Exposure to Robot Preachers Undermines Religious Commitment

Supplemental Materials

Kodaiji Temple Pilot Study

We conducted a brief online pilot study to test whether the presence of Mindar changes people's perceptions of the Kodaiji Temple in ways that could influence their likelihood of donating to the temple. We were particularly interested in two factors. First, participants could reason that the temple was wealthy because they could afford to build and host Mindar, leading them to donate less to the temple. Second, participants could reason that the temple employees had a lower work-ethic, leading them to delegate preacher duties to a robot. Both factors would presumably undermine our hypothesized credibility mechanism, so it was important to evaluate the likelihood of these confounds. Since we realized the potential significance of these factors following our field study, we evaluated them in an independent online sample.

Method

Participants. We recruited 100 participants (59 men, 41 women; M_{age} = 38.47, SD_{age} = 8.73) from Amazon Mechanical Turk using the Cloudresearch platform.

Manipulation. We manipulated knowledge of Mindar using a between-subjects design. All participants in the study read a description of the Kodaiji Temple adapted from the Temple's website (https://www.kodaiji.com/e_index.html). For 48 participants in the control condition, this description contained no information about Mindar. The remaining 52 participants in the experimental condition read that an android preacher by the name of "Mindar" had recently been introduced to the temple and read a brief description of Mindar's duties that read as follows and was adapted from press releases: "Recently, the temple introduced a robot named "Mindar" to preach sermons. The adult-sized android, modeled after Kannon Bodhisattva (the Buddhist Goddess of Mercy) was introduced to the Kodai-ji Temple in 2019."

Measures. Participants completed three measures. The first measure was a manipulation check, with the wording "Does the Kodaiji Temple employ a robot priest by the name of 'Mindar'?" Participants could respond "yes," "no," or "not sure" to this item. The remaining two items were statements, and participants could rate their agreement with each statement using a 1 ("Strongly Disagree") - 10 ("Strongly Agree") scale. The first item was "The Kodaiji Temple is very wealthy" and the second item was "The monks at the Kodaiji temple work hard."

Results

Manipulation Check

We first tested whether the manipulation was effective. After we excluded three participants who indicated that they were "not sure" for our manipulation check item, we found that 84% of participants in the experimental condition indicated knowledge of Mindar whereas only 11% in the control condition indicated knowledge of Mindar. In a logistical regression—which was appropriate given the binary nature of our manipulation check—we found that participants in the experimental condition were significantly more likely to indicate knowledge of Mindar compared to those in the control condition, $b = 3.79$, $SE = .61$, $z = 6.20$, $p < .001$.

Wealth and Work-Ethic Analyses

We next estimated experimental effects of condition on perceptions of Kodaiji Temple's wealth and the work ethic of the monks. In two independent samples t-tests, we found no significant differences between the control ($M = 7.63$) and experimental ($M = 7.39$) conditions on perceptions of wealth, $t(94.90) = 0.58$, $p = .56$. There were also no differences between the control ($M = 8.10$) and experimental ($M = 7.90$) conditions on perceptions of monk work ethic, $t(87.36) = 0.58$, $p = .56$. Since these parameters were highly similar (the *t*- and *p*-statistics were

identical at two-degrees of freedom), we also tested whether the outcomes were colinear. But work-ethic only correlated at $r = .16$ with perceptions of wealth.

One limitation of this study was that our participants were simply reading about the Kodaiji Temple, rather than visiting and worshipping at the Temple. Nevertheless, these results make it seem unlikely that awareness of Mindar is sufficient to change perceptions of the Kodaiji Temple's wealth or the work ethic of the temple's monks.

Additional Methodological Detail for Study 1

Table S1 presents the wording of key control variables from Study 1.

Table S1.

Religious Belief Exploratory Measures in Study 1

God Belief
Do you believe in God?
Karma Belief
1. I believe in the idea that past good deeds and bad deeds will eventually come back to you (karma).
2. When people are met with misfortune, they have brought it upon themselves by behavior in a past life.
3. When people experience good fortune, they have brought it upon themselves by behavior in a past life.
4. If a person does something bad, even if there are no immediate consequences, they will be punished for it in a future life.

5. When someone does a good deed, even if there are no immediate consequences, they will be rewarded for it in a future life.

Views of God

1. God is forgiving.
2. God is merciful.
3. God is gracious.
4. God is compassionate.
5. God is tolerant.
6. God is punishing.
7. God is strict.
8. God is wrathful.
9. God is stern.
10. God is commanding.

Study 1 Supplemental Results

Descriptive Statistics

Whereas the main text displays a histogram of donation rates, Figure S1 displays histograms showing the distribution of perceived credibility of Mindar and the human preacher. Perceived credibility of Mindar had a mean of 3.12 and a standard deviation of .88. Perceived credibility of the human preacher had a mean of 3.51 and a standard deviation of .78. As Figure S1 shows, both variables were normally distributed.

