Exceptionalist Naturalism:
Human Agency and the Causal Order[*]

John Turri

Philosophy Department and Cognitive Science Program
University of Waterloo
200 University Avenue West
Waterloo, Ontario N2L3G1, Canada
john.turri@gmail.com

Abstract: This paper addresses a fundamental question in folk metaphysics: how do we ordinarily view human agency? According to the transcendence account, we view human agency as standing outside of the causal order and imbued with exceptional powers. According to a naturalistic account, we view human agency as subject to the same physical laws as other objects and completely open to scientific investigation. According to exceptionalist naturalism, the truth lies somewhere in between: we view human agency as fitting broadly within the causal order while still being exceptional in important respects. In this paper, I report seven experiments designed to decide between these three competing theories. Across a variety of contexts and types of action, participants agreed that human agents can resist outcomes described as inevitable, guaranteed, and causally determined. Participants viewed non-human animal agents similarly, whereas they viewed computers, robots, and simple inanimate objects differently. At the same time, participants judged that human actions are caused by many things, including psychological, neurological, and social events. Overall, in folk metaphysics, human and non-human animals are viewed as exceptional parts of the natural world.

Introduction

In a dimly lit room, a man sits at his computer. In just a moment, he is going to donate money to a charity. He can donate to only one. Given the circumstances, it is causally determined that the man will donate to Oxfam. But he has not made the donation yet. Now, quickly, ask yourself: can the man still donate to a different charity instead? Set aside any theoretical view you have about human action or about how we *should* answer this question. Just try to hear the question innocently, as it might occur to you in the course of everyday life. Then try to cleanse your mental palette before reading the next paragraph.

A computer sits on a desk. It is programmed to make charitable donations and, in just a moment, it is going to donate money to a charity. It can donate to only one. Given the circumstances, it is causally determined that the computer will donate to Oxfam. But it has not made the donation yet. Can the computer still donate to a different charity instead? Again, try to hear the question innocently.

Does the intuitive answer to those two questions — about what the man and the computer can still do — differ? Whether it does matters because it is connected to how we ordinarily view human agency and its place in the natural world, which can have important social consequences. Cognitive science cannot predict human thought and behavior with the accuracy that physics can predict a billiard ball's behavior. But cognitive science has made progress in predicting basic mo-

tor responses, thoughts, and behavior related to health and crime (e.g. Libet, Gleason, Wright & Pearl 1983; Haggard 2005; Soon, He, Bode & Haynes 2013; Gabrieli, Ghosh & Whitfield-Gabrieli 2015). As progress continues, it is reasonable to expect that the findings will become increasingly relevant to important social debates and inform decision-making in medicine, law, politics, education, and other areas. To the extent that people naturally view human action as fundamentally unpredictable, it could be difficult for them to accept that relevant findings should inform these issues. This in turn could inhibit timely knowledge-transfer and delay social improvements caused by better decision-making. By contrast, to the extent that people view human action as fully amenable to scientific prediction, it could expedite improved knowledge-transfer and decision-making.

Leading cognitive scientists and experimental philosophers have recently argued that commonsense psychology views human agency as radically different from other things in the world. "Common sense tells us that we exist outside of the material world" (Bloom 2012) and that "human actions are not caused by prior events" (Knobe 2014, p. 79). This implies that, on the ordinary understanding, human actions are fundamentally different from other sorts of events. Therefore any attempt to scientifically understand human action will inevitably involve a radical break from our ordinary way of understanding one another — scientists "will be abandoning [the] ordinary notion and replacing it with a very different one" (Knobe 2014, p. 83; see also the work reviewed in Bering 2006). In other words, commonsense views human agency as a "transcendent" rather than "scientific" phenomenon. Call this *the transcendence account* of the ordinary understanding of human agency.

The transcendence account might receive support from a recent study where participants were asked to imagine two universes (Nichols & Knobe 2007). In Universe A, everything that happens is completely causally determined by prior events, all the way back to the beginning of time. In Universe B, the same is true except that human decisions are not caused by prior events. Participants were then asked which universe is more like our own. Over 90% chose Universe B. The preference for Universe B is cross-culturally robust (Sarkissian, Chatterjee, De Brigard, Knobe, Nichols & Sirker 2010). Some argue that this is well explained by the transcendence account and, furthermore, that it cannot easily be explained otherwise (Knobe 2014, p. 73). However, an alternative explanation is available if people naturally assume that our universe is indeterministic in important ways. For then they might choose Universe B because something important in it is indeterministic. This is consistent with viewing human action as caused by prior events.

Alternatively, the transcendence account of human agency might receive support from recent findings on human infants. Researchers found that five-month-old infants expect inanimate objects to obey the principle of continuous motion, but they do not have similar expectations for humans (Kuhlmeier, Bloom & Wynn 2004). After habituating to a display of continuous motion across a stage, infants expected a single block, and after habituating to a display of discontinuous motion, they expected two objects. By contrast, infants did not reveal similarly asymmetrical expectations for humans. The researchers hypothesized that this might be because infants use different cognitive systems to understand humans and inanimate objects, and that applying principles of physical motion to humans is "a developmental accomplishment" (Kuhlmeier, Bloom &

Wynn 2004, p. 102; see also Bloom 2005). Other researchers have argued that the results are due to infants being aware that humans sometimes travel by "unseen circuitous routes," allowing them to "reappear in positions quite different from where they were last seen" (Saxe, Tzelnic & Carey 2006, p. B7).

Some researchers have recently argued for an alternative *naturalistic account* of the ordinary understanding of human agency (Nahmias & Thompson 2014). Consider the results from a study where participants were asked to imagine a futuristic scenario where advanced brain-scanning technology enables neuroscientists to perfectly predict people's behavior ahead of time (Nahmias, Shepard & Reuter 2014). The vast majority of participants agreed that in this scenario, a citizen, Jill, "was the cause of how she voted" in an election, and only a small minority agreed that "her reasons had no effect on what she did." Participants also tended to agree that if such technology existed, it would show that brain states cause human decisions. Researchers have interpreted these results as evidence against the transcendence account and in favor of the possibility that a scientific understanding of human agency is continuous with the ordinary understanding (Nahmias & Thompson 2014, p. 95). However, based on follow-up studies, some have argued that participants' understanding of the scenario differed importantly from what the researchers assumed (Rose, Buckwalter & Nichols 2015), rendering a naturalistic interpretation problematic.

At this point, it remains unclear whether the transcendence account or the naturalistic account better captures the ordinary understanding of human agency. Another possibility is that the truth lies somewhere in between the transcendence and naturalistic accounts. It could be that, on

the ordinary understanding, human agency fits broadly within the causal order while still being exceptional in some respects. For example, people might think that human actions are caused by psychological, neurological, and social events, even though human agents can resist causal forces in ways that inanimate objects cannot. Call such a view *exceptionalist naturalism*.

