Supporting Information

Religion-Based Demographic Variables

Study 1

To examine whether participants' responses were associated with their religious background, we first coded attendance at religious services as zero (if participants/parents selected a response in the bottom half of the scale, i.e., reported attending services once a year or less) or one (if participants/parents selected a response in the top half of the scale). We also recoded participants' religious affiliation as zero if they identified as non-religious and one if they identified with any religious group. We split participants this way because individuals who did not believe in God may have experienced confusion responding to questions about God's mind or may otherwise have differed from religious participants. Finally, we conducted a 2 (Behavior: good vs. bad) x 3 (Agent: God vs. Pat vs. self) x 3 (Participant Age: 5- to 6-year-olds vs. 7- to 8-year-olds vs. adults) x 2 (Participant Religion: member of a religious group vs. not a member of a religious group) x 2 (Service Attendance: less frequent vs. more frequent) mixed ANOVA with repeated measures on the first two factors. Neither religion-based variable exerted a main effect or was involved in any interactions in our study ($ps \geq .089$).

In addition to these two variables, which were based on parental responses in the child sample, we also asked child and adult participants, "What do you think about God? Is God real, or not real?" and, "How sure are you that God is [real/not real, whichever participants selected]? Are you very sure, kind of sure, a little sure, or not sure at all?" We recoded these two variables into one scale where 1 = participant was very sure that God is not real and 8 = participant was very sure that God is real. Then, in each age group, we correlated this value with participants' attributions of beliefs about good and bad behaviors to themselves, another person, and God.

That is, we conducted six correlations in each age group (two behavior types x three agents, each correlated with participants' certainty in the existence or non-existence of God). Because this analysis included 18 correlations, uncorrected p values (reported here and in all subsequent analyses) needed to be .003 or lower to pass the Bonferroni-corrected significance threshold. No correlations reached significance ($|r|s \leq .33$, $ps \geq .010$). Therefore, it does not appear that certainty in the existence or non-existence of God was associated with participants' attributions of beliefs to God or other agents.

Study 2

As in Study 1, we analyzed participants' attributions of beliefs to agents using a 3 (Behavior: good vs. bad vs. controversial) x 3 (Agent: God vs. Pat vs. self) x 3 (Participant Age: 5- to 6-year-olds vs. 7- to 8-year-olds vs. adults) x 2 (Participant Religion: member of a religious group vs. not a member of a religious group) x 2 (Service Attendance: less frequent vs. more frequent) mixed design ANOVA with repeated measures. Religion-based demographics did not exert significant main effects, nor were they involved in any significant interactions ($ps \geq .099$). Additionally, as in Study 1, we created a continuous measure reflecting participants' certainty in God's existence or lack of existence. We then correlated this variable with participants' attributions of three different types of beliefs (concerning good, bad, and controversial behaviors) to three different ages (God, another person, and themselves) separately for each of the three age groups. Because this analysis included 27 correlations, p values needed to be .002 or lower to pass the Bonferroni-corrected significance threshold, and no correlations did so ($|r|s \leq .35$, $ps \geq .009$).

<div align="center">

Analyses Including Participant Gender

</div>

Study 1

In prior work (Richert, Shaman, Saide, & Lesage, 2016), girls anthropomorphized God more than did boys. Therefore, we re-ran the ANOVA reported in the main text including participant gender as a factor. That is, we analyzed participants' attributions of beliefs to agents using a 2 (Behavior: good vs. bad) x 3 (Agent: God vs. Pat vs. self) x 3 (Participant Age: 5- to 6-year-olds vs. 7- to 8-year-olds vs. adults) x 2 (Participant Gender: female vs. male) mixed ANOVA with repeated measures on the first two factors. Gender was involved in two three-way interactions, reported below. No other main effects or interactions including gender reached significance ($ps \geq .057$).