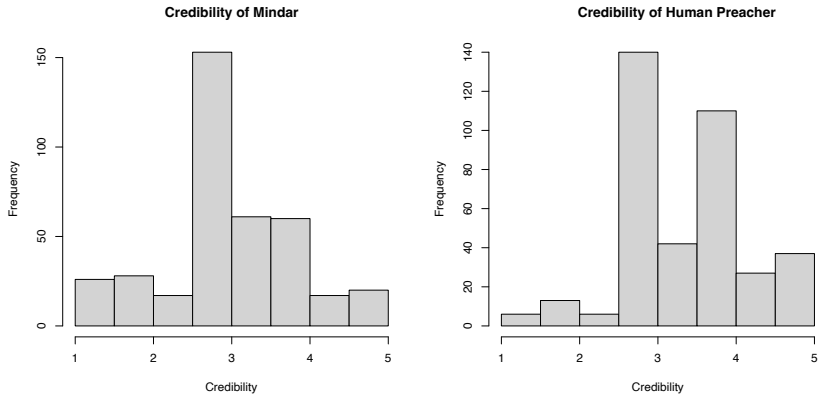

Figure S1. The distributions of credibility ratings when participants rated Mindar (left) and the human preacher (right).

We also examined the descriptive statistics of the additional religion variables. God belief had a mean score of .50 (50% of our Buddhist sample believed in God), with a standard deviation of .50. The mean Karma score was 2.84, with a standard deviation of 1.05. Loving God belief had a mean score of 3.66 (SD = .95), and punitive God belief had a mean score of 2.77 (SD = .89).

Results Including the Full Sample of Participants

Below we replicate our results while including the full sample of participants. In other words, we did not exclude people who came to the temple with the explicit purpose of seeing Mindar. With the full sample of participants, we again found that participants viewed the human preacher (M = 3.53) as significantly more credible than the robot preacher (M = 3.15), $t(801.51) = 6.37$, $p < .001$, $d = .43$.

We next analyzed participants' donations after seeing the robot preacher vs. the human preacher. We found that participants in the robot preacher condition were less likely to donate to the temple (70%) than participants in the human preacher condition (81%). This effect remained

significant when we controlled for prior donation (Model 2) and added other control variables (Model 3). Aside from viewing the robot preacher, the only other significant predictor in this second model was prior donation, which negatively predicted donation amount. All coefficients from these models are displayed in Table 1.

Table S1.

Multiple Regression Models Including All Participants Predicting Donation Likelihood in Study 1

Predictor	***df***	***b* (*SE*)**	***OR***	***t***	***p***	***95% CIs***
Model 1	420					
Robot Preacher		-.62 (.23)	.54	-2.69	.007	.34, .84
Model 2	413					
Robot Preacher		-.55 (.24)	.57	-2.34	.02	.36, .91
Prior Donation		-.68 (.24)	.51	-2.87	.004	.32, .81
Model 3	390					
Robot Preacher		-.53 (.25)	.59	-2.13	.03	.36, .96
Prior Donation		-.62 (.25)	.54	-2.49	.01	.33, .88
God Belief		.46 (.27)	1.59	1.72	.09	.94, 2.71
God as Loving		.14 (.14)	1.15	1.01	.31	.87, 1.52
God as Punitive		-.17 (.15)	.85	-1.13	.26	.63, 1.13
Karma		-.04 (.13)	.97	-.27	.78	.75, 1.24

Note. Confidence intervals are constructed around the odds ratio estimate, so intervals containing 1 are not statistically significant.

We still found no significant effect of condition when we examined the alternative dichotomized variable, which measured whether participants donated their entire pot (p = .34).

The interaction between condition and human preacher credibility still reached significance, b = -.75, SE = .37, OR = .47, t(396) = -2.04, p = .04, *95% CIs* [.23, .96]. When people viewed human preachers as low (-1 SD) in credibility, there was no significant difference in donation likelihood across the human preacher and robot preacher conditions, OR = .79, *95% CIs* [.38, 1.64]. However, when people viewed the human preachers as high (+1 SD) in credibility, they were more likely to donate in the human preacher condition compared to the robot preacher condition, OR = .25, *95% CIs* [.10, .56]. There was no interaction between condition and perceived credibility of robot preachers (p = .74).

Additional Methodological Detail for Study 2

In our main text, we describe the sermons that participants heard in Study 2. The text for both chapters is below:

Chapter 1 Text:

The tao that can be told

is not the eternal Tao.

The name that can be named

is not the eternal Name.

The unnamable is the eternally real.

Naming is the origin

of all particular things.

Free from desire, you realize the mystery.

Caught in desire, you see only the manifestations.

Yet mystery and manifestations

arise from the same source.

This source is called darkness.

Darkness within darkness.

The gateway to all understanding.

Chapter 8 Text:

The highest good is like water, nourishing all things and not contending with them

Dwelling in loathsome places and thus coming close to Tao

Dwelling among the lowly; Great-hearted; Loving in relationships; Sincere in speech;

Establishing order in government; Demonstrating ability in undertakings; Perfect in its timing.

Only where there is no contention can there be flawlessness.

Study 2 Supplemental Results

Descriptive Statistics

Our main text provides a histogram of willingness to donate. Figure S2 illustrates the distributions of the other key continuous variables in Study 2: perceived credibility of the sermon giver, perceived credibility of the temple, willingness to share the sermon's message.