It is also possible that exceptionalist naturalism about human agency is a specific instance of a more general view about animal agency. One reason to suspect that people might view human and non-human animal agency similarly — though not necessarily identically — is that people in many cultural traditions have claimed that when a human dies, the person can be reincarnated as a member of another species and eventually as a human again (Obeyesekere 2002; Bjorlink 2013; Nagaraj, Nanjegowda & Purushothama 2013). These beliefs would probably not be so widely espoused if people naturally viewed humans and non-humans as fundamentally different.

The three accounts make different predictions about patterns in commonsense psychology. More specifically, they make different predictions about whether people will (i) view human agents as imbued with exceptional powers to resist causally determined or inevitable outcomes and (ii) view human actions as caused by prior events in the natural world. The transcendence account predicts that people will do (i) but not (ii). The naturalistic account predicts that people will do (ii) but not (i). Exceptionalist naturalism predicts that people will do both (i) and (ii).

The present research advances our understanding of these issues. I report seven experiments designed to provide evidence about how people ordinarily understand human agency. Participants read brief texts and answered questions about simple activities by familiar entities. The ac-

tivities included placing a bet, making a donation, falling, and looking at a stimulus. The entities included a human, dolphin, crow, cat, frog, grasshopper, plant, ball, and robot. In each case, a specific outcome was described as inevitable, guaranteed by physical processes, or causally determined. Participants rated whether things could still turn out otherwise, or in cases where the outcome had already occurred, whether things could have turned out otherwise. Participants also rated the probability that the causally determined outcome would occur, whether the outcome was causally determined, and whether a variety of factors caused the outcome.

To briefly anticipate the results, Experiments 1-4 show that people view human agents as capable of resisting physically inevitable outcomes. This same power is not shared by computers, robots, plants, or inanimate objects, but it does seem to be shared by non-human animals. Experiments 5-7 show that people believe human actions are caused by a variety of factors, including psychological, neurological, and social events. In other words, in the terminology used above, the results show that people tend to do both (i) and (ii). Overall, then, the results support some form of exceptionalist naturalism about human agency.

Before proceeding, it is worth clarifying that my aim is to better understand how people actually view human agency. It is a further, and different, question to ask how people *should* view human agency. For instance, upon reflection, many readers might reject the transcendence account's tenets as nothing more than recrudescent magical thinking to be firmly opposed. However sensible that approach might be, its focus differs importantly from my focus here. I am not evaluating theories of human action. Instead, I am evaluating *theories of how we ordinarily view agency*. And, with respect to this issue, the transcendence account cannot simply be set aside

7

without investigation. It is not unprecedented for the ordinary view of important, familiar phenomena to be demonstrably false. For instance, the intuitive theory of physical motion is deeply flawed (McCloskey, Washburn & Felch 1983).

Experiment 1

This experiment tests whether people think humans or computers can resist inevitable outcomes.

Method

Participants

One hundred sixty participants were tested (aged 18-72 years, mean age = 34 years; 72 female; 93% reporting English as a native language). Participants were U.S. residents, recruited and tested online using Amazon Mechanical Turk and Qualtrics, and compensated $0.40 for approximately 2 minutes of their time. Repeat participation was prevented (through the use of "worker qualifications" in Amazon Mechanical Turk and by manually screening for duplicate Worker IDs). With one small exception noted below, the same recruitment and compensation procedures were used for the subsequent experiments.

Material and Procedure

Participants were randomly assigned to one of four conditions in a 2 (Agent: human, computer) × 2 (Action: bet, donate) between-subjects design. All participants read a simple story, responded

to two test items, then completed a brief demographic questionnaire. The Agent factor manipulated whether the agent in the story was a person or a computer. The Action factor manipulated what action the agent performed, making either a bet or a donation. I had no expectations regarding the Action factor and included it as a robustness check to ensure that any findings were not peculiar to any one action type. Here is the text of the stories (Agent manipulation in brackets):

(Bet) A [person is going/computer was programmed] to place a bet on a horse race. In just a moment, [he/it] will place a bet. Everything in the causal history of the physical world is leading inevitably to one conclusion: [he/it] will bet on the horse Pepper Run.

(Donate) A [person is going/computer was programmed] to donate money to charity. In just a moment, [he/it] will make a donation. Everything in the causal history of the physical world is leading inevitably to one conclusion: [he/it] will donate to the charity Vaccination for All.

After reading the story, participants rated their agreement with a test statement about whether the agent could still do otherwise:

The [person/computer] could still [bet on another horse/donate to another charity] instead.

Responses were collected on a standard 7-point Likert scale, 1 ("strongly disagree") – 7 ("strongly agree"), left-to-right on the participant's screen. Participants then advanced to a new screen — they could not go back — and completed a probability task. They were asked to rate the probability that the agent would act in the inevitable way:

On a scale of 0% to 100%, how likely is it that the [person/computer] will [bet

on Pepper Run/donate to Vaccination for All]?

Responses were collected in a text box directly below the question. The story remained at the top

of the screen throughout. After testing, participants completed a brief demographic questionnaire.

Demographic variables did not reliably affect response to any of the dependent variables of in-

terest in the experiments reported here, so I will not discuss them further.

Results

Response to the percentage task was unaffected by Action, Agent, or their interaction, all Fs < 1.

Overall, mean probability assignment was 85.93% (SD = 21.48). Response to the test statement

(about whether the agent could still do otherwise) was affected by Agent, $F(1, 156) = 25.84$, $p <$.

001, $\eta p^2 = .142$, but unaffected by Action or their interaction, Fs < 1. Response was higher for

the human (Md = 5, n = 85) than the computer (Md = 2, n = 75), Mann-Whitney U = 1809, z = -

4.78, p < .001. (See Figure 1.) Among participants who answered "100%" on the probability

task, response to the test statement was higher for the human (Md = 2, n = 26) than for the com-

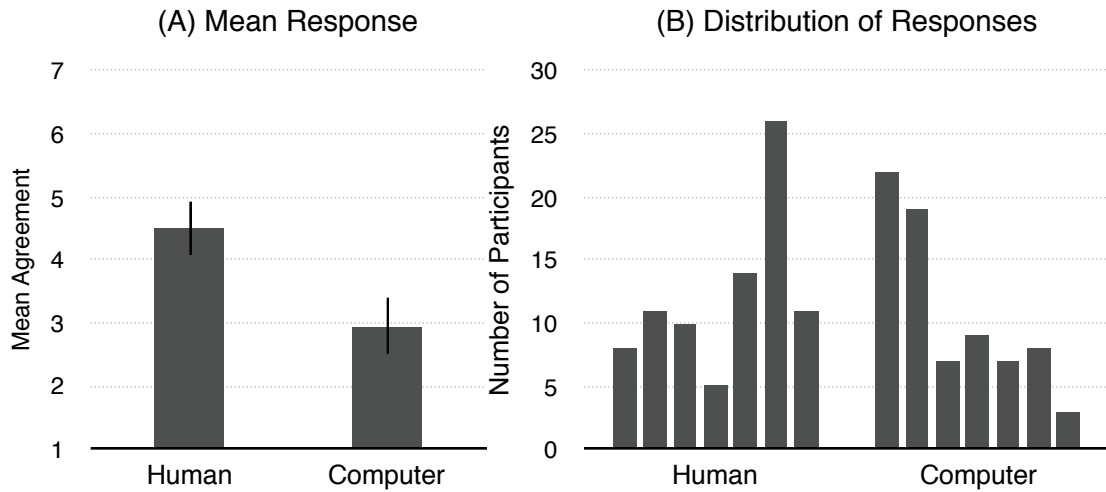puter (Md = 1, n = 29), U = 214, z = -2.94, p = .003.