First, we found a Behavior x Agent x Participant Gender interaction, $F(3.38, 256.93) = 3.48$, $p = .013$, $\eta_p^2 = .04$. To examine this interaction, we compared attributions of beliefs to agents separately within each Behavior condition for female and male participants (e.g., the extent to which females attributed beliefs about good behaviors differently to God versus Pat, the extent to which females attributed beliefs about good behaviors differently to God versus themselves, the extent to which males attributed beliefs about good behaviors differently to God versus themselves, etc.). This analysis included 12 comparisons; therefore, uncorrected p values needed to be .004 or lower to pass the Bonferroni-corrected significance threshold. Female participants were more likely to report that they ($M = .97$, $SD = .12$), rather than Pat ($M = .84$, $SD = .30$), thought that good behaviors were okay ($p = .002$, Cohen's $d = .40$). Female participants were also more likely to report that Pat ($M = -.75$, $SD = .43$), rather than God ($M = -.94$, $SD = .24$) or the participants themselves ($M = -.91$, $SD = .20$), thought that bad behaviors were okay ($ps < .001$, Cohen's $ds \geq .37$). Male participants were more likely to report that they themselves ($M = .95$, $SD = .14$), rather than God ($M = .78$, $SD = .50$), thought that good behaviors were okay ($p = .003$,

Cohen's d=.32). No other pairwise comparisons in this analysis passed the Bonferroni-corrected significance threshold (ps≥.005, Cohen's ds≤.35).

Second, we found an Agent x Participant Age x Participant Gender interaction, F(5.26, 266.67)=2.96, p=.011, η_p^2=.06. To examine this interaction, we compared females' and males' attributions of beliefs to agents separately within each age group. This analysis included 18 comparisons; therefore, uncorrected p values needed to be .003 or lower to pass the Bonferroni-corrected significance threshold. Adult men were less likely to report that God (M=-.08, SD=.38), rather than Pat (M=.15, SD=.23) or the participants themselves (M=.22, SD=.19), thought that behaviors were okay (ps<.001, |Cohen's d|s≥.49). The difference between adult men's attributions of beliefs to Pat and to themselves did not reach significance; similarly, no other pairwise comparisons in this analysis passed the Bonferroni-corrected significance threshold (ps≥.035, Cohen's ds≤.46).

Study 2

We analyzed participants' attributions of beliefs to agents using a 3 (Behavior: good vs. bad vs. controversial) x 3 (Agent: God vs. Pat vs. self) x 3 (Participant Age: 5- to 6-year-olds vs. 7- to 8-year-olds vs. adults) x 2 (Participant Gender: female vs. male) mixed ANOVA with repeated measures on the first two factors. Gender did not exert a significant main effect, nor was it involved in any significant interactions (ps≥.185).

Similarly, we analyzed the extent to which participants reported that agents would hold consistent beliefs over time using a 3 (Behavior: good vs. bad vs. controversial) x 3 (Agent: God vs. Pat vs. self) x 3 (Participant Age: 5- to 6-year-olds vs. 7- to 8-year-olds vs. adults) x 2 (Participant Gender: female vs. male) mixed ANOVA with repeated measures on the first two

factors. Again gender did not exert a significant main effect, nor was it involved in any significant interactions ($ps \geq .224$).

Study 1: Main Effects

As discussed in the main text, each variable exerted a significant main effect. First, we found a main effect of Behavior ($F(1, 176)=4416.55$, $p<.001$, $\eta_p^2=.96$). Participants reported that agents were more likely to think that good behaviors, rather than bad behaviors, were okay (Cohen's $d=3.82$). Second, we found a main effect of Agent ($F(1.79, 315.29)=9.87$, $p<.001$, $\eta_p^2=.05$). To examine this effect further, we compared each agent with each other agent. This resulted in a total of three comparisons; therefore, uncorrected p values needed to be .017 or lower to pass a Bonferroni-corrected significance threshold. Participants reported that God was less likely to think that behaviors were okay than either Pat ($p=.001$, Cohen's $d=-.26$) or participants themselves ($p<.001$, Cohen's $d=-.28$); participants did not distinguish between themselves and Pat ($p=.869$, Cohen's $d=0$). Third, we found a main effect of Participant Age ($F(2, 176)=8.06$, $p=.001$, $\eta_p^2=.08$). Again, because this analysis included three comparisons, uncorrected p values needed to be .017 or lower to pass the Bonferroni-corrected significance threshold. Adults were more likely to report that agents would think that behaviors were okay than either 5- to 6-year-olds ($p=.005$, Cohen's $d=.52$) or 7- to 8-year-olds ($p<.001$, Cohen's $d=.64$); the two age groups of children did not differ from each other ($p=.478$, Cohen's $d=.18$).