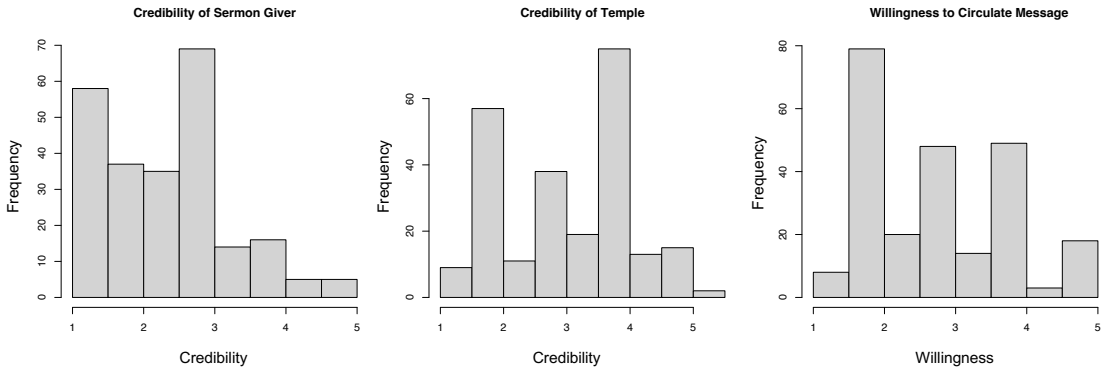

Figure S2. The distributions of credibility ratings of the sermon-giver (left), the Temple (center), and willingness to circulate the message of the sermon (right).

We also examined the descriptive statistics of the additional religion variables. Credibility of the sermon giver had a mean of 2.42 with a standard deviation of 1.02. Credibility of the temple had a mean of 3.14 with a standard deviation of .99. Willingness to circulate the flyer had a mean of .25 (*SD* = .43), meaning that approximately 25% of people said that they were willing to share a flyer about the temple. Willingness to circulate the message of the sermon had a mean of 2.88 with a standard deviation of 1.00. The variables were each normally distributed with a high proportion of low values (see Figure S2), which is because participants in the robot preacher condition gave low ratings on each of the credibility scores and indicated low willingness to circulate the message of the sermon.

Further Structural Equation Model (SEM) Information and Model-Fitting

We estimated the Study 2 path models in R using the lavaan package. SEM is a useful analytic strategy because it returns model-wide coefficients which indicate whether the model provides an appropriate fit to the data in addition to providing path-specific coefficients indicating the strength of the association between two variables. The model-wide coefficients can be used to adjudicate between different approaches to modeling the relationships between a

set of variables. We used these indices to evaluate whether our data were better suited to two kinds of models:

(a) **Multiple Mediation:** Preacher CREDs and temple CREDs served as two simultaneous mediators for the effect of robot preacher condition on religious commitment.

(b) **Serial Mediation:** Exposure to a robot preacher decreased perceived preacher CREDs, which in turn was associated with lower temple CREDs, which in turn was associated with lower commitment across our three religious commitment measures.

The overall fit statistics from each model are displayed in Table S2. The multiple mediation model showed the better fit statistics of the two alternative models. Directly comparing the models with a likelihood ratio test found that the multiple mediation had a significantly better fit than the serial mediation model, $\chi^2 = 36.46$, $p < .001$.

Table S2.

Fit Statistics of Serial Mediation and Multiple Mediation Models

Model	**Chi Squared Test**	**TLI**	**CFI**	**RMSEA**
Multiple Mediation	$\chi^2 = .60$, $p = .44$	1.02	1.00	.00
Serial Mediation	$\chi^2 = 37.06$, $p < .001$	.90	.52	.218

In Figure 5 within the main text, we present the regression path coefficients from this multiple mediation model as effect sizes with stars to indicate statistical significance. Table S3 presents the unstandardized beta, the standard error, the *z*-value, the *p*-value, and the confidence

intervals for each of these effects. For the sake of clarity, we re-display Figure 5 below in Figure S3, with letters on each path to correspond to effects in Table S3.

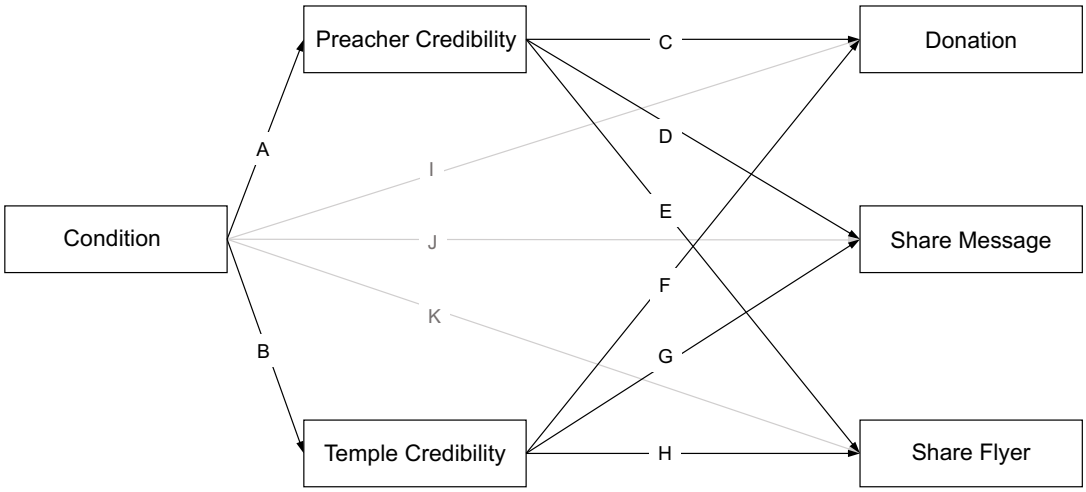

Figure S3. A reproduction of Figure 5 in the main text, with paths assigned letters. Each letter contains a full set of coefficients in Table S3.