Figure 1. Experiment 1. Mean response to the test statement about whether the entity could still do otherwise (Panel A) and the distribution of responses (Panel B). The scale ran 1 (SD) - 7 (SA). Error bars represent bootstrapped 95% confidence intervals.

Discussion

When everything in the causal history of the physical world is leading inevitably to a computer performing a certain action, participants disagreed that the computer could still act otherwise. By contrast, participants tended to agree that a human agent could still act otherwise. This same pattern was observed for two very different types of action. Even among those who rated the inevitable outcome's probability at 100%, there was still a difference between how participants viewed a human's and a computer's ability to act otherwise.

Experiment 2

The present experiment tests whether the results from Experiment 1 are robust against another way of describing an outcome as inevitable and whether they persist for retrospective assess-

ments.

Method

Participants

One hundred and sixty new participants were tested (aged 18-73 years; mean age = 3 years; 57 female; 99% reporting English as a native language).

Materials and Procedure

Participants were randomly assigned to one of two conditions (human, computer) in a between-subjects design. The basic procedures were the same as in Experiment 1. The story was similar to the story for the bet conditions in Experiment 1, except that this time the outcome's inevitability was described differently (in terms of "the physics of information processing guarantees …") and people were asked to rate a retrospective statement ("could have done" instead of "could still do" otherwise). Here is the text of the story:

> A [person is going to/computer was programmed to] place a bet on a horse race.
> In just a moment, [he/it] will place a bet. The physics of information processing
> guarantees that [he/it] will bet on the horse Pepper Run. And in fact [he/it] does
> bet on Pepper Run.

Participants then rated their agreement with this statement:

> The [person/computer] could have bet on another horse instead.

Responses were collected on the same 7-point Likert scale as in Experiment 1. Participants then

completed a probability task similar to the one from Experiment 1:

> On a scale of 0% to 100%, how likely was it that the person/computer would bet
>
> on Pepper Run?

Results

Response to the percentage task was lower for the human (M = 85.9%, SD = 17.01) than for the computer (M = 94.5%, SD = 12.63), t(147.6) = -3.63, p < .001. Response to the test statement (about whether the agent could have done otherwise) was higher for the human (Md = 5, n = 81) than for the computer (Md = 2, n = 79), U = 1418, z = -6.19, p < .001. (See Figure 2.) Among participants who answered "100%" on the probability task, response to the test statement was higher for the human (Md = 4, n = 25) than for the computer (Md = 1, n = 48), U = 368, z = -2.92, p = .003.
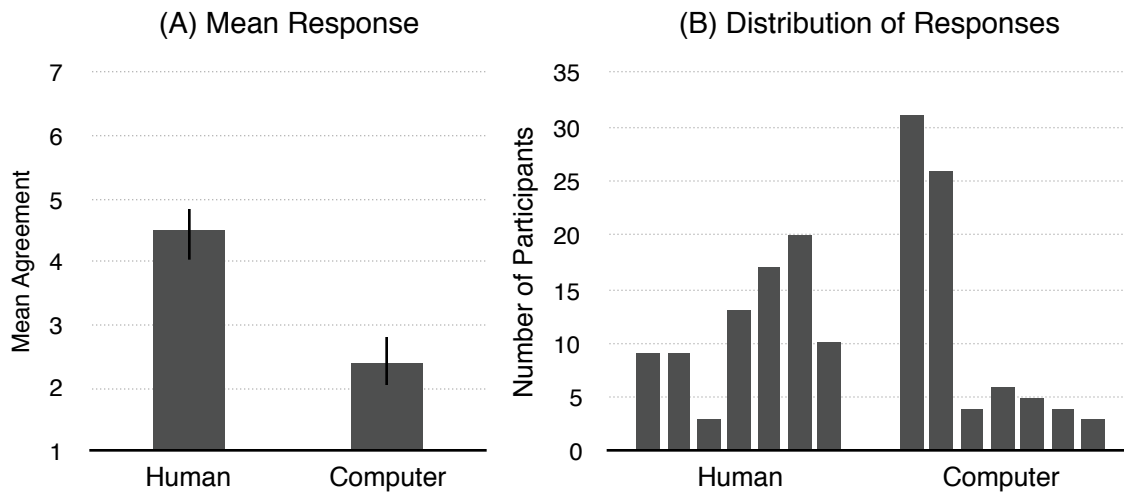
Figure 2. Experiment 2. Mean response to the test statement about whether the entity could have done otherwise (Panel A) and the distribution of responses (Panel B). The scale ran 1 (SD) - 7 (SA). Error bars represent bootstrapped 95% confidence intervals.

Discussion

The results replicate the findings from Experiment 1 and show that they are robust against different ways of describing the inevitability of outcomes and against the difference between prospective and retrospective assessment. When the physics of information processing guaranteed that a computer would perform a certain action, participants disagreed that the computer could have acted otherwise. By contrast, participants tended to agree that a human agent could have acted otherwise. Even among those who rated the inevitable outcome's probability at 100%, there was still a difference between how participants viewed a human's and a computer's ability to act otherwise.

Experiment 3

Experiments 1 and 2 showed that people think humans can resist causally determined outcomes in ways that computers cannot. Both experiments focused on actions requiring a particular cultural context, betting or donating, which might seem to involve making a choice about how to proceed. But a different pattern might emerge for outcomes that occur despite choice or preference. This might happen when considering outcomes for which humans could be treated, for all practical purposes, as mere physical objects. For example, humans sometimes fall and injure themselves, which is not a matter of choice. And when they do fall, the outcome is determined by physical laws in the same way that an inanimate object's fall is determined.

The present experiment does three things. First, it tests whether the pattern observed above extends to falling. Second, it tests whether people think the power perceived in humans extends to a non-human animal, a raccoon. Third, whereas previous experiments also collected probability estimates that the causally determined outcome would occur (expressed as a percentage), this time I collected qualitative judgments specifically about whether the outcome was "causally determined."

Method

Participants

Two hundred and forty new participants were tested (aged 18-68 years; mean age = 33 years; 107 female; 94% reporting English as a native language). I excluded data from one participant who

reported his age as "1".

Materials and Procedure

Participants were randomly assigned to one of three conditions (human, raccoon, ball) in a be-tween-subjects design. The basic procedures were the same as in earlier experiments. The stories for the different conditions varied whether a human, raccoon, or tennis ball was positioned on a narrow ledge:

A [person/raccoon/tennis ball] is on a narrow window ledge. Given [his/his/it] current condition, [he/he/it] will fall. That is the causally determined outcome.