Study 2: Main Effects for Questions About What Agents Believe

As in Study 1, and as reported in the main text, each variable exerted a significant main effect. First, we found a main effect of Behavior ($F(1.93, 319.70)=1562.25$, $p<.001$, $\eta_p^2=.90$). To examine this effect further, we compared each Behavior with each other Behavior. Because this analysis involved three comparisons, p values needed to be .017 or lower to pass the Bonferroni-

corrected significance threshold. Participants reported that agents were more likely to think that good behaviors, rather than controversial behaviors, were okay. Participants also reported that agents were more likely to think that controversial behaviors, rather than bad behaviors, were okay (ps for all pairwise comparisons<.001, Cohen's ds$\geq$1.46).

Second, we found a main effect of Agent ($F(1.79, 296.79)$=10.10, p<.001, η_p^2=.06). To examine this effect further, we compared each Agent with each other Agent. Because this analysis involved three comparisons, p values needed to be .017 or lower to pass the Bonferroni-corrected significance threshold. Participants reported that God was less likely to think that behaviors were okay than either Pat or participants themselves (ps$\leq$.003, Cohen's $|d|$s$\geq$.23); participants did not distinguish between themselves and Pat (p=.433, Cohen's d=.05).

Third, we found a main effect of Participant Age ($F(2, 166)$=3.39, p=.036, η_p^2=.04). To examine this effect further, we compared each age group with each other age group. Because this analysis involved three comparisons, p values needed to be .017 or lower to pass the Bonferroni-corrected significance threshold. Although no pairwise comparisons reached significance after applying this correction, the pattern of results was similar to Study 1. Namely, adults were somewhat more likely to report that agents would think that behaviors were okay than either 5- to 6-year-olds or 7- to 8-year-olds (ps$\leq$.028, Cohen's ds$\geq$.36). The two age groups of children did not differ from each other (p=.923, Cohen's d=0).

Study 2: Interactions for Questions About What Agents Believe

As in Study 1, each two-way interaction reached significance. First, we found an Agent x Participant Age interaction, $F(3.58, 296.79)$=4.14, p=.004, η_p^2=.05. To further investigate this interaction, we conducted nine pairwise comparisons (3 Agents x 3 Age Groups); therefore, p values needed to be .006 or lower to pass the Bonferroni-corrected significance threshold. As

seen in Figure 2 in the main text, children in both age groups did not distinguish among agents ($ps \geq .095$, |Cohen's d|s$\leq.26$). However, adults reported that Pat and they themselves were more likely than God to think that behaviors were okay, collapsing across good, bad, and controversial behaviors ($ps \leq .002$, Cohen's ds$\geq.36$). Adults did not distinguish between themselves and Pat ($p=.041$, Cohen's $d=.23$). These results are similar to Study 1. Importantly, the Agent x Participant Age interaction suggests that the propensity to distinguish between God's mind and human minds when attributing beliefs appears to emerge relatively late in development; we elaborate on this point in the General Discussion in the main text.

Second, we found a Behavior x Participant Age interaction, $F(3.85, 319.70)=6.65$, $p<.001$, $\eta_p^2=.07$. To further investigate this interaction, we conducted nine pairwise comparisons (3 Behavior Types x 3 Age Groups); therefore, p values needed to be .006 or lower to pass the Bonferroni-corrected significance threshold. As depicted in Figure 2 in the main text, 5- to 6-year-olds were more likely than adults to say that good behaviors were okay ($p=.002$, Cohen's $d=.56$); 7- to 8-year-olds did not differ from either of the other age groups ($ps \geq .016$, Cohen's ds$\leq.41$). However, adults were more likely than 5- to 6-year-olds and 7- to 8-year-olds to say that bad behaviors were okay ($ps \leq .002$, Cohen's ds$\geq.52$); 5- to 6-year-olds and 7- to 8-year-olds did not differ from each other ($p=.316$, Cohen's $d=.22$). Similarly, adults were more likely than 7- to 8-year-olds to say that controversial behaviors were okay ($p=.008$, Cohen's $d=.50$); 5- to 6-year-olds did not differ from either of the other age groups in response to controversial behaviors ($ps \geq .015$, Cohen's ds$\leq.50$).