Table S3.

Full set of coefficients for Study 2 Path Model

Path	***b* (SE)**	***z***	***p***	**95% CIs**
A	-1.15 (.11)	-10.80	< .001	-1.40, -0.94
B	-.80 (.12)	-6.72	< .001	-0.56, -0.80
C	-.09 (.06)	-1.39	.17	-0.21, 0.04
D	.42 (.07)	0.27, 0.56	< .001	0.27, 0.56
E	.09 (.03)	2.54	.01	0.01, 0.15
F	.71 (.06)	11.35	< .001	0.59, 0.83

G	.20 (.07)	2.85	.004	0.06, 0.33
H	.09 (.03)	3.37	.001	0.04, 0.15
I	-.40 (.15)	-2.71	.007	-0.68, -0.13
J	.39 (.15)	2.62	.009	0.09, 0.68
K	.05 (.07)	.80	.42	-0.08, 0.18

Note. The beta coefficients here are unstandardized, which is why they reflect different values than the standardized beta coefficients displayed in Figure 5.

We summarize the effects of the indirect effects of the better-fitting multiple mediation model in the main text. In this model, temple credibility mediated the effect of condition on donation amounts, *95% CIs* [-1.05, -0.62], flyer distribution, *95% CIs* [-0.17, -0.04], and willingness to share the sermon message, *95% CIs* [-0.41, -0.07]. Preacher credibility mediated the effect of condition on flyer distribution, *95% CIs* [-1.13, -0.02], and willingness to share the sermon message, *95% CIs* [-0.50, -0.19], but not on donation, *95% CIs* [-0.02, 0.18].

Here we summarize the indirect effects of the worse-fitting serial mediation model. This serial mediation model only had three indirect effects, none of which were significant: (a) condition → preacher credibility → temple credibility → donation, *95% CIs* [-0.09, 0.17], (b) condition → preacher credibility → temple credibility → flyer distribution, *95% CIs* [-0.01, 0.02], (c) condition → preacher credibility → temple credibility → willingness to share sermon message, *95% CIs* [-0.05, 0.03]. These indirect effects were non-significant because preacher credibility was not significantly associated with temple credibility after controlling for condition, $b = -0.05$, $SE = 0.08$, $t = -0.71$, $p = .48$. In other words, the effect of robot preacher exposure on temple credibility was independent of participants' perceptions of preacher credibility. This may

be because some participants focused on preacher credibility and others focused on temple credibility.

Credibility as Moderated by Robot Mind Perception

In Study 2, we included eight items measuring participants' general tendency to perceive mind in robots. We included this measure as an exploratory moderator testing whether the robot preacher's perceived credibility 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"). A factor analysis suggested that these items formed a strong one-factor solution, with the first eigenvalue explaining 59% of the total variance, and so we averaged the items together into a single mind perception index (α = .92). We employed linear regressions with gaussian estimation to test whether mind perception interacted with condition on perceptions of (a) preacher credibility and (b) temple credibility.

Our analyses revealed an interesting pattern of results. We observed a significant interaction of condition and robot mind perception on preacher credibility, b = -0.50, SE = 0.17, t = -2.87, p = .005. Participants always saw the robot preacher as less credible than the human preacher, but this gap was smaller among those who perceived more (+1 SD) mind in robots, b = -0.84, SE = 0.15, t = -5.48, p < .001, compared with those who perceived less (-1 SD) mind in robots, b = -1.48, SE = 0.16, t = -9.33, p < .001. On the other hand, there was no significant interaction of condition and robot mind perception on temple credibility, b = -0.14, SE = 0.19, t = -0.75, p = .46. In other words, even participants who perceived robots to have mental capacities

viewed the temple as less credible when it employed robots vs. human preachers. This may be because robots lack specific kinds of mental properties which are necessary for authentic religious belief, even if people perceive them to be capable of basic planning and communication. To this end, perceiving robots to be capable of "thinking" did moderate the effect of condition on temple credibility in the expected direction, b = -0.40, SE = 0.20, t = -2.07, p = .04. This was a small effect, and we encourage future research to explore whether there are certain mind perception qualities which people perceive to be especially important when evaluating religious elites and the institutions that employ them.

Additional Methodological Detail for Study 3

Sermon in Study 3

Below, we have pasted the sermon that appeared to participants in Study 3. The sermon was written by the first author of this paper.

> *God is almighty, and all power belongs to him. God's power may flow through his creatures, yet still it is His to keep and guard. The sun may burst from its chamber and race across the sky, yet it races but with the force of God, and its motions yet belong to God. None can stay the tides as they press towards the shore, yet they have neither might nor strength except as God gives to them. What little power belongs to men, what might we claim to wield, we owe it to the LORD God who has given us our small plot. As God sways his hand, so do nations move under glorious banners and empires rise and fall. And it is true that all power belongs to God.*

Study 3 Supplementary Results

Correlation Table with All Variables

Figure S4 is a correlation plot which shows the correlations between all key dependent variables in Study 3. To be as comprehensive as possible, we have not averaged together likability and charisma in this plot, nor have we averaged together agency and patiency.