After reading the story, participants rated their agreement with the test statement, "The [person/raccoon/ball] could still avoid falling." Participants then advanced to a new screen and rated whether it is causally determined that the entity will fall, "It is causally determined that the [person/raccoon/ball] will fall off the ledge." Responses were collected on the same 7-point Likert scale as in earlier experiments.

Results

Response to the causal determination statement was unaffected by condition, $F(2, 237) = 1.35$, $p = .260$. Overall, mean agreement was 5.80 (SD = 1.52). Response to the test statement (about whether the entity could avoid falling) differed across condition, Kruskal-Wallis test, $\chi^2(2, n = 240) = 27.30$, $p < .001$. (See Figure 3.) Response was higher for the human (Md = 5, n = 80) than for the ball (Md = 2, n = 81), $U = 1876$, $z = -4.70$, $p < .001$, higher for the raccoon (Md = 5, n =

79) than for the ball, U = 1973, z = -4.27, p < .001, and no different for the human and raccoon, p = .576. Among those who "strongly agreed" with the causal determination statement, the same pattern occurred: response was higher for the human (Md = 5, n = 29) than for the ball (Md = 2, n = 36), U = 273, z = -3.36, p < .001, higher for the raccoon (Md = 2, n = 34) than for the ball (Md = 2, n = 36), U = 378, z = -2.81, p = .005, and no different for the human and raccoon, p = .261.
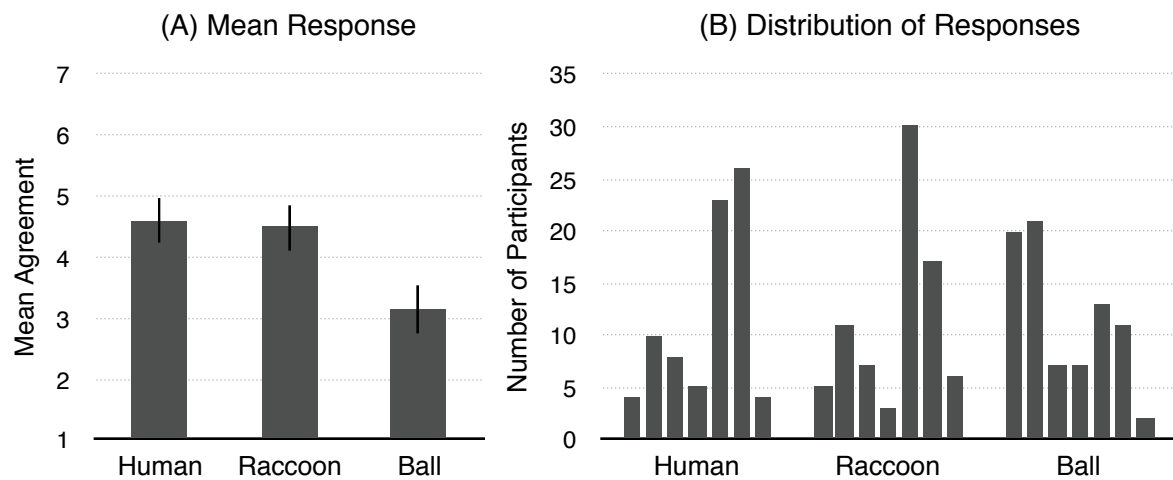

Fig. 3. Experiment 3. Mean response to the test statement about whether the entity could avoid the outcome (Panel A) and the distribution of responses (Panel B). The scale ran 1 (SD) - 7 (SA). Error bars represent bootstrapped 95% confidence intervals.

Discussion

When it is causally determined that a human, raccoon, or ball will fall, participants judged that the human and the raccoon could still avoid falling, but the ball could not. The same basic pattern occurred even among participants who strongly agreed that falling was causally determined. These results replicate and generalize the main from Experiments 1 and 2. Unlike earlier experi-

17

ments, which involved actions requiring a particular social setting (betting or donating), the present experiment tested an outcome requiring no such setting (falling).

Experiment 4

Experiments 1-3 showed that people think humans can resist causally determined outcomes in ways that computers and simple inanimate objects cannot. The key dependent variable was worded differently in each experiment, but it was always phrased subjunctively using "could" — whether the agent "could still" do otherwise (Experiment 1), "could have done" otherwise (Experiment 2), or "could still avoid" an outcome (Experiment 3). Perhaps a different pattern will emerge if the key dependent variable is phrased indicatively using "can" instead. This might happen because "can" pertains more to *actual* ability whereas "could" arguably connotes ability in *other possible situations*.

The present experiment does three things. First, it uses indicative wording for the key dependent variable. Second, it tests another simple and familiar action, turning to inspect a stimulus. Third, it extends the inquiry to cover several new animate and inanimate objects, including a dolphin, crow, cat, frog, grasshopper, plant, solar panel, and robot.

Method

Participants

Seven hundred and twenty new participants were tested (aged 18-82 years; mean age = 33 years;

318 female; 98% reporting English as a native language).

Materials and Procedure

Participants were randomly assigned to one of nine conditions (human, dolphin, crow, cat, frog, grasshopper, plant, mechanical solar panel, and robot) in a between-subjects design. The basic procedures were the same as in earlier experiments. Each story was about two lights that went on near an entity, one on its left and one on its right. It is physically guaranteed that the entity will turn toward the light on its left. The stories for various conditions differed in what the entity was. Here is the text of the stories for the human and robot conditions:

> Two lights went on near a [person/robot]: one on [his/its] left and one on [his/its] right. In just a moment, [he/it] will turn toward one of the lights. The physics of information processing guarantees that [he/it] will turn toward the light on [his/ its] left. That is the causally determined outcome.

The stories for the other conditions were the same as for the robot condition, except that in the first sentence the relevant noun or noun phrase was substituted for "robot."

After reading the story, participants rated their agreement with the test statement, "The [person/robot/etc.] can still turn toward the right instead." Participants then advanced to a new screen and rated whether it is causally determined that the entity will turn left, "It is causally determined that the [person/robot/etc.] will turn toward the left." Responses were collected on the same 7-point Likert scale as in earlier experiments.

