Exposure to Robot Preachers Undermines Religious Commitment

Joshua Conrad Jackson[1], Kai Chi Yam[2], Pok Man Tang[3], Ting Liu[4], and Azim Shariff[5]

[1] Management and Organizations Department, Kellogg School of Management, Northwestern University
[2] Department of Management and Organization, National University of Singapore
[3] Department of Management, Terry School of Business, University of Georgia
[4] Department of Business Administration, Graduate School of Management, Kyoto University
[5] Department of Psychology, University of British Columbia

Over the last decade, robots continue to infiltrate the workforce, permeating occupations that once seemed immune to automation. This process seems to be inevitable because robots have ever-expanding capabilities. However, drawing from theories of cultural evolution and social learning, we propose that robots may have limited influence in domains that require high degrees of "credibility"; here we focus on the automation of religious preachers as one such domain. Using a natural experiment in a recently automated Buddhist temple (Study 1) and a fully randomized experiment in a Taoist temple (Study 2), we consistently show that religious adherents perceive robot preachers—and the institutions which employ them—as less credible than human preachers. This lack of credibility explains reductions in religious commitment after people listen to robot (vs. human) preachers deliver sermons. Study 3 conceptually replicates this finding in an online experiment and suggests that religious elites require perceived minds (agency and patiency) to be credible, which is partly why robot preachers inspire less credibility than humans. Our studies support cultural evolutionary theories of religion and suggest that escalating religious automation may induce religious decline.

Public Significance Statement

The rise of robots and artificial intelligence is changing how people work and live. The economic and political consequences of this trend have been widely studied, but less is known about the cultural consequences. Here, we study how automation may instigate cultural change as robots move into a new domain: religion. We argue that automated preachers and sermon-writers may reduce religious commitment. According to cultural evolutionary theories of religion, religious elites are effective not only because they can capably carry out their duties, but also because they are viewed as highly credible. Robots are capable of writing and delivering sermons, but people do not view them as credible religious elites, a critical perception for sustaining religious commitment in adherents. Two field studies and an online experiment support our key claims and suggest that automation of religious duties may prompt declines in religious commitment.

Every year, automation and artificial intelligence (AI) further expand into domains of work and play held exclusively by human beings. Automated agents have become increasingly capable in the sectors of manufacturing, customer service, finance, and gaming (Brynjolfsson & McAfee, 2014; Ford, 2015; Frank et al., 2019).

The coming years promise deeper inroads into medicine (Topol, 2019), journalism (GPT-3, 2020), and even psychotherapy (Fitzpatrick et al., 2017) and prostitution (Devlin, 2020)—occupations that once seemed immune to automation because of the need for human understanding and contact. These inroads raise a

Joshua Conrad Jackson ![ORCID] https://orcid.org/0000-0002-2947-9815

Correspondence concerning this article should be addressed to Joshua Conrad Jackson, Kellogg School of Management, Northwestern University, Kellogg Global Hub, 2211 Campus Dr, Evanston, IL 60208, United States. Email: joshua.jackson@kellogg.northwestern.edu

provocative question about the impact and future of automation: Is there anything that robots cannot do?

We propose that answering this question requires putting aside domains of capability—where robots are continually improving—and instead focusing on domains of *special credibility*. Although people traffic in physical goods and services, they also invent and transmit intellectual goods such as ideas, beliefs, and norms. In these domains, success is just as dependent on the *credibility* of who is advancing ideas as the content of the ideas themselves. Nowhere is this more pronounced than within the domain of religion. For thousands of years, religions have elevated people to positions of moral and epistemic leadership. Clergy, shamans, and other religious elites have served as cultural models, who do not merely recite and transmit the tenets of their faiths but embody and legitimize them. Influential theories of religion suggest that the credibility of religious elites might have been instrumental in religious institutions' ability to maintain high levels of commitment over history and into the contemporary era (Henrich, 2009).

Now, as in many occupations, robots are beginning to take on roles of religious professionals. Mindar, a robot designed to look like the Buddhist deity of Mercy (see Figure 1), made headlines in 2019 after it began giving sermons in Japan's Kodaiji Temple (CNN, n.d.). In 2017, the robot "Bless-U-2" was installed in the Lutheran State Church at Wittenberg where it reads blessings in five languages The AI-enabled robot SanTO was distributed in 2019 to South American Catholics with the promise of listening to confessions and delivering personalized encouragement from the Bible (Deus Ex Machina: Religions Use Robots to Connect With the Public—WSJ, n.d.). These robots are still rare, and only time will tell if they spread widely or remain novelty attractions. But even these isolated cases raise the provocative question of whether robots can serve as effective religious professionals. Robot preachers are clearly capable of carrying out priestly duties, but are they viewed as credible? The automation of religion creates a unique opportunity for us to test whether robots can emulate human credibility in a critical domain, and whether automation's impact on the credibility of religious elites could result in declining religious commitment.

Figure 1

A View of Mindar During a Public Service at Kodaiji Temple in Kyoto

Note. See the online article for the color version of this figure.

Background Theory: Credibility Enhancing Displays and Religious Commitment

Research on cultural evolution has revealed the psychological features that underlie cultural transmission. Dual-inheritance theories of cultural evolution emphasize that humans thrive in many different environments because they benefit not just from a genetic inheritance of evolved adaptations, but a cultural inheritance of knowledge and skills that have been acquired over many generations (Boyd & Richerson, 1988; Cavalli-Sforza & Feldman, 1981). As a result, humans have evolved to be keen social learners, adopting several social learning biases which dictate the kind of information we are most likely to learn, and the kinds of people from whom we are most likely to learn (Henrich & McElreath, 2003; Kendal et al., 2018).

One critical example of these biases is an outsized tendency to learn from people who display that they are *credible*. Such "credibility enhancing displays" (CREDs) can be diverse, but they typically represent behaviors that would prove costly if an agent did not actually believe what they professed (a canonical example is eating a blue mushroom to indicate it is not poisonous—a costly act if the eater did not truly believe the mushroom was harmless; Henrich, 2009). More generally, these behaviors function to show that someone authentically holds their beliefs and that they can be trusted as credible messengers of this belief. People show a learning bias toward credible targets (Henrich, 2009), and rely on credibility when deciding whether to copy someone who advocates for behavior change (Kraft-Todd et al., 2018).

Religions, which depend on the spread and sustainment of complex sets of beliefs and practices, have been an empirical proving ground for much of the theorizing about cultural evolution (Henrich, 2009; Norenzayan et al., 2016). Early theories about the cultural spread of religious ideas exclusively focused on the psychological appeal of the contents of religious beliefs (Atran, 1998; Barrett, 2004; Boyer, 2007). But more recent theorizing has suggested that contextual factors—such as the credibility of the messenger—are indispensable to understanding why some religions catch on while others fall by the wayside of history (Gervais & Henrich, 2010). Many of the seemingly costly requirements of religious adherence, from frequent prayer to dietary restrictions to priestly celibacy, serve as CREDs of religious commitment. When religious elites do not meet these requirements—failing to demonstrate credibility—religious commitment may wither (Lanman, 2012; Lanman & Buhrmester, 2017).

Several studies support the credibility theory of religious commitment. For example, Lanman and Buhrmester (2017) found that childhood exposure to credible models predicted religiosity in adulthood. More recent surveys have supported this finding and showed that exposure to CREDs predicts deconversion better than other proposed mechanisms such as analytic thinking or existential insecurity (Gervais et al., 2021). Indeed, even among the formerly religious, greater childhood exposure to CREDs predicted remaining religious for longer (Langston et al., 2020). However, these studies are limited because they conflate the credibility of religious models with the religiosity of the household environment, so it is difficult to disentangle whether credibility uniquely increases religious commitment or whether people are simply more likely to remain religious as adults if they adopted religious belief from a young age and experienced more social pressure to remain religious as they grew older. These limitations are difficult to overcome because it is not easy to

systematically manipulate credibility in religious elites, especially in real religious communities that have high external validity. The emergence of robot preachers represents a rare case where we can ethically and feasibly conduct such an experimental manipulation.

Although robot preachers may be able to fulfill the role of transmitting the content of religion, they may nonetheless falter with credibility. Specifically, people perceive robots as having less mind than humans, especially in their capacity to feel and understand but also in their ability to think and decide (K. Gray & Wegner, 2012; H. M. Gray et al., 2007). These mind perception qualities are essential to cultivating credibility—a robot cannot authentically believe in supernatural agents if they do not have the capacity to believe, and they cannot engage in potentially costly behavior such as celibacy if they are not able to feel the cost. Unlike human religious elites, who profess a deep commitment to their faith which leads them to sacrifice time and material goods, robots are simply programmed to give sermons or blessings without an authentic understanding of, commitment to, or suffering for their religious group. While such displays of credibility may be unimportant in many of the other professions being outsourced to robots, they could be crucial for professions in the religious sphere. As a result, exposure to robot (vs. human) preachers may reduce perceptions of credibility, and this reduction in credibility should translate to less religious commitment.

Current Research

We test our hypothesis across two field studies and one online experiment, with a research program that aims to balance internal and external validity. Study 1 is a quasiexperimental field study conducted in the Buddhist Kodaiji Temple in Kyoto, Japan. We survey people after they watched a sermon delivered by the robot preacher Mindar versus a sermon delivered by a human preacher. We measure how much participants donate to the Temple at the end of the study as a measure of religious commitment. Study 2 is a preregistered experimental study in a Taoist temple in Singapore, where we randomly assign people to listen to the same sermon delivered by a robot or human preacher and measure multiple behavioral indicators of religious commitment. Study 3 is a preregistered online experiment where people read a sermon that they believe is either composed by a human preacher or an AI program. This paradigm gives us an opportunity to extend our research program to disembodied AI bots which now genuinely have the ability to write religious sermons (ChatGPT, 2022), and to include a broader set of measures. For example, we test whether mind perception explains why robot preachers are ascribed less credibility than human preachers, and we measure other potential mechanisms, such as likability, which could explain why exposure to robot preachers decreases religious commitment (Schjoedt et al., 2011; Sperber, 2010).

