: Psychological Well-Being

Table S2. Results for Study 2 including filter trials.

Dependent variables	Analysis	Effect	F	p value	η_p^2
GLM1: Belief updating	Repeated ANOVA with Target (self vs. Good vs. Bad agent) and Feedback (desirable vs. undesirable)	Target	1.096	0.336	0.011
		Target × Feedback	2.450	0.089 †	0.023
GLM2: Optimism	Univariate analysis	SVO type	0.617	0.434	0.006
GLM2: Vicarious optimism	Mixed ANOVA with Agent (Good vs. Bad) as within-subject variable, SVO type as between-subject variable	Agent × SVO type	4.136	0.045*	0.039

$p < 0.10$†, $p < 0.05$*, $p < 0.01$**, $p < 0.001$***.