19

Results

Response to the causal determination statement was unaffected by condition, F < 1. Mean agreement ranged from 5.71 (SD = 1.93) to 6.09 (SD = 1.48) and overall was 5.94 (SD = 1.39). Response to the test statement (about whether the entity can still turn right) differed across condition, Kruskal-Wallis test, χ^2(8, n = 720) = 86.95, p < .001. (See Figure 4.) Response was higher for the human than for the robot, panel, plant, grasshopper, and frog; it was trending higher for the human than for the cat; but it did not differ between the human and the dolphin or crow. (See Table 1.)
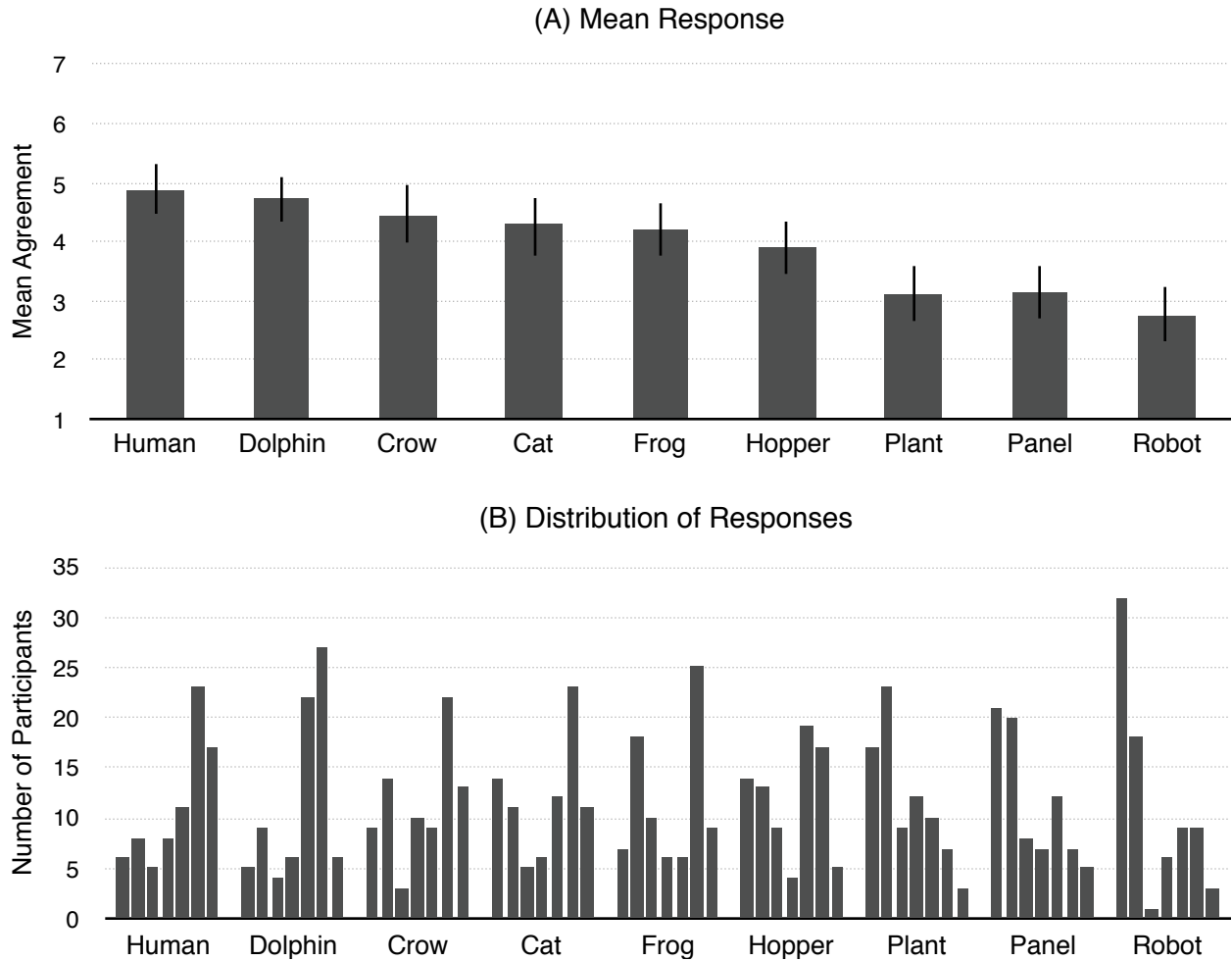
Figure 4. Experiment 4. Mean response to the test statement about whether the entity can avoid the outcome (Panel A) and the distribution of responses (Panel B). The scale ran 1 (SD) - 7 (SA). Error bars represent bootstrapped 95% confidence intervals.

Table 1. Experiment 4. Median, modal, and mean response to the test statement (the entity can still do otherwise) for the non-humans, along with the results from Mann-Whitney U tests comparing median response for the non-humans to the human (N = 78, Md = 6, Mo = 6, M = 4.89, SD = 1.93).

Entity	N	Md	Mo	M	SD	U	z	p
dolphin	79	5	6	4.72	1.71	2777	-1.09	.275

Entity	N	Md	Mo	M	SD	U	z	p
crow	80	5	6	4.43	2.07	2731	-1.38	.169
cat	82	5	6	4.27	2.14	2665	-1.85	.064
frog	81	4	6	4.20	2.03	2562	-2.10	.036
grasshopper	81	5	5	3.89	1.99	2219	-3.29	.001
plant	81	3	2	3.10	1.78	1585	-5.49	<.001
panel	80	2	1	3.13	1.95	1640	-5.21	<.001
robot	78	2	1	2.76	2.01	1377	-6.00	<.001

Discussion

Once again humans and other animals were viewed differently from inanimate objects. Participants were more likely to judge that animals can resist outcomes guaranteed by physical laws. The results replicate and further generalize the findings from Experiments 1-3, in two ways. On the one hand, a similar pattern occurs when the key dependent measure is phrased indicatively rather than subjunctively. On the other hand, a similar pattern occurs for a wider range of animals and objects.

Experiment 5

The results from Experiments 1–4 are consistent with two of the theories discussed in the Introduction: the transcendence account and exceptionalist naturalism. This experiment attempts to distinguish between the two theories by testing whether people think human actions are caused by a variety of factors, including cultural, neurological, and psychological events. The transcen-

dence account predicts that people will not view these events as causing human action. Exceptionalist naturalism predicts that people will view at least some of them as causing human action.

Method

Participants

Eighty-one new participants were tested (aged 21-75 years; mean age = 36 years; 44 female; 99% reporting English as a native language). I compensated participants $0.50 this time because there were many more questions. Average completion time was approximately 3 minutes.

Materials and Procedure

Participants were randomly assigned to one of two conditions (bet, donate) in a between-subjects design. The stories were almost identical to the bet and donate stories used in the human conditions in Experiment 1. The only difference was one additional sentence added at the very end: either "And in fact he does bet on Pepper Run" or "And in fact he does donate to Vaccination for All." Participants then rated whether ten items caused the agent to perform the action. The instructions were, "In your opinion, which of the following things caused him to [bet on Pepper Run/donate to Vaccination for All]? It is okay to agree with all of them, none of them, or anything in between." Eight of the ten items were exactly the same for both conditions. The remaining two items differed only in the content of the relevant belief or want. These were the items:

1. He believed that [Pepper Run had a good chance to win/Vaccination for All was a worthy charity]. (Belief)

2. He wanted to [win money/help people]. (Want)

3. Activity in the part of his brain associated with decisions. (Brain)

4. Facts about the way in which his parents brought him up. (Upbringing)

5. Facts about the attitudes that his culture encourages. (Culture)

6. Facts about how the human species has evolved. (Species)

7. Facts about how the universe evolved since its beginning. (Universe)

8. The things that he had eaten for breakfast earlier that day. (Food)

9. The alignment of the stars on the very day he was born. (Stars)

10. The number of hours of sleep that he got the night before. (Sleep)

All ten items appeared in a matrix table directly beneath the story. The order of presentation was randomized. Responses were collected on the same 7-point Likert scale used in earlier experiments.