Across these studies, we measure religious commitment through (a) monetary donations to religious places of worship (Studies 1–2), (b) willingness to circulate information about the message of a sermon (Study 2), and (c) self-reported commitment to one's religious identity (Study 3). This operationalization is consistent with other field research on religious commitment (Xygalatas et al., 2013). But we also acknowledge that past literature has conceptualized and defined religious commitment in other ways, including through strength of religious beliefs (Cornwall et al., 1986), sense of closeness to a religious community (Wesselmann et al., 2016), investment in religious morals and values (Worthington Jr. et al., 2012), and participation in

religious rituals (Cornwall et al., 1986). Our measurement strategy means that we cannot make claims about how robot preachers affect these other aspects of religious commitment.

Transparency and Openness

All materials, data, and code are available on our OSF page: https://osf.io/4vfqa/?view_only=ba7463f035464d688c058d1dd38de30f. We preregistered the measures, procedure, and some analyses associated with Studies 2 and 3, and the preregistrations are also hosted on our OSF page in the folder "Preregistrations."

Study 1: Exposure to a Real-World Robot Preacher in a Buddhist Temple

Our first field study took place in Kodaiji Temple. In operation for over 400 years, the temple gained fame in 2019 when it introduced Mindar to preach sermons. We surveyed individuals leaving the Temple who had either seen Mindar give a sermon or had seen a human preacher give a sermon (no participant saw both). We had no a priori predictions, but we reasoned that a negative effect of robot preacher exposure on religious commitment (i.e., donation) would support dual inheritance theories of religion that stress the importance of credibility and would speak to the limits of robots in credibility-based professions. We conducted this data collection in conjunction with another project that had unrelated hypotheses. Some of our measures were therefore unrelated to the present investigation.

Method

Sample and Setting

Kodaiji Temple is a large Buddhist Temple in Kyoto's Higashiyama District. We selected the Kodaiji Temple as the site to carry out data collection because it is perhaps the most famous example in the world of a robot who regularly preaches religious sermons. Some people visit Kodaiji Temple specifically to see Mindar, but since the temple is already well-established, many people come to the temple with no intent of visiting Mindar. It is worth noting that, unlike in Western countries, Japanese are less likely to consider themselves loyal to any single tradition such as Buddhism or Shinto (Kavanagh & Jong, 2019; Reader, 1991). They may participate in activities associated with both religious traditions or other religious traditions.

Mindar and the human preacher were housed in different adjacent buildings, thus constituting a natural quasi-experimental design where we surveyed participants as they exited either building. Mindar and the human preacher give different sermons, which is a limitation of our study. Mindar's sermon uses consistent language; Mindar consistently delivers a 25-min Heart Sutra while turning its head and torso toward different parts of the room. Mindar's sermon also incorporates audiovisual special effects (see Figure 1; instances of the sermon can be viewed on YouTube at https://www.youtube.com/watch?v=hLoF5_-OUKY). Human preachers give more variable sermons with no special effects.

With permission from the temple, we sampled as many visitors as possible during the 6-week window in which we had permission to collect data. The hall that houses Mindar only opens from Fridays to Sundays, and Mindar delivered the sermon once every hour. However, with permission from the temple, we were able to open

it for visitors during weekdays for the full 6 weeks when we ran this study. In addition, we were able to get Mindar to deliver the sermons every 30 min to ensure guests were exposed to its full effects. We ultimately surveyed a total of 422 participants as they left the temple. Of these, we excluded 24 participants who said that they had visited the temple to specifically see Mindar. This number is somewhat low because we ran this study during the pandemic when Japan is closed to all foreign tourists. This left 398 participants—M_{age} = 45.89, standard deviation (SD_{age}) = 15.10, 228 men, 165 women, one non-binary—in our analyses. Results are substantively identical if we include the full sample, and we report these in the online supplemental materials.

Procedure

Our data collection was carried out by two research assistants (RAs), who stood just outside the two exits where Mindar and the human preacher were housed (we counter-balanced the RAs by day to avoid an experimenter selection effect). RAs consented people as they left the temple, and then handed participants packets which contained our survey measures. The donation paradigm came at the end of the survey. We note that there are also opportunities to donate to Kodaiji throughout the temple, and 64% of our sample indicated that they had already donated during their visit. For this reason, we controlled for prior donation in all our analyses, but it did not affect the results.

It is plausible that knowledge of Mindar changes participants' perceptions of the Kodaiji Temple in ways that are outside the theoretical framework of our study but nevertheless reduce donations. Awareness of Mindar might lead participants to see the temple as wealthy—and less in need of donations—since the temple can afford to construct and regularly use an android preacher. Awareness of Mindar might also lead participants to view the temple's monks as less hard-working—and less deserving of donations—since they have delegated their preaching duties to an android. We evaluated these potential factors with a pilot study which we summarize in the online supplemental materials. In this study, we found no evidence that awareness of Mindar affects perceptions of monk work ethic or wealth of the Kodaiji Temple.

Measures

Religious Commitment

We provided each participant with 1,000 yen (approximately 8 USD) for completing the survey. Participants could give as much of this money as they wished back to the temple, in increments of 100 yen. Donation amounts had a very strong bimodal distribution, with 56% of participants either giving nothing (25%) or the entire amount (31%) to the temple. Figure 1 displays this bimodal distribution. Given this distribution, we created two dichotomous versions of the donation variable. One of these dichotomous variables measured whether (75%) or not (25%) participants had donated anything to the temple. The other dichotomous variable measured whether (31%) or not (69%) participants had donated all their money to the temple.

Credibility

We measured the perceived credibility of both the human and robot preachers using a modified version of Lanman and Buhrmester's (2017) scale. Our items were (a) "The robot

[human] priest acts as a good religious role model"; (b) "The robot [human] priest avoids harming others because religion told them so"; (c) "The robot [human] priest acts fairly to others because religion told them so"; and (d) "The robot [human] priest believes the message of the sermon that they give." Participants used a 1 (strongly disagree) to 5 (strongly agree) scale to rate each item. The credibility scale was reliable when people rated the human preacher (α = .93) and the robot preacher (α = .89), and so we averaged the items into composite indices in which higher values indicated higher perceived credibility.

Religious Beliefs

We measured participants' various religious beliefs as control variables, since Study 1 was not a truly randomized design. These included whether participants believed in god with a dummy-coded variable, whether they believed in Karma with a three-item scale (adapted from White, Norenzayan, & Schaller, 2019), and their views of god as loving or punitive with a 10-item scale (adapted from Johnson et al., 2015). Belief in gods, karma, and a punitive view of god have all been linked with generosity and cooperation with coreligionists (Shariff et al., 2016; Shariff & Norenzayan, 2011; White, Kelly, et al., 2019), and so it was important to control for these religious beliefs.

Karma belief and views of God items were rated using a 1–5 scale from strongly disagree (1) to strongly agree (5). Table S1 in the online supplemental materials contains items from these measures. A reliability analysis showed that the karma items were highly reliable (α = .92), and so we averaged them into a single composite index in which greater numbers indicated higher belief in karma. As in past research, a factor analysis of the views of God scale showed a clear two-factor solution with two factors explaining 70% of the item variance. The first factor contained the five loving items (loadings > .70) and the second factor contained the five punitive items (loadings > .70). We, therefore, averaged the sets of items into two composite scales where higher values, respectively, indicated greater belief in a loving and punitive god. The loving and punitive factors correlated at .24, $p < .001$, indicating that belief in a loving God was not mutually exclusive with belief in a punitive God.

Other Variables

We included other variables in this survey related to moral values, which were adapted from the Moral Foundations Questionnaire (Graham et al., 2009). These items were included to test a different research question, and we did not analyze them here.

Analytic Plan

We tested three main questions in this study. First, did participants find Mindar versus the human preacher more credible? We tested this question using a within-subjects t-test since all participants evaluated the credibility of the robot and of the human preacher. Second, did participants donate more in the robot preacher versus the human preacher conditions? We tested this question with three logistical regressions using binomial estimation in which robot condition was the focal fixed effect. Binomial estimation was appropriate since our donation variable was dichotomized. The first logistical regression included no control variables. The second included

prior donations as an additional fixed effect, and the third included God belief, views of God, and karma beliefs as additional fixed effects. This allowed us to show that the condition effects were not driven by different religious profiles among those who viewed Mindar versus those who viewed the human preacher.

Our final question was: did participants' perceptions of credibility moderate whether they donated more after witnessing the robot versus the human preacher? We tested this question by interacting perceptions of (a) robot credibility and (b) human credibility with robot condition in our logistical regressions, and sequentially probing whether these interactions were significant while modeling the same covariates as in our prior analyses. Throughout our results, we report confidence intervals (CIs) of logistical regressions around the odds ratios (ORs) of each model rather than the estimate. All analyses are conducted in R. All statistical tests are two-tailed.

Results

Prior to testing hypotheses, we conducted two logistic regressions testing whether participants who saw Mindar were likely to have less belief in God or have given less of a prior donation. We found that participants who viewed Mindar (39%) were indeed less likely to believe in God than participants who viewed the human preacher (62%), $b = -0.93$, standard error $(SE) = 0.21$, $OR = 0.40$, $t(387) = -4.44$, $p < .001$, 95% CIs [0.26, 0.59], and participants who viewed Mindar (26%) had given a significantly lower donation prior to the beginning of the study compared to participants who viewed the human preacher (44%), $b = -0.80$, $SE = 0.22$, $OR = 0.45$, $t(387) = -3.68$, $p < .001$, [0.29, 0.69]. This underscores the importance of controlling for these variables when testing our hypotheses.