The results for good and bad behaviors are similar across Studies 1 and 2. In both studies, adults may have considered mitigating factors when judging bad behaviors (e.g., they considered it okay to hit someone who was threatening to harm another person), which may have been less

salient to younger participants (Jambon & Smetana, 2014; Schleifer, Schultz, & Lefebvre-Pinard, 1983).

Third, we found a Behavior x Agent interaction, $F(3.46, 574.98)=6.72$, $p<.001$, $\eta_p^2=.04$. To further investigate this interaction, we conducted nine pairwise comparisons (3 Behavior Types x 3 Agents); therefore, p values needed to be .006 or lower to pass the Bonferroni-corrected significance threshold. Participants responded that they themselves were more likely to think that good behaviors were okay than either God or Pat ($ps\leq.003$, Cohen's $ds\geq.22$), who did not differ from each other ($p=.418$, Cohen's $d=.06$). Participants also responded that Pat was more likely to think that bad behaviors were okay than they themselves ($p<.001$, Cohen's $d=.28$); conceptions of God did not differ from conceptions of either Pat or the participants themselves ($ps\geq.012$, Cohen's $ds\leq.20$). Finally, participants responded that they were more likely than God to think that controversial behaviors were okay ($p<.001$, Cohen's $d=.31$); attributions of beliefs about controversial behaviors to Pat did not differ from attributions to God or to the participant themselves ($ps\geq.023$, Cohen's $ds\leq.18$).

Overall, these two-way interactions revealed broadly similar results as Study 1 for concepts of good and bad behaviors.

Replication of Adult Data from Study 2

The purpose of this replication attempt was to obtain more precise estimates of (a) the extent to which individuals report that God, Pat, and they themselves think that good and bad behaviors are okay and (b) the extent to which individuals think that these agents will hold consistent beliefs across time. To do so, we recruited a larger sample of adults to answer the same questions used in Study 2. Because Study 2 did not uncover differences in children's

attributions of beliefs to God, themselves, and another person, the replication included only adults.

Method

Participants. Participants were 118 adults (M_{age}=32.05 years, SD_{age}=8.76 years; 39% female; 75% White; 38% Christian, 55% non-religious, remainder other). Using the same scales as in Study 1, participants reported attending services, on average, "less than once a year" (M=2.00, SD=1.54). We also tested and excluded data from six adults because they failed to correctly answer the attention check question (n=5; the wording of this question was the same as in Study 1) or because they participated in Study 2 (n=1). Procedures for recruiting adult participants and obtaining demographic data were identical to those used in Studies 1 and 2, except that participants received greater compensation ($2.67).

Procedure. The procedure was identical to Study 2, with one exception: after all experimental items, adults completed two additional tasks to measure their theory of mind. These measures were exploratory and not of central focus to the current study; thus, they are not discussed further.

Results and Discussion

What Do Agents Believe? To investigate these participants' responses to questions about what agents believed, we conducted a 3 (Behavior: good vs. bad vs. controversial) x 3 (Agent: God vs. Pat vs. self) repeated measures ANOVA. As in Study 2, this analysis revealed a main effect of Behavior ($F(1.65, 193.36)$=565.92, p<.001, η_p^2=.83) and of Agent ($F(2, 234)$=20.54, p<.001, η_p^2=.15). These main effects were qualified by a Behavior x Agent interaction ($F(3.60, 420.61)$=22.41, p<.001, η_p^2=.16).