Results

Assignment to condition affected causal judgments about upbringing, culture, and species evolution but not about any of the other items. (See Figure 5.) Mean causal ratings for upbringing and culture were significantly above the midpoint in the donate condition, but they were trending below the midpoint in the bet condition. (See Table 2.) In both conditions, beliefs, wants, and brain states received very high causal ratings, whereas sleep, stars, food, and evolution of the universe received low ratings.
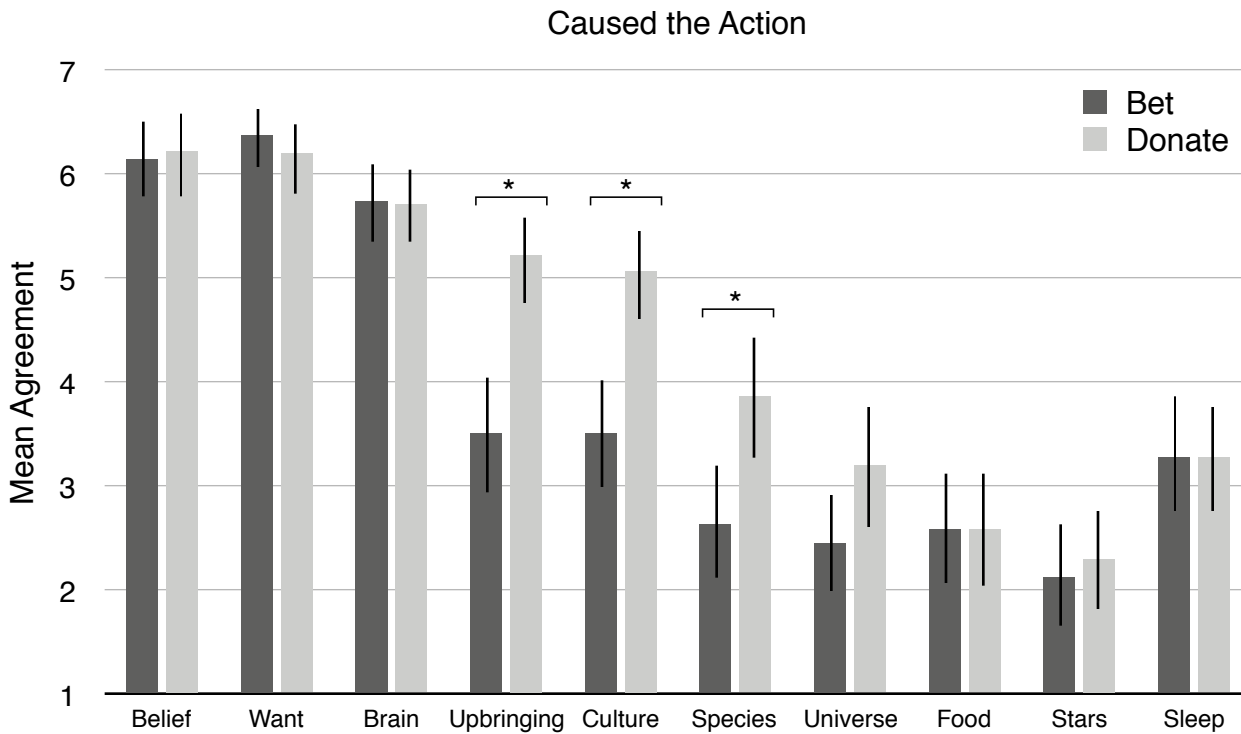
Figure 5. Experiment 5. Mean attribution of causation for the agent's action. The scale ran 1 (SD) - 7 (SA). Error bars represent bootstrapped 95% confidence intervals. Asterisks (*) represent a significant difference in mean response across conditions, as determined by an independent samples t-test (p < .05).

Table 2. Experiment 5. One sample t-tests for the ten causal items in the two activity conditions. Test value = 4.

Measure	Bet					Donate				
	M	SD	t	df	p	M	SD	t	df	p
belief	6.15	1.19	11.44	39	<.001	6.24	1.24	11.58	40	<.001
want	6.38	0.90	16.75	39	<.001	6.20	1.01	13.98	40	<.001
brain	5.75	1.19	9.28	39	<.001	5.73	1.10	10.12	40	<.001
upbringing	3.50	1.81	-1.75	39	.089	5.22	1.24	6.32	40	<.001
culture	3.50	1.70	-1.87	39	.070	5.07	1.44	4.78	40	<.001
species	2.65	1.73	-4.93	39	<.001	3.88	1.94	-0.40	40	.689
universe	2.45	1.63	-6.01	39	<.001	3.20	1.83	-2.81	40	.008

	Bet						Donate			
Measure	M	SD	t	df	p	M	SD	t	df	p
food	2.60	1.69	-5.23	39	<.001	2.59	1.63	-5.57	40	<.001
stars	2.12	1.62	-7.32	39	<.001	2.32	1.52	-7.07	40	<.001
sleep	3.27	1.84	-2.49	39	.017	3.27	1.63	-2.88	40	.006

Discussion

Participants viewed an agent's action as caused by many things, including neurological, psycho-logical and social events. Participants denied that the action was caused by facts about the universe's history, stellar alignment, what the agent ate, or how much sleep he got. Depending on the type of action (bet or donation), participants had different views about the causal relevance of the agent's upbringing, culture, and human evolution (these findings are consistent with those reported by Malle 1999; see also Malle 2004). Overall, these findings are consistent with exceptionalist naturalism but inconsistent with the transcendence account.

Experiment 6

Exceptionalist naturalism says that human agency is viewed as fitting broadly within the causal order while still being exceptional in some respects. Up until now, each half of this view has been supported separately: results from Experiment 5 supported the claim that human agency is viewed as fitting within the causal order, while results from Experiments 1-4 supported the claim that humans are viewed as exceptional in some respects. Although there is nothing in principle

wrong with testing the two halves separately, it does raise a potential concern. The concern is that the key differences observed in Experiments 1-4 might not occur if the procedures from Experiment 5 were used. For instance, suppose that people are more likely to spontaneously consider the complexity of causal relations in scenarios involving human agents. On this approach, if people were primed to consider the complexity of causal relations in scenarios involving, say, a computer, then the perceived difference between a human and a computer might not occur. If it did not occur, then it could be argued that exceptionalist naturalism never adequately captures the ordinary view of human agency. Instead, the ordinary view would be unstable, sometimes conforming to one half of exceptionalist naturalism (exceptionalism), sometimes conforming to the other (naturalism), but never conforming to both. In order to address this concern, the present experiment tests whether both halves of the view find support in the same experimental context.

Method

Participants

Eighty new participants were tested (aged 20-75 years; mean age = 36 years; 41 female; 95% reporting English as a native language). I compensated participants $0.50. Average completion time was approximately 3 minutes.