Robot Versus Human Credibility

Our first analysis tested evaluated whether participants found Mindar versus the human preacher more credible. We found that participants viewed the human preacher ($M = 3.51$) as more credible than the robot preacher ($M = 3.12$), $t(751.98) = 6.51$, $p < .001$, $d = 0.45$.

Robot Condition and Religious Commitment

We next analyzed participants' donations after seeing the robot preacher versus the human preacher. We found that participants in the robot preacher condition were less likely to donate to the temple (68%) than participants in the human preacher condition (80%). This effect remained significant when we controlled for prior donation (Model 2) and our full set of religious belief measures (Model 3). Aside from viewing the robot preacher, the only other significant predictor in this complete model was prior donation, which negatively predicted donation amount. The model showed no evidence of multicollinearity (Variance Inflation Factors < 2.00), which is important because variables such as belief in God and belief in a loving God could plausibly covary strongly. All model coefficients are displayed in Table 1.

We found no significant effect of condition when we examined the alternative dichotomized variable, which measured whether participants donated their entire pot ($p = .33$). Therefore, preacher characteristics appear to affect people's likelihood of donating versus not donating rather than their likelihood of highly generous donations versus moderately generous donations (see Figure 2 for

Table 1

Multiple Regression Models Predicting Donation Likelihood in Study 1

| Predictor	df	b (SE)	OR	t	p	95% CIs
Model 1	393					
Robot preacher		−0.65	0.52	−2.78	.005	[0.33, 0.82]
Model 2	387					
Robot preacher		−0.59	0.56	−2.40	.02	[0.35, 0.90]
Prior donation		−0.65	0.52	−2.64	.008	[0.32, 0.85]
Model 3	369					
Robot preacher		−0.57	0.56	−2.24	.03	[0.34, 0.93]
Prior donation		−0.59	0.55	−2.28	.02	[0.33, 0.92]
God belief		0.35	1.42	1.27	.20	[0.83, 2.46]
God as loving		0.16	1.17	1.09	.28	[0.88, 1.56]
God as punitive		−0.16	0.85	−1.05	.29	[0.63, 1.15]
Karma		−0.03	0.97	−0.22	.83	[0.75, 1.26]

Note. CIs are constructed around the OR estimate, so intervals containing 1 are not statistically significant. SE = standard error; OR = odds ratio; CIs = confidence intervals.

the distribution of donation amounts by condition). We write more about this pattern more in the general discussion of the article.

Our final analysis tested whether perceived credibility of robot and human preachers moderated the relationship between preacher condition and donation. We found that the interaction between condition and human preacher credibility reached significance when we included no control variables (as in Table 1, Model 1), $b = -0.97$, $SE = 0.37$, $OR = 0.38$, $t(383) = -2.65$, $p = .01$, 95% CIs [0.18, 0.76], when we controlled for prior donations (as in Table 1, Model 2), $b = -0.87$, $SE = 0.37$, $OR = 0.42$, $t(378) = -2.35$, $p = .02$, [0.20, 0.85], and when we additionally controlled for God belief, belief in God as loving, belief in God as punitive, and belief

Figure 2

A Histogram of Donation Amounts, Organized by 10 Different Levels of Donation
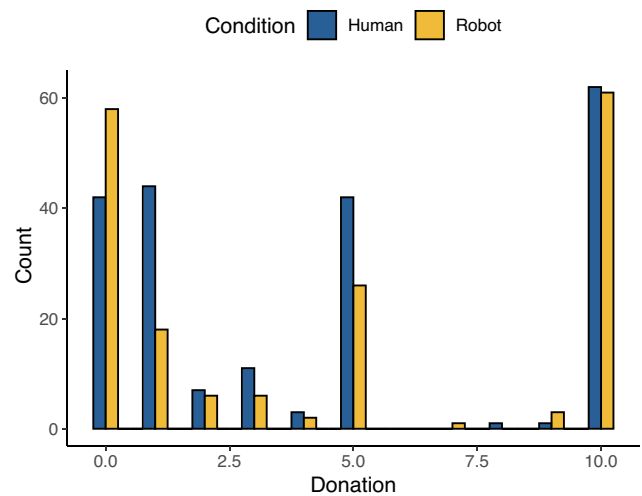
Note. A score of "10" indicates that participants donated their full 1,000 yen. Dark/blue indicates "Human Preacher" condition. Light/goldenrod indicates "Robot Preacher" condition. See the online article for the color version of this figure.

in God as loving (as in Table 1, Model 3), $b = -0.90$, $SE = 0.38$, $OR = 0.41$, $t(377) = -2.39$, $p = .02$, [0.19, 0.84].

Simple slopes, derived from our third model with all controls, showed that there was a strong negative effect of viewing the robot preacher on donation rates among people who viewed human preachers as relatively high (+1 SD) in credibility, $OR = 0.23$, 95% CIs [0.09, 0.53], whereas the effect of viewing the robot preacher on donation rates did not reach significance for people who viewed human preachers as relatively low (−1 SD) in credibility, $OR = 0.95$, [0.45, 2.01]. Figure 3 illustrates this moderation.

In these same models, we also probed for possible moderation by perceived robot credibility. However, there were no significant interactions between condition and perceived robot credibility in any of the models ($ps > .60$).

Discussion

We found evidence in a unique field study that robot preachers are perceived as less credible than human preachers, and that witnessing a robot preacher predicts less religious commitment (donations to a temple) than witnessing a human preacher. Study 1 had high ecological validity and unlike most studies of religious commitment, it featured a non-Western sample (Henrich et al., 2010). However, Study 1 was not a pure experiment with random assignment. To address these limitations, we preregistered a second study which used a fully experimental design in a Taoist temple.

Study 2: Manipulating Preacher Characteristics in a Taoist Temple

Study 2 was a preregistered experiment which we conducted in a Taoist temple in Singapore. All participants who visited this temple during our study period witnessed a short sermon, but we manipulated whether the sermon was delivered by a human or robot preacher. Unlike Study 1, participants were randomly assigned to engage with the human versus robot preacher, and the content of the sermon was identical across conditions. We hypothesized that participants exposed to a robot preacher would show less religious commitment compared to participants exposed to a human preacher, and this effect would be mediated by credibility.

Study 2 also featured improved measurements of credibility and religious commitment. For example, the Study 2 design allowed us to measure credibility in a way which was more consistent with our theoretical

Figure 3
Participants' Likelihood of Donating Based on Their Study Condition and Their Perceptions of Human Preacher Credibility

Note. See the online article for the color version of this figure.

model. In Study 1, Mindar and the human preacher were housed in different buildings, but participants were aware of both preachers and their role in the temple. As such, we chose to ask all participants about the credibility of Mindar and the human preacher, and to use these ratings as moderators. In Study 2, however, the temple was occupied by either the robot or human preacher, so we asked participants to rate the credibility of the preacher with whom they had interacted. We could then use these ratings of credibility as mediators, which was more consistent with our theoretical model in which exposure to a robot preacher may undermine religious commitment *through* lower perceived credibility.

We also measured credibility more broadly in Study 2 than in Study 1. In Study 1, we only measured the credibility of the preacher. However, participants could also see the temple as a whole as less credible when they use a robot to deliver sermons. This may especially be the case for participants who do not see robots as agents, and simply see them as instruments of the temple, similar to choosing to play a sermon from a laptop. These participants could still show less religious commitment after exposure to a robot preacher, and credibility would play an important role in their lower commitment, but it would be the credibility of the temple rather than the preacher. Separately measuring credibility of the robot and the temple therefore allowed us to more fully capture how exposure to a robot preacher may reduce perceived credibility, and in turn, religious commitment.

Our measurement of religious commitment in Study 2 was more thorough than in Study 1. In addition to measuring donations to the temple—our focal dependent variable—we also measured participants' willingness to circulate flyers advertising the temple and self-reported likelihood of spreading the message of the sermon as measures in the online supplemental materials. This allowed us to measure religious commitment with some convergent validity, rather than relying solely on donations to the temple as a measure of commitment.

Method

Sample and Setting

We conducted this study in a Taoist temple in Singapore. Like most Taoist temples, this temple has a Taoist master (priest) who regularly delivers sermons to adherents (Dean, 2014). We chose this site because one of the authors of this article was able to gain permission for us to conduct an experimental field study. We first conducted briefing sessions with all the Taoist disciples working on the site during which we explained the details and procedures of the current research without disclosing the hypotheses.