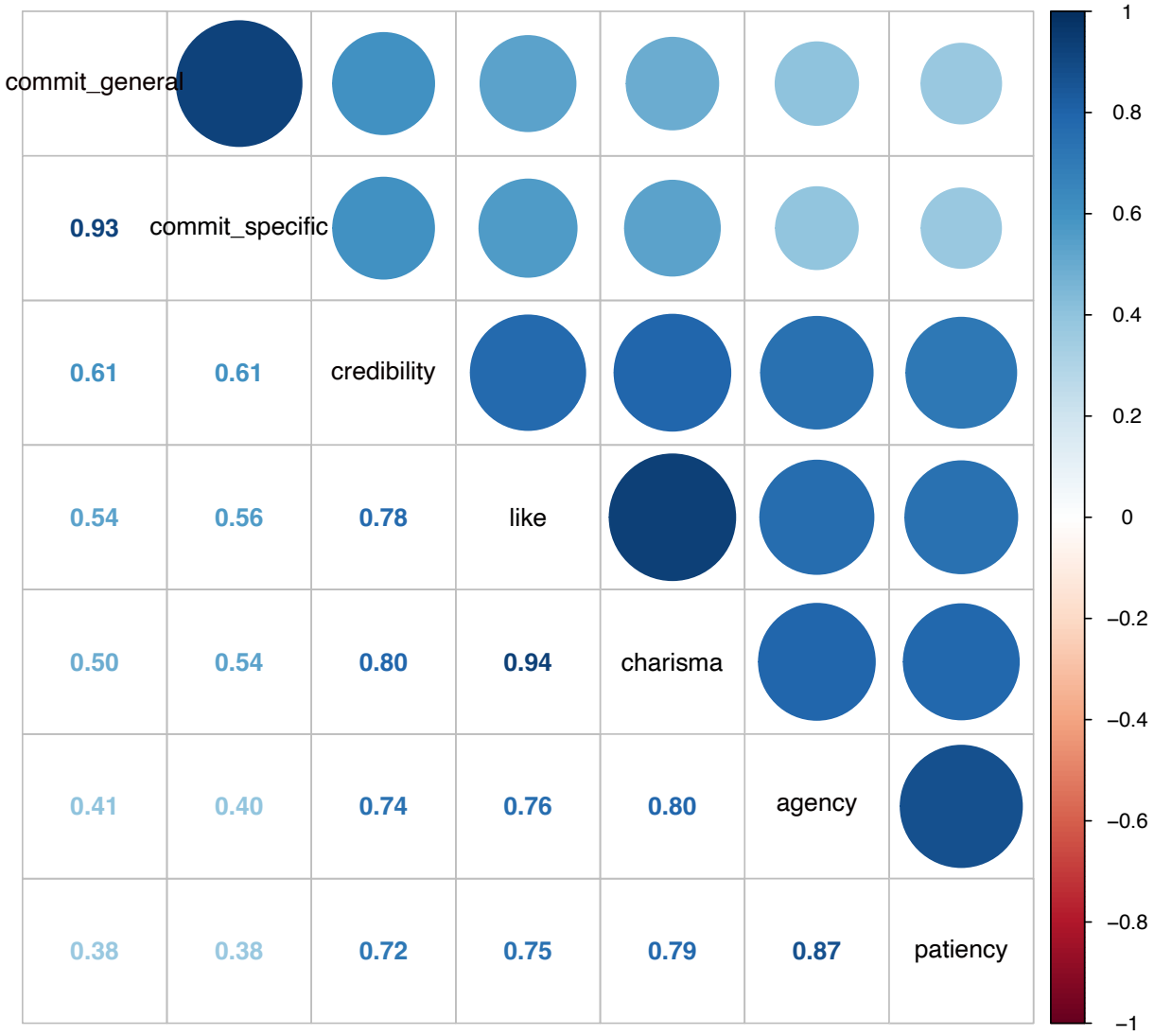

Figure S4. Correlations of all measured variables in Study 3. The bottom diagonal displays the correlation coefficients, and the top diagonal visualizes the magnitude of each correlation using colored circles where large and more deeply blue circles have stronger correlations.

Full Set of Path Coefficients for SEM Model

In Figure 6 within the main text, we present the regression path coefficients from this multiple mediation model as effect sizes with stars to indicate statistical significance. Table S4 presents the unstandardized beta, the standard error, the *z*-value, the *p*-value, and the confidence intervals for each of these effects. For the sake of clarity, we re-display Figure 6 below in Figure S5, with letters on each path to correspond to effects in Table S4.

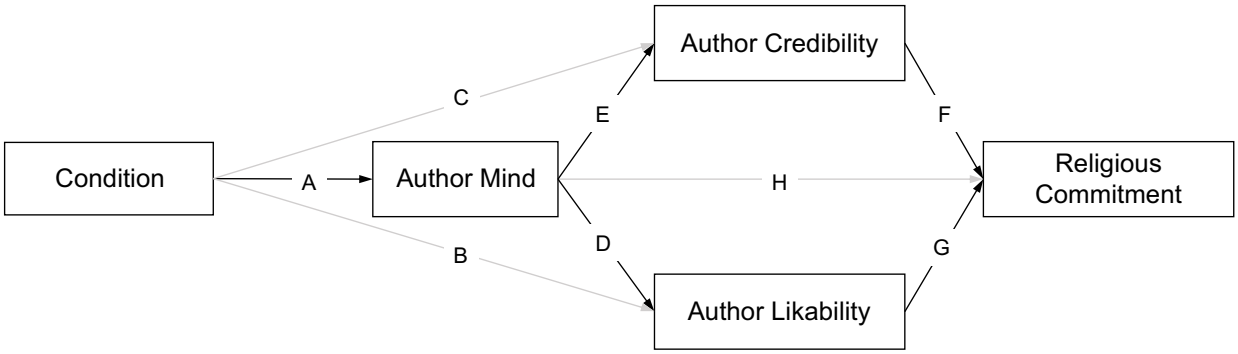

Figure S5. A reproduction of Figure 6 in the main text, with paths assigned letters. Each letter contains a full set of coefficients in Table S4.