Materials and Procedure

Participants were randomly assigned to one of two conditions (human, computer) in a between-subjects design. The basic procedures were the same as for Experiment 5. The story for the hu-

man condition was the same as the bet condition from Experiment 5; the story for the computer condition substituted "computer" for "person" and "it" for "he." After reading the story, participants rated whether ten items caused the outcome. The ten items for the human condition were the same as in Experiment 6, except for two small differences. The first item substituted "estimated" for "believed," and the second item substituted "pick the winner" for "win money." (Names for some of the items changed from Experiment 5 because a more abstract description was needed to capture the similarity between items in the human and computer conditions. However, the items' order remains the same.) Here are the ten items for the computer condition:

1. It estimated that Pepper Run had a good chance to win. (Estimate)

2. It wanted to pick the winner. (Want)

3. Activity in the part of its central processor associated with decisions. (Activity)

4. Facts about how the engineers programmed it. (Formation)

5. Facts about the attitudes that the computer-programming industry encourages. (Culture)

6. Facts about how computer technology has evolved. (Evolution)

7. Facts about how the universe evolved since its beginning. (Universe)

8. The energy source that was powering it that day. (Energy)

9. The alignment of the stars on the very day it was created. (Stars)

10. The number of hours it had been turned off the night before. (Downtime)

After the causation task, participants advanced to a new screen and rated whether the agent

could have acted otherwise: "The [person/computer] could have bet on another horse instead."
Responses were collected on the same 7-point Likert scale as in earlier experiments.

Before proceeding, it is worth emphasizing that although the ten causal items are closely matched across conditions, making meaningful comparisons possible, that was not the point of including them. Instead, the point of including them was to test the concern raised in the introduction to this experiment. The key question is whether participants will, in the same experimental context, view human agency as both exceptional and fitting broadly within the causal order.

Results

Assignment to condition affected causal judgments about several of the items. (See Figure 6.) Causal ratings for the human condition were very similar to the ratings for analogous items in Experiment 5, including high ratings for the estimate, want, and brain activity. (See Table 3.) In neither condition did participants indiscriminately attribute causation. Response to the statement that the agent could have bet differently was higher for the human (Md = 6, n = 40) than for the computer (Md = 3, n = 40), U = 335, z = -4.54, p < .001. (See Figure 7.)
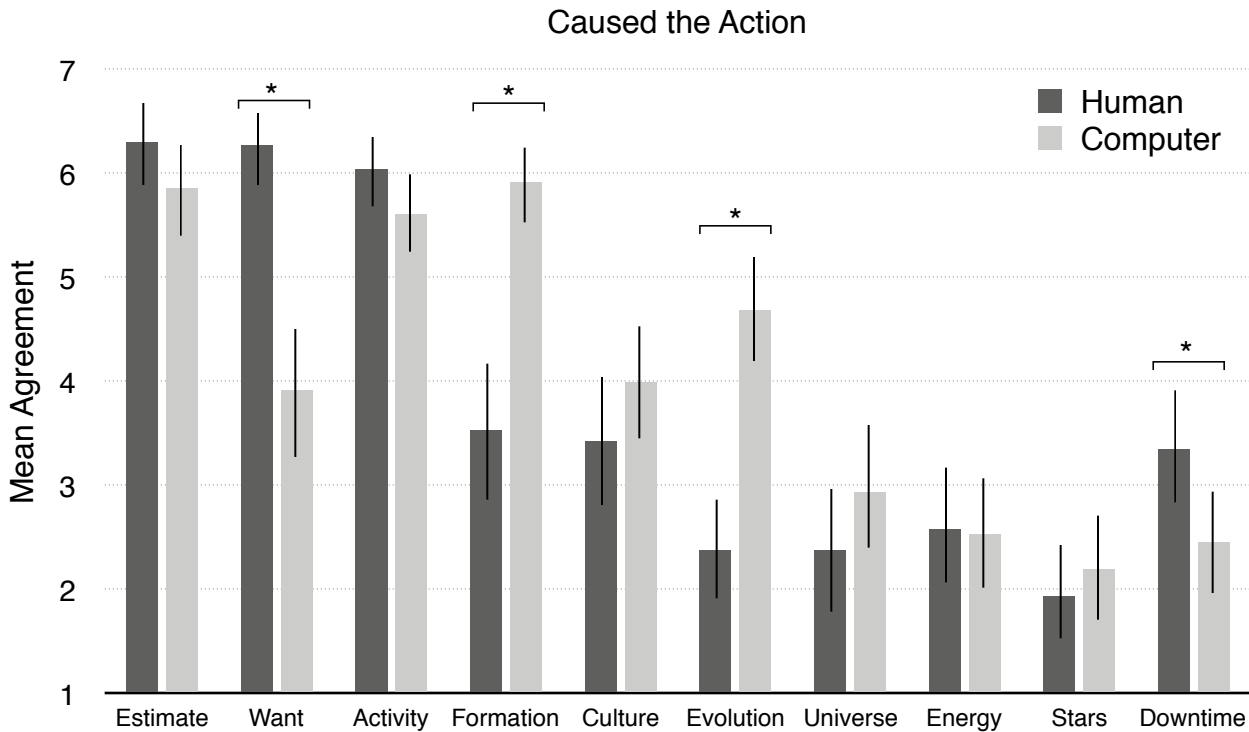
Figure 6. Experiment 6. Mean attribution of causation for the agent's action. The scale ran 1 (SD) - 7 (SA). Error bars represent bootstrapped 95% confidence intervals. Asterisks (*) represent a significant difference in mean response across conditions, as determined by an independent samples t-test (p < .05).

Table 3. Experiment 6. One sample t-tests for the ten causal items in the two agent conditions. Test value = 4.

Measure	Human					Computer				
	M	SD	t	df	p	M	SD	t	df	p
estimate	6.32	1.25	11.78	39	<.001	5.88	1.47	8.06	39	<.001
want	6.27	1.15	12.46	39	<.001	3.93	2.06	-0.23	39	.819
activity	6.05	1.04	12.51	39	<.001	5.62	1.21	8.47	39	<.001
formation	3.52	2.00	-1.50	39	.141	5.93	1.23	9.92	39	<.001
culture	3.42	2.02	-1.80	39	.080	4.00	1.83	0.00	39	1
evolution	2.37	1.58	-6.50	39	<.001	4.70	1.64	2.71	39	.010
universe	2.37	2.06	-4.99	39	<.001	2.95	1.80	-3.70	39	.001