As with Study 1, we recruited as many participants as we could within a 2-week time window in which we had permission to collect data. We hoped to collect data on 400 participants, but we were only able to gather a sample of 239 ($M_{age} = 39.80$, $SD_{age} = 8.17$, 128 men, 111 women). This lower-than-expected sample size was not due to a low response rate, but rather relatively low traffic to the temple because of COVID-19 social distancing restrictions imposed by the local government (only five people were allowed in the temple at a time). These participants were Taoist temple visitors whom we recruited as they were leaving the temple. We compensated each visitor with 5 SGD (i.e., approximately 3.7 USD), and over 90% of participants who we approached agreed to do the study and nobody dropped out after starting the study. This is important to note because a high attrition or agreement rate could lead to a failure of randomization if one condition had a higher attrition rate than the other.
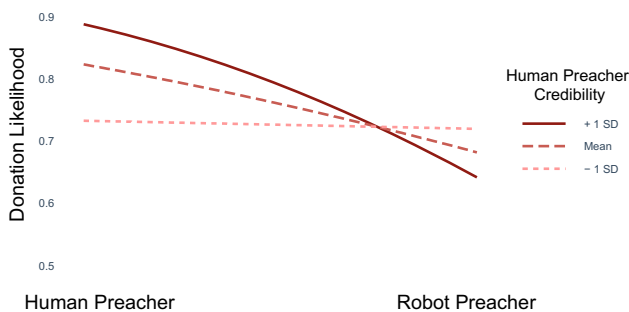
Manipulation

We arranged the temple so that some days featured the robot preacher and other days featured the human preacher. The days were randomly assigned and did not vary in any meaningful way (e.g., weekend vs. weekday) and there was a similar number of participants in the robotic preacher condition ($n = 118$) and the human preacher condition ($n = 120$). In both conditions, the preacher delivered an identical sermon using content from Chapters 1 and 8 from the Tao Te Ching, which is the Bible of Taoism. These chapters are particularly important within the Tao Te Ching because they represent the most important Taoist principles in the book (LaFargue, 1994). The text for both chapters is presented in the online supplemental materials.

Measures

Credibility

We used the same credibility measure as in Study 1. However, we expanded the measure to focus on two sources of credibility: The credibility of the preacher, and the credibility of the temple. The temple-focused items were modified slightly to read (a) The alter provides a good environment to practice faith, (b) the alter embodies the values of Taoism, and (c) the alter believes the messages of Taoism. The credibility scale was reliable when people rated the preacher ($\alpha = .96$) and the temple ($\alpha = .70$), and so we averaged the items into composite indices in which higher values indicated higher perceived credibility.

Religious Commitment

As in Study 1, we measured religious commitment using donations to the temple. Participants could choose to give away any amount of the money they earned for completing the study, in six single-dollar increments ranging from 0 to 5 SGD. Unlike Study 1, donation amounts did not show a large bimodal distribution (see Figure 4), so we analyzed donations as a continuous variable.

We also included two measures in the online supplemental materials of religious commitment in this study (we note that we did not preregister analyses on these measures). First, we asked participants if they would be willing to take and distribute flyers related to the Taoist temple throughout Singapore, with participants' responses dummy coded as 0 (did not wish to distribute flyers) and 1 (agreed to distribute flyers). Second, we asked participants how much they planned to circulate the message of the sermon across three items: (a) I was moved to pass on the sermon to people I know, (b) I intend to teach the message of the sermon to my children or other family members, and (c) I will encourage more people to experience today's sermon. Due to a RA error, participants responded to these items using a scale from 5 (strongly agree) to 1 (strongly disagree)—in other words, the higher numbers appeared on the left instead of the right side of the scale and were reverse coded in our raw data. For this reason, we transformed the variables prior to analysis so that higher values represented stronger agreement. The scale appeared to be reliable ($\alpha = .92$), and so we averaged the items into a composite index in which higher values indicated more willingness to share the message of the sermon.

All forms of religious commitment were correlated significantly but weakly. Participants who donated more to the temple were also more likely to agree to handing out flyers, $r = .16$, $p = .02$, and circulating the message of the sermon, $r = .13$, $p = .04$, and participants who agreed to hand out flyers were also more likely to circulate the message of the sermon, $r = .26$, $p < .001$. The positive correlations between willingness to circulating the message of the

Figure 4
Main Effects of Preacher Condition in Study 2
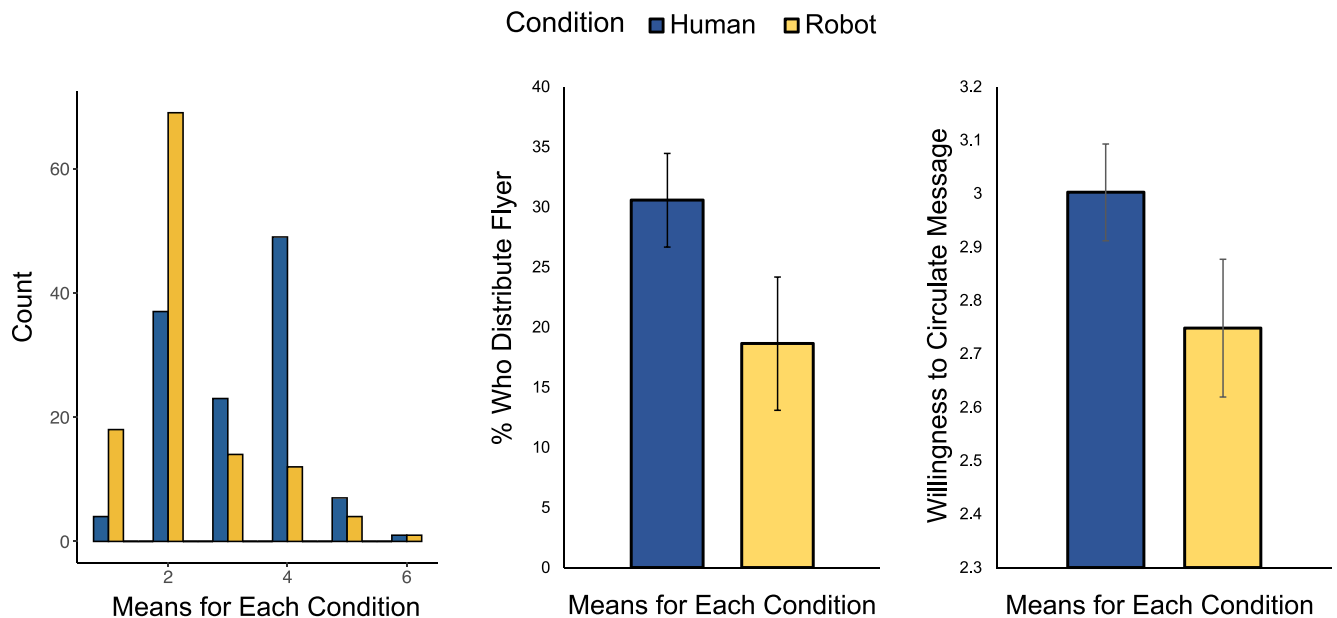
Note. Left: A histogram of donation amounts, organized by six different levels of donation. A score of "6" indicates that participants donated their full 5 SGD. Center: Percentages of people willing to distribute flyers by condition. Right: Mean willingness to circulate sermon message by condition. Dark/blue indicates "Human Preacher" condition. Light/goldenrod indicates "Robot Preacher" condition. The error bars represent standard errors. See the online article for the color version of this figure.

sermon and the other religious commitment measures suggest that participants had interpreted the willingness to circulate scale appropriately and that our decision to reverse-score the scale values was justified.

Perceptions of Robots' Minds

We included eight items measuring participants' general tendency to perceive mind in robots. We included this measure as an exploratory moderator testing whether participants' perceptions of credibility across the robot versus human preacher conditions would depend on whether people perceived robots as having mind. Four of these items were designed to measure robots' experience ("Robots can feel pain"; "Robots can feel fear"; "Robots can have desires"; "Robots can be happy"). The remaining four items were designed to measure robots' agency ("Robots can communicate with others"; "Robots can think"; "Robots can plan their actions"; "Robots can remember things"). We write more about this measure—and present the results of the measure—in the online supplemental materials. We also conducted a follow-up moderation in Study 3 using a measure of anthropomorphism which we also provide in the online supplemental materials.

Analytic Plan

We used a similar analytic plan as in Study 1. First, we tested whether credibility ratings of the preacher and temple varied significantly across the human and robot preacher conditions. We conducted a linear regression with Gaussian estimation for this analysis, with robot preacher condition entered as a dummy-coded fixed effect. Second, we tested whether religious commitment varied across conditions using generalized linear models with Poisson estimation (for the donation variable, which was a count model), binomial estimation (for the flyer variable) and Gaussian estimation (for the message circulation variable). Finally, we fit a 5,000-sample bootstrapped structural equation model (SEM) to evaluate whether credibility mediated the relationship between condition and each measure of religious commitment, controlling for covariation between these commitment measures. We preregistered the 5,000-sample bootstrap format of the SEM and the prediction that perceptions of credibility would mediate the impact of condition on religious commitment, but we did not specify an exact constellation of our model because we did not have strong a priori predictions about how preacher and temple credibility would relate to one another. In the main text, we present a well-supported model where temple credibility and preacher credibility are independent mediators. In the online supplemental materials, we present an alternative serial mediational model in which perceptions of preacher credibility are linked to lower commitment through perceptions of temple credibility. We ultimately did not find support for mediation in this serial model.

Results

Robot Versus Human Credibility

Our credibility findings largely mirrored those from Study 1. Participants rated the preacher as having lower CREDs in the robot compared to the human condition, $b = -1.15$, $\beta = -.57$, $SE = 0.11$, $t(237) = -10.58$, $p < .001$, 95% CIs [-1.37, -0.94].

Participants also rated the temple as having lower CREDs in the robot condition compared to the human condition, $b = -0.80$, $\beta = -.41$, $SE = 0.12$, $t(237) = -6.82$, $p < .001$, [-1.03, -0.57].

Robot Condition and Religious Commitment

Participants in the robot preacher condition donated less, $b = -0.32$, $SE = 0.08$, $t(238) = -4.03$, $p < .001$, 95% CIs [-0.48, -0.17], agreed to distribute flyers less often, $b = -0.65$, $OR = 0.73$, $SE = 0.31$, $t(238) = -2.12$, $p = .03$, [0.62, 0.85], and were less likely to say that they would circulate the message of the sermon, $b = -0.25$, $\beta = -.13$, $SE = 0.13$, $t(237) = -1.97$, $p = .049$, [-0.50, -0.00001] (see Figure 4). The latter two effect sizes were small, but each effect was consistent with our hypotheses. The effect on message circulation was particularly interesting because all participants heard the exact same sermons in Study 2. This resembled our Study 1 finding that robot preacher exposure was more likely to lead to any kind of donation but did not increase the likelihood that people donated their entire pot of money.