To further examine this Behavior x Agent interaction, we conducted nine pairwise comparisons to determine whether adults distinguished among agents when answering questions about each type of behavior (Figure S1). Thus, uncorrected p values needed to be .006 or lower to pass the Bonferroni-corrected significance threshold. Several results replicated findings from Study 2. Adults were *more* likely to report that they themselves, rather than Pat, thought that good behaviors were okay ($p<.001$, Cohen's $d=.50$), and they were *less* likely to report that they themselves, rather than Pat, thought that bad behaviors were okay ($p=.001$, Cohen's $d=-.32$). Other pairwise comparisons for such beliefs failed to reach significance after applying a Bonferroni correction (ps$\geq.008$, Cohen's ds$\leq.25$). Additionally, adults were *less* likely to report that God, rather than Pat and the participants themselves, thought that controversial behaviors were okay (ps$\leq.001$, |Cohen's d|s$\geq.30$). However, unlike in Study 2, adults were also less likely to report that Pat, rather than they themselves, thought that controversial behaviors were okay ($p<.001$, Cohen's $d=-.56$).
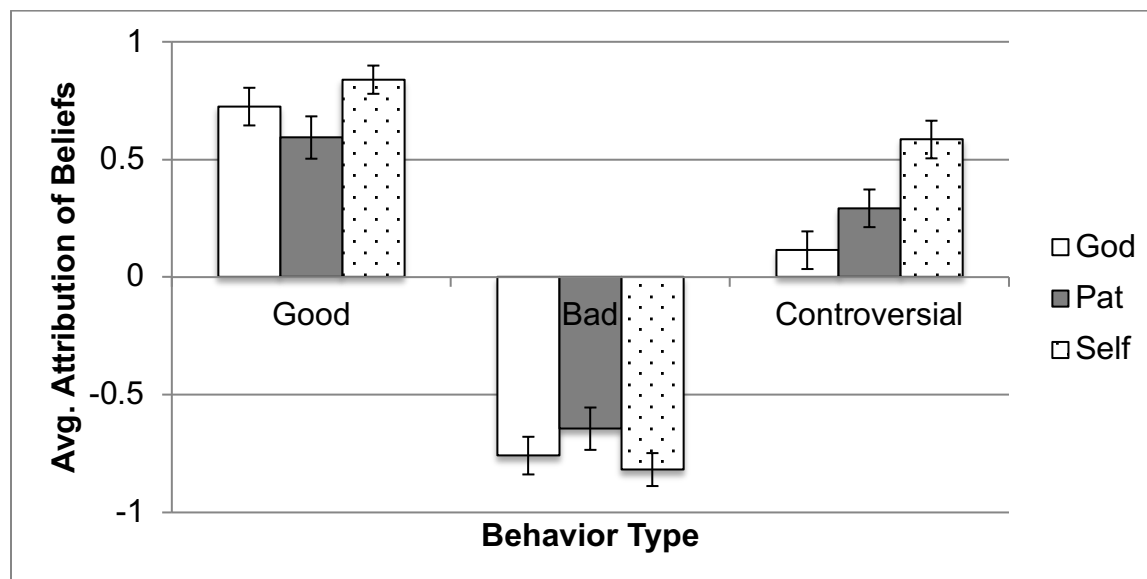
Figure S1. Average extent to which adults reported that each agent thought that good, bad, and controversial behaviors were okay in a replication of Study 2. Error bars represent 95% confidence intervals. Belief attributions were coded such that No=-1, Maybe=0, and Yes=+1.

Will Agents Always Believe The Same Thing? We analyzed responses using a 3 (Behavior: good vs. bad vs. controversial) x 3 (Agent: God vs. Pat vs. self) repeated measures ANOVA (Figure S2). This analysis revealed a main effect of Behavior ($F(1.89, 221.52)=40.53$, $p<.001$, $\eta_p^2=.26$). Adults were less likely to report that agents would always hold the same beliefs about controversial behaviors than about good or bad behaviors (ps$<.001$, |Cohen's d|s$\geq.59$); adults were equally likely to report that agents would always hold the same beliefs about good and bad behaviors ($p=.141$, Cohen's $d=-.14$). As in Study 2, the omnibus ANOVA also revealed a main effect of Agent ($F(2, 234)=30.08$, $p<.001$, $\eta_p^2=.21$). As in Study 2, adults were less likely to report that Pat, rather than God or they themselves (ps$<.001$, |Cohen's d|s$\geq.43$), would always hold the same beliefs. Unlike in Study 2, adults were also less likely to