Table S4.

Full set of coefficients for Study 3 SEM

Path	***b* (SE)**	***z***	***p***	**95% CIs**
A	-5.29 (0.29)	-18.46	< .001	-5.84, -4.72
B	0.23 (0.32)	0.71	0.48	-0.43, 0.85
C	0.26 (0.28)	0.95	0.34	-0.30, 0.77

D	0.71 (0.05)	14.91	< .001	0.71, 0.81
E	0.65 (0.04)	15.74	< .001	0.56, 0.72
F	0.51 (0.07)	6.93	< .001	0.36, 0.65
G	0.28 (0.07)	43.80	< .001	0.14, 0.42
H	-0.07 (.06)	-1.12	0.26	-0.19, 0.05

Note. The beta coefficients here are unstandardized, which is why they reflect different values than the standardized beta coefficients displayed in Figure 6.

Analyzing Trainer Characteristics

In addition to analyzing perceptions of the sermon author in Study 3, we also analyzed perceptions of the individual who trained the author. In the AI sermon condition, this was the person who trained the AI program that generated the sermon. In the human sermon condition, this was the person who trained the preacher. While our main analyses focused on whether people viewed an AI preacher as less credible than a human preacher, we were also interested in whether people also perceived the trainer of the AI preacher as less credible than the trainer of the human preacher. Each of our models was a general linear model using Gaussian estimation, with AI condition as a dummy-coded fixed effect and no other fixed effects.

Unsurprisingly, we found no significant effect of AI sermon condition on mind perception, $b = 0.34$, $SE = 0.21$, $t = 1.66$, $p = .10$. There was no significant evidence that people perceived the trainer of a human preacher as having more mind than the trainer of an AI algorithm. More surprisingly, we also saw no significant effect of AI sermon condition on trainer credibility, $b = -0.43$, $SE = 0.23$, $t = -1.85$, $p = .07$. Participants may have presumed that the AI's trainer was themselves a religious elite who used the AI program to compose sermons.

Nevertheless, these analyses offer no evidence that exposure to automated preachers affects how people perceive religious elites that may construct and train these robots or AI algorithms.

Moderation by Anthropomorphism

We argue that participants perceive robots as less credible than human preachers because they are less capable of thinking and feeling. In turn, these perceptions of credibility translate to religious commitment. In the main text, we support this mechanistic account by showing that mind perception mediates the effect of condition on perceived credibility. However, it is also possible to test this account using moderation by AI anthropomorphism. People who view AI as humanlike should, by definition, ascribe more mind to AI and perceive AI as more credible than people who view AI algorithms as mere lines of code. We tested this moderation in Study 3 by measuring perceptions of AI anthropomorphism using Bartneck and colleagues' (2009) measure, in which participants rated the extent to which they believed artificial intelligence could be human-like along five key dimensions using a 7-point scale: "fake" (1) vs. "natural" (5), "machinelike" (1) vs. "humanlike" (5), "unconscious" (1) vs. "conscious" (5), and "artificial" (1) vs. "lifelike" (5). Our analyses were general linear models with gaussian estimation.

Results were consistent with the moderation effects in Study 2. In Table S5, we show that participants lower in anthropomorphism perceived AI vs. human sermon authors as especially devoid of mind (Model 1), credibility (Model 2a), and reported especially low religious commitment in the AI vs. human sermon condition (Model 3a) compared with participants who were higher in anthropomorphism. Critically, the interaction between condition and AI anthropomorphism on credibility (Model 2b) and religious commitment (Model 2b) no longer reached significance when we controlled for mind perception, which suggests that participants who anthropomorphize AI are more likely to see AI as credible sermon authors—and are more

likely to anticipate religious commitment after hearing an AI sermon—because they perceive AI algorithms to possess more mind. These results further suggest that mind perception is a key reason why people do not perceive AI algorithms and robot preachers to be credible religious elites.

Table S5.

Moderations by Anthropomorphism

Variable	***b* (SE)**	***t***	***p***	**95% CIs**
Model 1: Mind Perception				
AI Condition	-5.28 (.25)	-21.19	< .001	-5.78, -4.79
AI Anthropomorphism	-0.10 (0.10)	-1.03	0.30	-0.28, 0.09
AI Condition × AI Anthropomorphism	0.94 (0.14)	6.84	< .001	0.67, 1.21
Model 2a: Credibility				
AI Condition	-3.14 (0.26)	-12.05	< .001	-3.65, -2.63
AI Anthropomorphism	0.18 (0.10)	1.77	0.08	-0.02, 0.37
AI Condition × AI Anthropomorphism	0.68 (0.14)	4.75	< .001	0.40, 0.97
Model 2b: Credibility				
AI Condition	-0.21 (0.36)	-0.59	0.56	-0.92, 0.50
AI Anthropomorphism	0.23 (0.09)	2.72	0.007	0.06, 0.40
Mind Perception	0.55 (0.05)	10.25	< .001	0.45, 0.66
AI Condition × AI Anthropomorphism	0.16 (0.13)	1.22	0.23	-0.10, 0.42
Model 3a: Commitment				
AI Condition	-0.99 (0.27)	-3.66	< .001	-1.52, -0.45