In sum, exposure to a robot preacher decreased participants' perceptions of the preacher's as well as the temple's credibility. Exposure to a robot preacher also decreased religious commitment, resulting in fewer donations to the temple and less willingness to spread the message of the sermon. Our final analysis was an SEM which estimated whether credibility ratings mediated the relationship between preacher condition and religious commitment, controlling for covariation between the three forms of religious commitment.

The full SEM with effect sizes is displayed in Figure 5, and all coefficients are provided in the online supplemental materials. Our model found significant mediation of condition on religious commitment through both preacher and temple credibility. Temple credibility mediated the effect of condition on donation amounts, 95% CIs [-1.05, -0.62], flyer distribution [-0.17, -0.04], and willingness to share the sermon message [-0.41, -0.07]. Preacher credibility mediated the effect of condition on flyer distribution [-1.13, -0.02], and willingness to share the sermon message [-0.50, -0.19], but not on donation [-0.02, 0.18]. This may be because donation was directly awarded to the temple, so participants were especially tentative to the credibility of the temple when they made this judgment.

We note that there were also significant direct effects of condition on two of our outcome variables. After residualizing on credibility, exposure to a robot preacher still predicted less donation to the temple, suggesting that there may be an unmodeled mechanism by which robot preacher exposure reduces religious donation. There was also a suppression effect of condition on willingness to share the sermon's message, such that exposure to the robot preacher *increased* willingness after residualizing on credibility. This may be due to the novelty effect of seeing a robot deliver a sermon, but we did not measure perceived novelty, so this is a speculative explanation. These are interesting effects, but they are less theoretically relevant than our hypothesized model. We also note that we tested a serial mediation model in which condition and preacher credibility influenced religious commitment via temple credibility, but we did not find evidence for significant mediation in this model. See online supplemental materials for more information about the results of this model.

Figure 5
Regression Estimates From Our Study 2 SEM
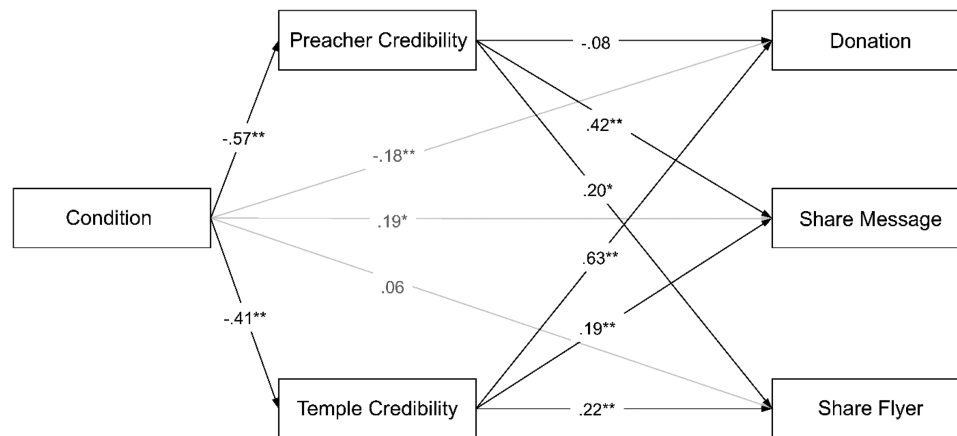
Note. All estimates are standardized. Single-starred effects represent significance at the *p* < .05 level. Double-starred effects represent significance at the *p* < .005 level. Gray paths were estimated but not included in the indirect effect estimation. Condition is dummy-coded such that "1" represents participants who viewed the robot preacher and "0" represents participants who viewed the human preacher. A full set of coefficients are displayed in the online supplemental materials. SEM = structural equation model.

Discussion

We found that adherents randomly assigned to a robot—rather than human—preacher showed decreased religious commitment. In addition to decreasing donation to the temple, participants who witnessed a robot preacher were also less willing to spread the message of the sermon that they heard, and to distribute flyers for their temple. These effects on religious commitment were mediated by perceptions of credibility, both with respect to the temple and the preacher. Study 2 built on Study 1 in several important ways. For example, Study 2 replicated our findings using a fully randomized experimental design. Study 2 also used a broader set of measures of religious commitment and showed that the relationship between exposure to a robot preacher and religious commitment is jointly mediated by the preacher's perceived credibility but also the credibility of the religious institution that employs the preacher.

However, several key limitations remained. Most notably, we did not measure other key factors such as mind perception and likability. Mind perception is a plausible mechanism for why robot preachers are perceived as less credible than human preachers. Having a mind, with the capacity to think and feel, is likely an important prerequisite to being a credible religious elite. Charisma and likability are potential confounding variables. Apart from appearing less credible, robots may also appear less charismatic and likable than human preachers. Since charismatic religious elites inspire greater trust and devotion (Schjoedt et al., 2011; Sperber, 2010), they could plausibly explain why robot preachers inspired less commitment in Studies 1–2 compared to human preachers. In Study 3, we measured these additional factors to rule them out as possible confounds to credibility.

Study 3

Study 3 was a preregistered online between-subjects experiment in which religious participants were told that a sermon was either generated by a highly advanced AI program or by a human preacher.

Participants then rated their perceptions of the sermon author's credibility, mental properties, likability, charisma, and estimated their religious commitment if they were to hear the sermon at their place of worship. While Study 3 did not have the same ecological validity as our field studies, it allowed us to survey a novel population (Christians), counterbalance our measures within an electronic survey, and test whether our effects would generalize beyond embodied robots who read sermons to disembodied AI programs who can generate sermons.

We tested whether participants would anticipate less religious commitment after witnessing an AI-generated sermon versus a human-generated sermon, and this effect would be mediated by perceptions of AI credibility, which would in turn be mediated by mind perception properties. However, we did not have strong a priori predictions about the exact nature of this mediation, and we explored various different model constellations, which means that the SEM models that we present here were not preregistered. In Study 3, we separately measured religious commitment to participants' specific religious institution and to their Christian identity more generally. We did not have strong a priori hypotheses about whether effects would markedly differ across these referents.

Method

Participants

We recruited 300 participants from Amazon Mechanical Turk using the CloudResearch platform. Using the CloudResearch filters, we specifically sought to recruit Christian individuals, but we nevertheless found that 23 participants identified as Atheist, Agnostic, or "none" as their religious affiliation, so we excluded these participants from analyses. We also found that three participants failed our manipulation check (selecting "gardening" from a list of hobbies at the experiment's request), which left a total of 274 participants (130 men, 144 women; $M_{age} = 43.80$, $SD_{age} = 12.64$) in our sample for analysis.

Manipulation

After consenting to participate in the study, all participants read that "in this survey, we are interested in how people evaluate new technologies which are changing religion." Participants in the human sermon condition then read "Many people are now reading sermons on computers rather than listening to them in churches or reading them in books…Below we have pasted an example of a real sermon which was written by a preacher and published online. Please take your time and read the sermon." Participants in the AI sermon condition instead read "New innovations in artificial intelligence mean that computer programmers can produce sermons which have been generated entirely by machines … Below we have copied an example of one of these 'AI sermons.' Please take your time and read the sermon." The content of the sermon was the same in both conditions (see online supplemental materials). It was written by the first author of this article so that participants could not discover the sermon and its true origins on their browser during the study (we conducted this study before the release of Chat GPT, which can spontaneously write realistic sermons).

After reading the sermon, participants in the AI sermon condition responded to the question, "Who do you think is more responsible for writing this sermon: The AI program that generated the sermon or the human who trained the AI program?" Participants responded to the question using a 1–10 scale anchored at 1 (*The AI Program*) and 10 (*The Human Trainer*). This question was intended to assess whether humans could view an artificial agent as capable of being responsible for a sermon, since we did not measure perceived responsibility in Studies 1–2. We found that the mean response was 5.82, and a one-sample t-test suggested that this mean was not significantly different than the 5.50 mid-point of the scale, $t(131) = 0.32$, $p = .25$, suggesting that the human trainer was not deemed significantly more or less responsible than the AI program for generating the sermon.

Measures

Religious Commitment

To contextualize the religious commitment ratings, participants were instructed "We would like you to think about how you would feel if this sermon was read while you were attending your place of worship. Please respond to each statement using the scale that we have provided." We then gave participants six statements to rate on a 1 (*strongly disagree*) to 10 (*strongly agree*) scale. Three of the items were intended to measure religious commitment to participants' specific place of worship ("I would feel committed to my church"; "I would consider donating money to my church"; "I would consider telling strangers to join my place of worship"). The other three items were intended to measure generalized religious commitment ("I would feel committed to my Christian identity"; "I would consider donating to a Christian organization"; "I would consider telling people about Christianity").

A factor analysis found evidence for a single-factor solution, with a single factor explaining 82% of variance and no other factors exceeding an eigenvalue of 0.29. We, therefore, averaged the specific and generalized items into a single composite index in which higher values indicated greater religious commitment ($\alpha = .97$). We also note results separately for the specific and generalized indices for the sake of transparency.

Credibility

We measured credibility using a three-item scale which was very similar to the scale we used in Studies 1–2: (a) The person [AI] who wrote this sermon is probably a good religious role model, (b) The person [AI] who wrote this sermon embodies religious values, and (c) The AI who wrote this sermon believes the message of the sermon. We adapted slightly so that it conveyed the same meaning as the original scale but had less anthropomorphic language. The scale was highly reliable ($\alpha = .95$), so we averaged the items into a composite index where higher values indicated more credibility.