report that God, rather than they themselves, would always hold the same beliefs (p=.006, Cohen's d=-.26). As in Study 2, the Behavior x Agent interaction did not reach significance (F(3.63, 424.22)=1.83, p=.129, η_p^2=.015).
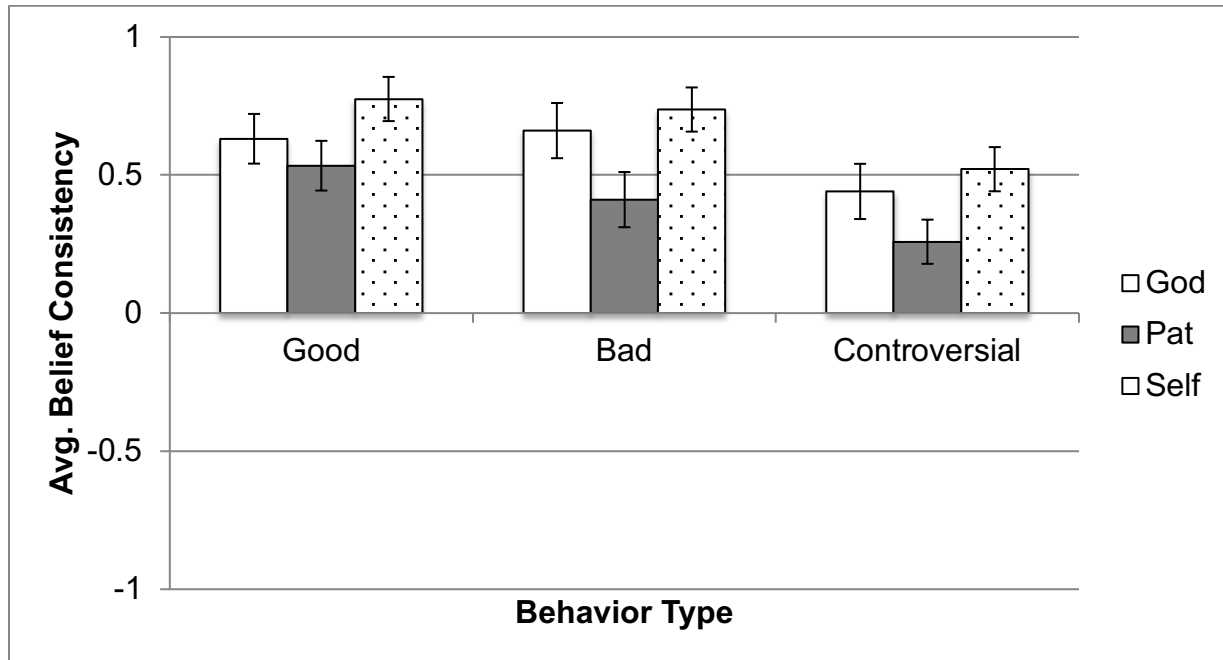

Figure S2. Average extent to which adults reported that agents would hold consistent moral beliefs across time in a replication of Study 2. Error bars represent 95% confidence intervals. Attributions of consistency were coded such that No=-1, Maybe=0, and Yes=+1.

In sum, Study 2 and this replication among adults yielded largely consistent results. In both datasets, adults were more likely to report that they themselves, rather than God or Pat, thought that good behaviors were okay (note that in this replication, differences in adults' attributions of beliefs about good behaviors to themselves and to God passed the traditional .05 significance threshold but failed to remain significant after applying a Bonferroni correction). Furthermore, in both data collections, adults were less likely to report that they, rather than Pat, thought that bad behaviors were okay. Finally, in both data collections, adults were less likely to

report that God, rather than Pat or they themselves, thought that controversial behaviors were okay. Overall, adults appear to distinguish among agents by viewing themselves as most likely to think that good behaviors are okay and least likely to think that bad behaviors are okay; adults also view God as least likely to think that controversial behaviors are okay. Furthermore, adults distinguish among agents and belief types when answering questions about belief stability. In both data collections, adults reported that agents were less likely to always hold the same beliefs about controversial behaviors rather than prototypically good behaviors, and adults also reported that another person (Pat) was less likely to always hold the same beliefs than either themselves or God.

References

Jambon, M., & Smetana, J. G. (2014). Moral complexity in middle childhood: Children's evaluations of necessary harm. *Developmental Psychology, 50,* 22-33. doi: 10.1037/a0032992

Richert, R. A., Shaman, N. J., Saide, A. R., & Lesage, K. A. (2016). Folding your hands helps God hear you: Prayer and anthropomorphism in parents and children. In A. Village & R. Hoods (Eds.), *Research in the Social Scientific Study of Religion* (pp. 140-157). Leiden: Brill.

Schleifer, M., Schultz, T. R., & Lefebvre-Pinard, M. (1983). Children's judgments of causality, responsibility, and punishment in cases of harm due to omission. *British Journal of Developmental Psychology, 1*, 87-97. doi: 10.1111/j.2044-835X.1983.tb00546.x