AI Anthropomorphism	0.45 (0.10)	4.38	< .001	0.25, 0.66
AI Condition × AI Anthropomorphism	0.44 (0.15)	2.93	< .001	0.14, 0.73
Model 3b: Commitment				
AI Condition	0.65 (0.42)	1.55	0.12	-0.18, 1.48
AI Anthropomorphism	0.48 (0.10)	4.86	< .001	0.29, 0.68
Mind Perception	0.31 (0.06)	4.91	< .001	0.18, 0.43
AI Condition × AI Anthropomorphism	0.14 (0.15)	0.93	0.35	-0.16, 0.45

Results with a Combined Likeability-Charisma Index

In our main text, we note that perceptions of author likability and charisma correlated at $r = 0.94$. In the main text, we chose to analyze likability 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). However, there was a similar effect of condition when we used a combined index where we averaged the likability and charisma items together, $b = -3.69$, $\beta = -0.60$, $SE = 0.30$, $t(272) = -12.19$, $p < .001$, *95% CIs* [-4.29, -3.10], or when we simply analyzed charisma alone, $b = -3.87$, $\beta = -0.60$, $SE = 0.31$, $t(272) = -12.35$, $p < .001$, *95% CIs* [-4.49, -3.25]. We also found highly similar results when we reproduced the main SEM from Study 3 (see Figure 6) using a likability-charisma composite instead of only using likability. We display this SEM below as Figure S6. In this SEM, as in the main text, we found that the indirect effects of credibility, $b = -0.69$, *95% CIs* [-1.21, -0.19], and the likability composite index, $b = -1.05$, *95% CIs* [-1.85, -0.39], on religious commitment were both statistically significant.

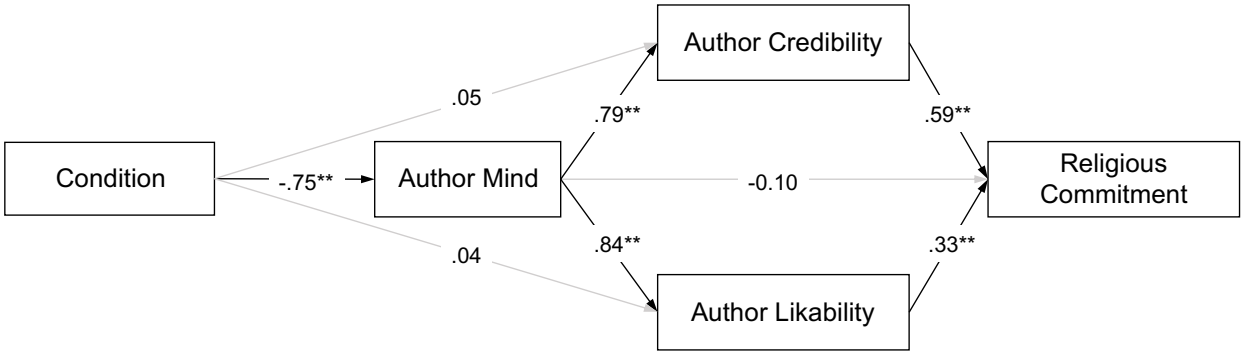

Figure S6. Regression estimates from the Study 3 SEM with a composite likability variable that includes charisma. All estimates are standardized. Single-starred effects represent significance at the $p < .05$ level. Double-starred effects represent significance at the $p < .005$ level. Gray paths were 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.

SEM Without Mind Perception Items

Our SEM in Study 3 (see Figure 6) is different from the SEM in Study 2 (see Figure 5) because we include mind perception as the first mediator in a serial mediation analysis. Below, we provide Figure S7 which Figure 6 without mind perception as the first mediator of this serial mediation, and therefore closely resembles the SEM in Study 2. This model reproduces many of the findings from our main text. We found a significant indirect effect of condition through both credibility, $b = -1.52$, *95% CIs* [-2.08, -1.06], and likeability, $b = -0.79$, *95% CIs* [-1.25, -0.36]. Standardized effects for individual paths are summarized in Figure S7, and starred based on their statistical significance.