Mind Perception

Mind perception is traditionally measured along two dimensions of agency (capable of thinking and acting) and patiency (capable of feeling) (H. M. Gray et al., 2007). However, these two dimensions can sometimes collapse into a single dimension of "more mind" versus "less mind" (Waytz et al., 2010). We developed separate 1 (*strongly disagree*)–10 (*strongly agree*) items in this study to measure agency ("The person [AI] who wrote this sermon is probably capable of thinking and planning") and patiency ("The person [AI] who wrote this sermon is probably capable of hunger and thirst"). We created one item for agency and patiency because the agency and patiency items correlated highly with one another in our previous studies, and because we had already increased the number of questions in Study 3 (including items about likability and charisma, religious commitment items, and asking all items twice—once for the author of the sermon and again for the trainer of the sermon's author) and wanted to manage participant fatigue.

We found that the agency and patiency items correlated highly with each other ($r = .87$), so we averaged them into a single scale in which higher numbers indicated greater perceived mind. These items likely correlated highly because participants were using them to reflect the belief that robots generally have less mind than humans. In the online supplemental materials, we show that our findings are highly similar when we use each of the individual items.

Likability and Charisma

We measured likability using the item "The person [AI] who wrote this sermon is probably likable" and we measured charisma using the item "The person [AI] who wrote this sermon is probably charismatic." Both items were rated using a 1 (*strongly disagree*) to 10 (*strongly agree*) scale. The two items correlated at $r = .94$. Here we present results using the "likability" subscale, because charisma can mean different things to Christian participants (e.g., either having a charismatic personality or belonging to a doctrine in which charismatic displays such as speaking in tongues is normative). In the online supplemental materials, we show that our findings are virtually identical when we average together charisma and likability into a composite scale, or replace likability with charisma.

Perceptions of Trainer

As exploratory measures, we also measured mind perception and credibility of the person who trained the author of the sermon using the same items that we used for the sermon's author. In the human sermon condition, this was the mentor of the preacher who wrote the sermon. In the AI sermon condition, it was the programmer

who developed the sermon-generating AI. In the online supplemental materials, we provide more theoretical background for these measures, and we also show that people perceived the human trainers' minds and credibility similarly across the study conditions.

Anthropomorphism

We also included a measure of AI anthropomorphism, previously developed by Bartneck et al. (2009) in which participants rated the extent to which they believed AI could be human-like along five key dimensions using a seven-point scale: *fake* (1) versus *natural* (5), *machinelike* (1) versus *humanlike* (5), *unconscious* (1) versus *conscious* (5), and *artificial* (1) versus *lifelike* (5). We used this measure of anthropomorphism as an exploratory moderator (meaning that we did not have a preregistered predictions concerning this measure), and we present more reasoning behind its inclusion—and results—in the online supplemental materials.

Procedure

Participants began the study by viewing our manipulation. They then completed the religious commitment measures and rated the sermon author's attributes (credibility, mind perception, likability, and charisma) in counterbalanced order. We counterbalanced the measures to ensure that rating credibility prior to religious commitment was not introducing any researcher demand.

Analytic Plan

We followed a similar analytic strategy to Study 2. First, we tested whether the AI program was perceived as having less mind, less credibility, and less likability than the human preacher. We conducted this test using three linear regressions with Gaussian estimation in which AI sermon condition was entered as a fixed effect with no other covariates (we anticipated that participants in the AI sermon condition would ascribe less credibility than participants in the human sermon condition). Second, we tested whether people reported less religious commitment in the AI sermon condition than the human sermon condition (we anticipated that participants in the AI sermon condition would report lower religious commitment than participants in the human sermon condition). Again, we conducted this test using three linear regressions with Gaussian estimation, where condition was a dummy-coded fixed effect and there were no other covariates. We note that our preregistration specified t-tests for these analyses. Our linear regression models include these preregistered t-tests while also accommodating control variables and other statistics (e.g., standardized betas). t-test results are also displayed in Table 2.

Table 2
Mean and SD *of Key Study 3 Variables Across Conditions*

Variable	AI sermon Mean (*SD*)	Human sermon Mean (*SD*)	Difference
Religious commitment	5.41 (2.60)	6.41 (2.52)	$t = -3.25, p < .001$
Mind perception	3.28 (2.71)	8.57 (1.88)	$t = -18.89, p < .001$
Credibility	4.40 (2.88)	7.55 (1.89)	$t = -10.79, p < .001$
Likability	4.13 (2.96)	7.65 (2.07)	$t = -11.45, p < .001$

Note. AI = artificial intelligence; *SD* = standard deviation.

Finally, we fit a 5,000-sample bootstrapped SEM to evaluate possible causal pathways between AI sermon condition and religious commitment. As in Study 2, we did not preregister a fixed structure to this SEM. We only preregistered that this model would include credibility, likability/charisma, and mind perception. Our exact language in the preregistration is that we would use "Path model to test whether CREDs, mind perception, likability, or charisma can better explain the relationship between condition and religious commitment." For the sake of transparency, we present two plausible models in this main text—one model in which likability is entered as a control variable and another model in which likability is modeled as a separate mediator along with credibility.

Results

Table 2 contains means of credibility, mind perception, and likability in both the robot and human preacher conditions.

Robot Versus Human Credibility

As with Studies 1–2, AI authors of the sermon were perceived as less credible than human authors, $b = -3.15$, $\beta = -.54$, $SE = 0.29$, $t(272) = -10.79$, $p < .001$, 95% CIs $[-3.73, -2.58]$. AI authors were also perceived as having less mind than human authors, $b = -5.29$, $\beta = -.75$, $SE = 0.28$, $t(272) = -18.89$, $p < .001$, $[-5.85, -4.74]$, and as being less likable than human authors, $b = -3.52$, $\beta = -.57$, $SE = 0.31$, $t(272) = -11.45$, $p < .001$, $[-4.12, -2.91]$.

Robot Condition and Religious Commitment

Participants reported less religious commitment in the AI sermon condition ($M = 5.41$) versus the human sermon condition ($M = 6.41$), $b = -1.01$, $\beta = -.19$, $SE = 0.31$, $t(272) = -3.25$, $p = .001$, 95% CIs $[-1.61, -0.40]$. This was true for both the items that concerned participants' specific religious institutions, $b = -0.98$, $\beta = -.19$, $SE = 0.31$, $t(272) = -3.15$, $p = .002$, $[-1.59, -0.37]$, and the items that targeted their religious commitment more broadly, $b = -1.03$, $\beta = -.19$, $SE = 0.32$, $t(272) = -3.22$, $p = .001$, $[-1.66, -0.40]$. The effect of condition on religious commitment was stronger when participants rated religious commitment before author credibility ($b = -1.46$) than when they rated credibility before commitment ($b = -0.44$). This difference was not statistically significant, $b = -1.02$, $\beta = -.17$, $SE = 0.62$, $t(270) = -1.64$, $p = .10$, $[-2.24, 0.20]$, but it does help rule out the possibility that rating credibility before commitment amplifies the effect of condition on religious commitment because of perceived experimenter demand.

Our final analysis was a SEM in which we estimated the serial mediation from condition → author mind perception → author credibility → participant religious commitment. We fit this model in two different ways. Our first formulation entered author likability as a covariate in each of the paths. In this model, the indirect effect of robot preacher exposure on participant religious commitment through author mind perception and author credibility was statistically significant, $b = -0.86$, 95% CIs $[-1.36, -0.45]$.

In the second model, displayed in Figure 6, we included author likability as a mediator in the online supplemental materials alongside author credibility to test whether author likability and author credibility were redundant or whether they could be distinct mechanisms for why religious commitment was lower in the AI sermon condition. In support of the second possibility, we found that the

Figure 6
Regression Estimates From the Study 3 SEM
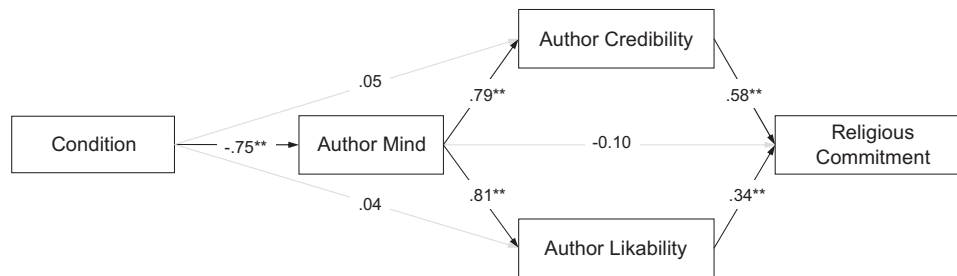
Note. All estimates are standardized. Single-starred effects represent significance at the $p < .05$ level. Double-starred effects represent significance at the $p < .005$ level. Gray paths were estimated but not included in the indirect effect estimation (one gray path, the direct effect of condition on religious commitment, is not displayed for visual purposes but is .24**). Condition is dummy-coded such that "1" represents participants in the AI sermon condition and "0" represents participants in the human sermon condition. A full set of coefficients are displayed in the online supplemental materials. SEM = structural equation model.

indirect effects of credibility, $b = -1.72$, 95% CIs $[-2.31, -1.17]$ and likability, $b = -1.04$, $[-1.63, -0.48]$, were both statistically significant. This model, whose individual paths are displayed in Figure 6, suggested the participants reported lower religious commitment in the AI versus human sermon condition because they perceived the AI as having less mind, which facilitated perceptions of less likable and less credibility.

Examining the direct effects of this model suggested full mediation of several key effects. For example, the direct effect of author mind perception on religious commitment was not statistically significant, $b = -0.07$, 95% CIs $[-0.20, 0.05]$, suggesting that the path was fully mediated by author credibility and author likability. Likewise, there was no significant direct effect of AI sermon condition on author credibility, $b = 0.26$, $[-0.29, 0.78]$, nor on author likability, $b = 0.23$, $[-0.41, 0.83]$, after controlling for author mind perception, suggesting that mind perception explains why AI sermon-writers are perceived as less credible and less likable than human preachers. Interestingly, the direct effect of AI sermon condition on religious commitment was *positive* after removing the variance associated with author mind perception, credibility, and likability, $b = 1.20$, 95% CIs $[0.73, 1.64]$. As with Study 2, this may be because of a novelty effect associated with AI programs or robots serving as religious elites. Nevertheless, this positive direct effect was overwhelmed by the strong negative indirect effects through credibility and likability. This explains why the total effect of condition on commitment was negative in our linear regression.

In the online supplemental materials, we present a version of this SEM that does not include mind perception as the first mediator. This model closely resembles the SEM from Study 2, where mind perception was not a mediator, and it replicates the findings from Study 2.

Discussion

Study 3 replicated our key effects in Studies 1–2 with an online experimental paradigm in which participants read a sermon either generated by an AI program or a human. We found that participants anticipated feeling less religious commitment after hearing the AI-generated sermon versus the human-generated sermon and perceived the AI program to have less credibility than the human preacher, which mediated the effect of condition on religious commitment.

One of the strengths of Study 3 was that we were able to model a more complex set of mechanisms for why robot preachers engender less religious commitment than their human counterparts. For example, we found that robots are perceived to have lower credibility because they are ascribed less mind (agency and patiency) than human preachers. We also found that perceptions of deficient mind perception among robots also leads to less perceived likability, which further erodes religious commitment following robot preacher exposure.

The major limitation of Study 3 was that we measured subjective and hypothetical religious commitment rather than immediate behavioral measures of religious commitment. In other words, participants *imagined* their level of religious commitment if they were to hear our sermon at their place of worship, which may not correspond to their actual level of religious commitment if they heard an AI-generated sermon at a real service. In this sense, Studies 1–3 have complementary strengths and limitations since Studies 1–2 had higher ecological validity with behavioral measures of religious commitment.

General Discussion

Predicting where humans will and will not be replaced by machines is a popular pastime. Much of this prognostication has focused on what skills lie within or beyond the capabilities of emerging technology (Frey & Osborne, 2017). Here we focused on a different domain in which agents' humanity may matter: serving as credible cultural models. By comparing exposure to human versus robot preachers, we found that exposure to a robot preacher led to less religious commitment. Exposure to robot preachers may reduce religious commitment at least partly because they are perceived as less credible than human preachers. In Study 1, the effect was largest for people who perceived human preachers as highly credible, and in Studies 2–3, credibility mediated the effect of robot preacher exposure on religious commitment. In Study 2, we found evidence that people also see religious institutions as less credible when they employ robot versus human preachers—an effect which further contributes to lower religious commitment. And in Study 3, we found that robot preachers are seen as less credible than human preachers

because they are perceived to have less mind—in other words, less capacity to think and feel. The online supplemental materials provide checks that further support our conclusions and show descriptive statistics for key variables.

Our findings show a unique domain where robots may not easily displace humans. The results also support cultural evolutionary theories of religion. According to these theories, people remain committed to their religious beliefs at least partly because of religious elites who maintain credibility by engaging in displays that would be extremely costly if they did not truly hold their beliefs (Henrich, 2009). By replacing these religious elites with robots, who cannot truly hold beliefs or perceive costs, religious institutions may risk declining commitment among their congregants. This effect seems to be compounded by the fact that robot (vs. human) preachers are also less likable and charismatic—additional factors that are important for religious commitment (Schjoedt et al., 2011; Sperber, 2010). This research also adds to a growing scientific study of source credibility, which explains why scientists are widely trusted across cultures whereas spiritual gurus are more trusted by religious versus nonreligious individuals (Hoogeveen et al., 2022; van der Miesen et al., 2022).

An interesting nuance of our first two studies was that preacher characteristics seemed more likely to affect moderately committed individuals than highly committed individuals. In Study 1, individuals who witnessed a human preacher were more likely to donate something rather than nothing, but no more likely to donate their entire pot of money. Similarly, we found in Study 2 that exposure to a robot preacher appeared most likely to increase donations from 0–1 SGD to 2–3 SGD, while people in both conditions were equally likely to give 4–5 SGD (see Figure 4). This pattern is a puzzle to us. One possibility is that *witnessing* displays by religious elites may increase commitment among moderate adherents, whereas *participating* in religious displays is more effective for increasing commitment for devout adherents (see Xygalatas et al., 2013, for research on religious participation and devout adherents). However, this is very much a speculation that we encourage future research to explore empirically.

Strengths, Limitations, and Future Directions

Our studies have several unique strengths. First, we show convergent approaches across three different paradigms which balance high internal and high external validity: a field study involving a real-life robot preacher, a randomized experimental paradigm in the field, and an online experiment. Second, whereas past studies have supported the role of credibility in religious commitment using correlational evidence (Gervais et al., 2021; Willard & Cingl, 2017), the current work provides causal experimental evidence. Third, whereas past research has focused on credibility associated with single individuals, Study 2 showed that the credibility of religious elites and of religious institutions jointly predict religious commitment, which broadens signaling models of the cultural evolution of religion. Finally, whereas past research on credibility has focused on Americans and Europeans, we were able to test our hypotheses in both Western (Study 3) and non-Western, non-Christian samples (Studies 1–2).

A limitation of our work is that we do not explore conditions under which robots could be made more credible or likable as religious figures. In our view, such conditions may not exist. So long

as robots are perceived to have less mind than humans, there may be hard barriers to their perceived credibility. In analyses in the online supplemental materials, we found that people who tend to anthropomorphize robots and AI do perceive automated preachers as more credible than people who view robots and AI as less human (see Studies 2 and 3 in the online supplemental materials). But even high-anthropomorphizing participants viewed robots as significantly less credible than humans.

One possibility is that making robots appear more human may increase their perceived credibility as religious elites. However, this is no easy challenge considering the difficulty of manufacturing an authentically human android (Roese & Amir, 2009). Moreover, the appearance of humanness can trigger an uncanny valley response which could create a further aversion to robot preachers (K. Gray & Wegner, 2012). Another interesting avenue could explore whether agency (the ability to think and do) or patiency (the ability to feel) is more critical to why robots are perceived to have low credibility. In our AI sermon paradigm, the two dimensions correlated extremely highly, but other studies have found that embodied robots are ascribed more agency than patiency (H. M. Gray et al., 2007). In cases where agency and patiency come apart, it would be interesting to test which dimension of mind perception is more crucial to being a credible religious figure.

We also see other future directions in the study of robot preachers, many which explore different forms of religiosity and religious commitment. For example, research could explore whether people are less likely to internalize the moral values which are preached by robot priests, and whether they view rituals as less successful when they involve automated participants. Future studies could also explore whether robot preachers affect cohesion within religious communities by failing to provide empathic and credible leadership. Shared moral values (Graham & Haidt, 2010), cohesion with coreligionists (Whitehouse et al., 2014), and perceived ritual efficacy (Lawson & McCauley, 1993) are all important components to religious commitment that we did not examine in these studies, but which could be jeopardized by rising exposure to robot preachers.

Constraints on Generality

These findings are intended to be relevant for academics, religious practitioners, and members of religious communities. We present data collected from naturalistic field studies as well as controlled laboratory studies so that our findings have a combination of ecological and internal validity. We have provided support for our hypotheses with data from Japan, Singapore, and the United States. We encourage more work to test whether our findings are also valid in other cultural contexts.

Implications and Conclusion

The current findings have implications for the future of both automation and religion. For automation, our research reveals how recent insights from psychological research on social learning and cultural transmission can help predict which occupations can be successfully automated, and which need remain human. Domains like religion, which rely on agents modeling their epistemic and moral commitment to belief systems and each other, may not be easily outsourced to robots. We include a longer summary of implications for robot outsourcing in the online supplemental materials.

This research also speaks to the future of religion, which is declining in many world regions, including East Asia, North America, and Europe (Jackson et al., 2021). Adherents' commitment to religions has been sustained over the centuries by many factors, but one critical factor may have been credible religious exemplars who sacrificed their resources, health, and money for their faith. These historical figures draw a sharp contrast to the scandal-prone television evangelists of today and—potentially—the unfeeling robot preachers of tomorrow. We cannot forecast what the future holds for religion, but if religious leaders continue to become less credible figures in society, their followers could continue to become less committed.

References

Atran, S. (1998). Folk biology and the anthropology of science: Cognitive universals and cultural particulars. *Behavioral and Brain Sciences*, *21*(4), 547–569. https://doi.org/10.1017/S0140525X98001277

Barrett, J. L. (2004). *Why would anyone believe in God?* AltaMira Press.

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, *1*(1), 71–81. https://doi.org/10.1007/s12369-008-0001-3

Boyd, R., & Richerson, P. J. (1988). *Culture and the evolutionary process*. University of Chicago Press.

Boyer, P. (2007). *Religion explained: The evolutionary origins of religious thought*. Basic Books.

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. WW Norton & Company.

Cavalli-Sforza, L. L., & Feldman, M. W. (1981). *Cultural transmission and evolution: A quantitative approach*. Princeton University Press.