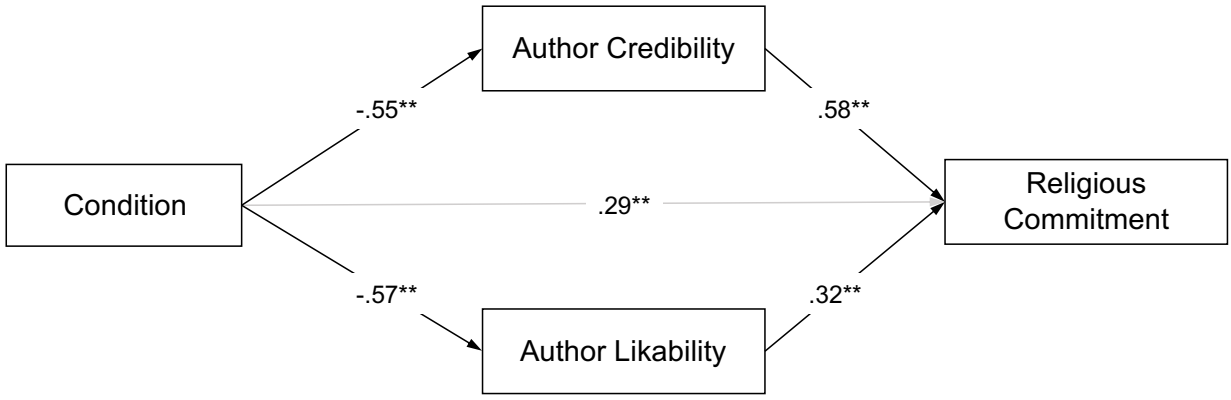

Figure S7. Regression estimates from an SEM which does not include mind perception in a serial mediation. All estimates are standardized. Single-starred effects represent significance at the $p < .05$ level. Double-starred effects represent significance at the $p < .005$ level. Gray paths were estimated but not included in the indirect effect estimation. Condition is dummy-coded such that "1" represents participants in the AI sermon condition and "0" represents participants in the human sermon condition.

Separately Analyzing Agency and Patiency

In our main text, we present analyses that combine our agency and patiency item. In Table S6, we analyze these items separately. These analyses show that, when separately analyzed, agency and patiency have nearly identical associations with religious commitment, credibility, and likability. This suggests that the participants of our study were using these items interchangeably to contrast robot vs. human qualities.

Table S6.

Separately Analyzing Agency and Patiency

Variable	***b* (SE)**	***t***	***p***	**95% CIs**

Associations with Patiency				
Model 1: Religious Commitment	0.38 (0.06)	6.58	< .001	0.27, 0.49
Model 2: Author Credibility	0.54 (0.05)	11.00	< .001	0.43, 0.63
Model 3: Author Likability	0.59 (0.05)	11.85	< .001	0.49, 0.69
Associations with Agency				
Model 1: Religious Commitment	0.41 (0.06)	7.05	< .001	0.29, 0.52
Model 2: Author Credibility	0.58 (0.04)	12.29	< .001	0.49, 0.68
Model 3: Author Likability	0.64 (0.05)	12.98	< .001	0.54, 0.74

Note. Each row of this table is a separate regression model in which either patiency or agency was entered as a fixed effect controlling for the effect of condition, which is why we have listed model names along with the outcome variables. The estimates are similar when we do not control for the effect of condition.

Implications of This Research for Automation

This work attempts to map out domains of work in which human involvement may remain essential. Most research in this area has carved jobs into those whose tasks fall within the capabilities of existing or emerging technologies, and those whose tasks require levels of flexibility, creativity and social intelligence that lie beyond reasonable horizons (Autor et al., 2003; Frey & Osborne, 2017). However, the success of robots will depend not just on how well the work is performed, but also how well it is received (Shariff et al., 2017). Although robots may improve their competencies in being able to mimic (and exceed) human behavior, there are some areas in which people will respond poorly to even superior performance.

In some cases, people reject robots simply because they misperceive robots' capabilities. For instance, recent research showing people's preference for a human doctor over an artificial intelligence for medical diagnoses finds that it is because people overestimate the capabilities that come with human understanding (Cadario et al., 2021; Longoni et al., 2019). Participants overestimated the insight into their own unique health circumstances—and thus the accuracy in diagnosis—that came from having a human doctor. Other studies find that people would remain hesitant to ride in self-driving cars that exceed the level of safety of the average human driver because most individuals tend to believe themselves even safer, overestimating their own capabilities (Shariff et al., 2021).

In other cases, however, it is not the capability of the artificial agent that is lacking, but the ability of these agents to communicate credibility and commitment. Our work is novel in that it incorporates recent insights from psychological research on social learning and the transmission of culture to understand how people respond to, learn from, and trust other agents.

In doing so, it helps predict roles in which the humanity of agents may be indispensable. Domains like religion, which rely on agents modeling their epistemic and moral commitment to belief systems and each other, may not be easily outsourced. Although automation and AI continue to creep into nearly every aspect of our lives, that creep—or at least its effectiveness—may stop at the temple door. Future research may next turn to other areas in which an emphasis on credibility or other actor features are critical to job performance. Holding the performance of a task constant, when does the existence of a mentalizing, committing, suffering mind behind the performance matter?

Supplemental References

Autor, D. H., Levy, F., & Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly Journal of Economics*, *118*(4), 1279–1333.

Cadario, R., Longoni, C., & Morewedge, C. K. (2021). Understanding, explaining, and utilizing medical artificial intelligence. *Nature Human Behaviour*, *5*(12), 1636–1642.

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.

Longoni, C., Bonezzi, A., & Morewedge, C. K. (2019). Resistance to medical artificial intelligence. *Journal of Consumer Research*, *46*(4), 629–650.

Shariff, A., Bonnefon, J.-F., & Rahwan, I. (2017). Psychological roadblocks to the adoption of self-driving vehicles. *Nature Human Behaviour*, *1*(10), 694–696.

Shariff, A., Bonnefon, J.-F., & Rahwan, I. (2021). How safe is safe enough? Psychological mechanisms underlying extreme safety demands for self-driving cars. *Transportation Research Part C: Emerging Technologies*, *126*, 103069.