ChatGPT: Optimizing Language Models for Dialogue. (2022, November 30). *OpenAI*. https://openai.com/blog/chatgpt/

CNN. (n.d.). *The android priest that's revolutionizing Buddhism*. CNN. Retrieved March 28, 2022, from https://www.cnn.com/travel/article/mindar-android-buddhist-priest-japan/index.html

Cornwall, M., Albrecht, S. L., Cunningham, P. H., & Pitcher, B. L. (1986). The dimensions of religiosity: A conceptual model with an empirical test. *Review of Religious Research*, *27*(3), 226–244. https://doi.org/10.2307/3511418

Dean, K. (2014). *Taoist ritual and popular cults of Southeast China*. Princeton University Press.

Deus Ex Machina: Religions Use Robots to Connect With the Public—WSJ. (n.d.). Retrieved April 11, 2022, from https://www.wsj.com/articles/deus-ex-machina-religions-use-robots-to-connect-with-the-public-11553782825

Devlin, K. (2020). *Turned on: Science, sex and robots*. Bloomsbury Publishing.

Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial. *JMIR Mental Health*, *4*(2), Article e19. https://doi.org/10.2196/mental.7785

Ford, M. (2015). *Rise of the robots: Technology and the threat of a jobless future*. Basic Books.

Frank, M. R., Autor, D., Bessen, J. E., Brynjolfsson, E., Cebrian, M., Deming, D. J., Feldman, M., Groh, M., Lobo, J., & Moro, E. (2019). Toward understanding the impact of artificial intelligence on labor. *Proceedings of the National Academy of Sciences*, *116*(14), 6531–6539. https://doi.org/10.1073/pnas.1900949116

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, *114*, 254–280. https://doi.org/10.1016/j.techfore.2016.08.019

Gervais, W. M., & Henrich, J. (2010). The Zeus problem: Why representational content biases cannot explain faith in gods. *Journal of Cognition and Culture*, *10*(3–4), 383–389. https://doi.org/10.1163/156853710X531249

Gervais, W. M., Najle, M. B., & Caluori, N. (2021). The origins of religious disbelief: A dual inheritance approach. *Social Psychological and Personality Science*, *12*(7), 1369–1379. https://doi.org/10.1177/1948550621994001

GPT-3. (2020). *A robot wrote this entire article. Are you scared yet, human?* The Guardian.

Graham, J., & Haidt, J. (2010). Beyond beliefs: Religions bind individuals into moral communities. *Personality and Social Psychology Review*, *14*(1), 140–150. https://doi.org/10.1177/1088868309353415

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, *96*(5), 1029–1046. https://doi.org/10.1037/a0015141

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, *315*(5812), 619. https://doi.org/10.1126/science.1134475

Gray, K., & Wegner, D. M. (2012). Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition*, *125*(1), 125–130. https://doi.org/10.1016/j.cognition.2012.06.007

Henrich, J. (2009). The evolution of costly displays, cooperation and religion: Credibility enhancing displays and their implications for cultural evolution. *Evolution and Human Behavior*, *30*(4), 244–260. https://doi.org/10.1016/j.evolhumbehav.2009.03.005

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Henrich, J., & McElreath, R. (2003). The evolution of cultural evolution. *Evolutionary Anthropology: Issues, News, and Reviews*, *12*(3), 123–135. https://doi.org/10.1002/evan.10110

Hoogeveen, S., Haaf, J. M., Bulbulia, J. A., Ross, R. M., McKay, R., Altay, S., Bendixen, T., Berniūnas, R., Cheshin, A., & Gentili, C. (2022). The Einstein effect provides global evidence for scientific source credibility effects and the influence of religiosity. *Nature Human Behaviour*, *6*(4), 523–535. https://doi.org/10.1038/s41562-021-01273-8

Jackson, J. C., Caluori, N. E., Gray, K., & Gelfand, M. (2021). The new science of religious change. *American Psychologist*, *76*(6), 838–850. https://doi.org/10.1037/amp0000818

Johnson, K. A., Okun, M. A., & Cohen, A. B. (2015). The mind of the Lord: Measuring authoritarian and benevolent God representations. *Psychology of Religion and Spirituality*, *7*(3), 227–238. https://doi.org/10.1037/rel0000011

Kavanagh, C., & Jong, J. (2019). *Is Japan Religious?* https://psyarxiv.com/qyt95

Kendal, R. L., Boogert, N. J., Rendell, L., Laland, K. N., Webster, M., & Jones, P. L. (2018). Social learning strategies: Bridge-building between fields. *Trends in Cognitive Sciences*, *22*(7), 651–665. https://doi.org/10.1016/j.tics.2018.04.003

Kraft-Todd, G. T., Bollinger, B., Gillingham, K., Lamp, S., & Rand, D. G. (2018). Credibility-enhancing displays promote the provision of nonnormative public goods. *Nature*, *563*(7730), 245–248. https://doi.org/10.1038/s41586-018-0647-4

LaFargue, M. (1994). *Tao and method: A reasoned approach to the Tao Te Ching*. SUNY Press.

Langston, J., Speed, D., & Coleman, T. J., III. (2020). Predicting age of atheism: Credibility enhancing displays and religious importance, choice, and conflict in family of upbringing. *Religion, Brain & Behavior*, *10*(1), 49–67. https://doi.org/10.1080/2153599X.2018.1502678

Lanman, J. A. (2012). The importance of religious displays for belief acquisition and secularization. *Journal of Contemporary Religion*, *27*(1), 49–65. https://doi.org/10.1080/13537903.2012.642726

Lanman, J. A., & Buhrmester, M. D. (2017). Religious actions speak louder than words: Exposure to credibility-enhancing displays predicts theism. *Religion, Brain & Behavior*, *7*(1), 3–16. https://doi.org/10.1080/2153599X.2015.1117011

Lawson, E. T., & McCauley, R. N. (1993). *Rethinking religion: Connecting cognition and culture*. Cambridge University Press.

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

Reader, I. (1991). *Religion in contemporary Japan*. University of Hawaii Press.

Roese, N. J., & Amir, E. (2009). Human—Android interaction in the near and distant future. *Perspectives on Psychological Science*, *4*(4), 429–434. https://doi.org/10.1111/j.1745-6924.2009.01150.x

Schjoedt, U., Stødkilde-Jørgensen, H., Geertz, A. W., Lund, T. E., & Roepstorff, A. (2011). The power of charisma—Perceived charisma inhibits the frontal executive network of believers in intercessory prayer. *Social Cognitive and Affective Neuroscience*, *6*(1), 119–127. https://doi.org/10.1093/scan/nsq023

Shariff, A. F., & Norenzayan, A. (2011). Mean gods make good people: Different views of God predict cheating behavior. *The International Journal for the Psychology of Religion*, *21*(2), 85–96. https://doi.org/10.1080/10508619.2011.556990

Shariff, A. F., Willard, A. K., Andersen, T., & Norenzayan, A. (2016). Religious priming: A meta-analysis with a focus on prosociality. *Personality and Social Psychology Review*, *20*(1), 27–48. https://doi.org/10.1177/1088868314568811

Sperber, D. (2010). The guru effect. *Review of Philosophy and Psychology*, *1*(4), 583–592. https://doi.org/10.1007/s13164-010-0025-0

Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, *25*(1), 44–56. https://doi.org/10.1038/s41591-018-0300-7

van der Miesen, M. M., van der Lande, G. J. M., Hoogeveen, S., Schjoedt, U., & van Elk, M. (2022). The effect of source credibility on the evaluation of statements in a spiritual and scientific context: A registered report study. *Comprehensive Results in Social Psychology*, *6*(1–3), 59–84. https://doi.org/10.1080/23743603.2022.2041984

Waytz, A., Gray, K., Epley, N., & Wegner, D. M. (2010). Causes and consequences of mind perception. *Trends in Cognitive Sciences*, *14*(8), 383–388. https://doi.org/10.1016/j.tics.2010.05.006

Wesselmann, E. D., VanderDrift, L. E., & Agnew, C. R. (2016). Religious commitment: An interdependence approach. *Psychology of Religion and Spirituality*, *8*(1), 35–45. https://doi.org/10.1037/rel0000024

White, C. J., 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

White, C. J., Norenzayan, A., & Schaller, M. (2019). The content and correlates of belief in Karma across cultures. *Personality and Social Psychology Bulletin*, *45*(8), 1184–1201. https://doi.org/10.1177/0146167218808502

Whitehouse, H., Lanman, J. A., Downey, G., Fredman, L. A., Swann, W. B., Jr., Lende, D. H., McCauley, R. N., Shankland, D., Stausberg, M., & Xygalatas, D. (2014). The ties that bind us: Ritual, fusion, and identification. *Current Anthropology*, *55*(6), 674–695. https://doi.org/10.1086/678698

Willard, A. K., & Cingl, L. (2017). Testing theories of secularization and religious belief in the Czech Republic and Slovakia. *Evolution and Human Behavior*, *38*(5), 604–615. https://doi.org/10.1016/j.evolhumbehav.2017.01.002

Worthington, E. L., Jr., Wade, N. G., Hight, T. L., Ripley, J. S., McCullough, M. E., Berry, J. W., Schmitt, M. M., Berry, J. T., Bursley, K. H., & & O'Connor, L. (2012). Religious Commitment Inventory—10. *Journal of Counseling Psychology*, *50*(1), 84–96. https://doi.org/10.1037/0022-0167.50.1.84

Xygalatas, D., Mitkidis, P., Fischer, R., Reddish, P., Skewes, J., Geertz, A. W., Roepstorff, A., & Bulbulia, J. (2013). Extreme rituals promote prosociality. *Psychological Science*, *24*(8), 1602–1605. https://doi.org/10.1177/0956797612472910