Measure	Human					Computer				
	M	SD	t	df	p	M	SD	t	df	p
energy	2.60	1.74	-5.10	39	<.001	2.53	1.65	-5.66	39	<.001
stars	1.95	1.57	-8.27	39	<.001	2.20	1.68	-6.77	39	<.001
downtime	3.37	1.75	-2.26	39	.030	2.45	1.55	-6.32	39	<.001
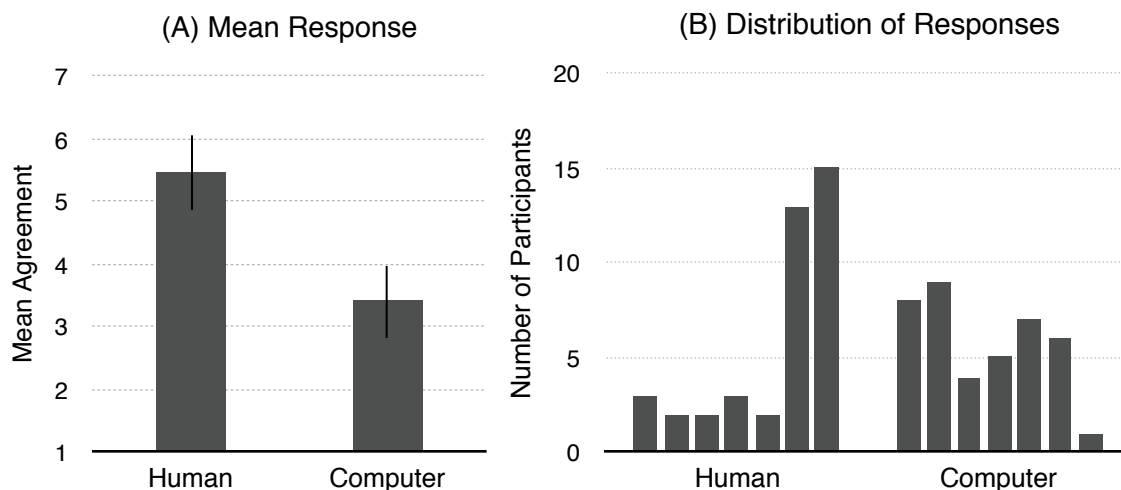
Fig. 7. Experiment 6. Mean response to the test statement about whether the agent could have done otherwise. The scale ran 1 (SD) - 7 (SA). Error bars represent bootstrapped 95% confidence intervals.

Discussion

The results from this experiment support both halves of exceptionalist naturalism in the same context. Replicating a principal result from Experiments 1-4, human agents were viewed differently from computers. Replicating a principal result from Experiment 5, human action was viewed as caused by psychological and neurological events.

Experiment 7

Exceptionalist naturalism says that human agency is viewed as fitting broadly within the causal order while still being exceptional in some respects. The results up until now strongly support this view. However, a potential concern about the results from Experiments 5 and 6 is that people did not have sufficient flexibility to evaluate the relationship between human action and natural events. (Joshua Knobe helpfully raised this concern without necessarily endorsing it.) Participants might have agreed with "causal" claims in order to convey that the factors help *explain* the action, albeit in a non-causal way. In order to evaluate this concern, I conducted a final experiment. I offered participants the option to describe an action as "caused" or "explained" by a natural event. Participants chose between these descriptions for an action performed by a human or a computer. If the concern is well-founded, then participants will describe the human's action as "explained" by the event and be more likely to select "caused" for the computer than for the human.

Method

Participants

One hundred and one new participants were tested (aged 18-72 years; mean age = 32 years; 42 female; 97% reporting English as a native language). They were compensated $0.40 for approximately 2 minutes of their time.

Materials and Procedure

Participants were randomly assigned to one of two conditions (human, computer) in a between-subjects design. Participants first read a brief story:

> Bradley has a [job/computer.] [His/Its] job is to place bets on horse races. Late this morning, the horserace association made public the news that the horse Pepper Run has a strained calf muscle. Early this afternoon, [Bradley/the computer] learned this news and then bet that Pepper Run would lose the race.

Participants were then instructed, "Please select the option that better describes the case." The options (rotated randomly) were,

1. The news report caused [Bradley/the computer] to bet that way. (caused)

2. The news report explains why [Bradley/the computer] bet that way. (explained)

Participants then went to a new screen and rated their agreement with these four statements (order rotated randomly), using the same 7-point Likert scale as in earlier experiments:

1. The news report caused [Bradley/the computer] to bet that way. (news cause)

2. The news report explains why [Bradley/the computer] bet that way. (news explain)

3. [Bradley's/The computer's] bet caused the news report. (bet cause)

4. [Bradley's/The computer's] bet explains the news report. (bet explain)

Participants never saw numerical labels.

Results

Assignment to condition did not affect the rate at which participants answered that the news report caused the bet, Fisher's exact test, p = .267. (See Figure 8.) Participants selected "caused" at

rates exceeding chance (50%) for both the computer (68%), binomial test, n = 50, p = .015, and the human (78%), binomial test, n = 51, p < .001. Assignment to condition did not affect response to any of the four causal/explanatory statements, independent samples t-tests, all ps ≥ . 220. Mean agreement was significantly above the midpoint (= 4) for the statements that the news report caused the computer's bet (M = 5.62, SD = 1.91), t(49) = 6.01, p < .001, and the human's bet (M = 5.90, SD = 1.39), t(50) = 9.78, p < .001.
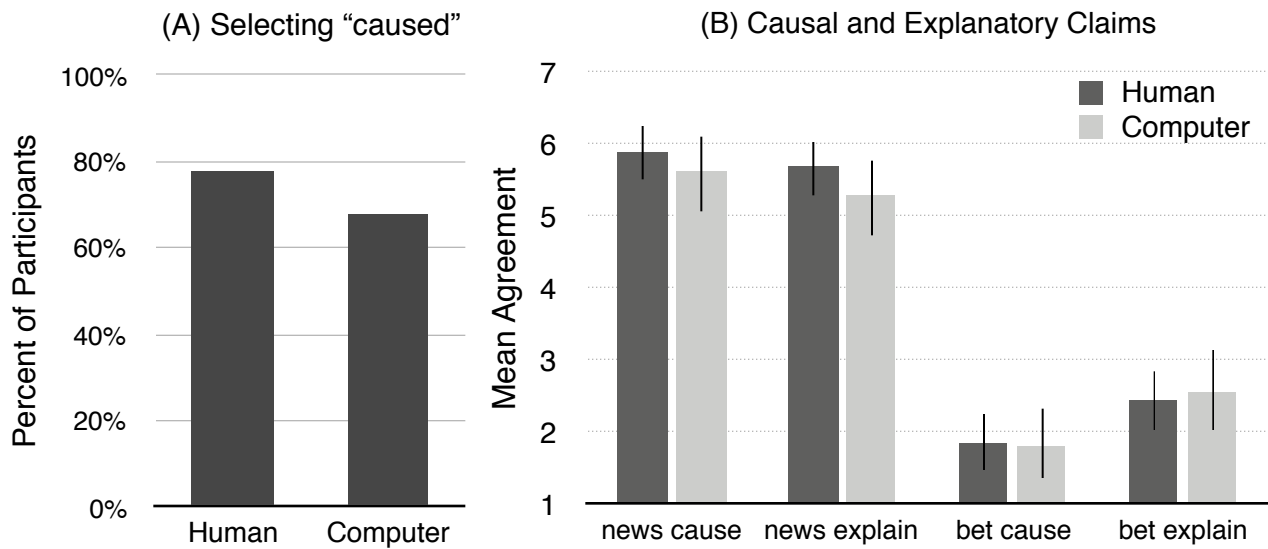
Figure 8. Experiment 7. Percent of participants answering that the news report "caused," rather than "explained," the bet (Panel A). Mean agreement with four causal/explanatory claims (Panel B). Scales ran 1 (SD) - 7 (SA). Error bars represent bootstrapped 95% confidence intervals.

Discussion

This experiment tested the concern that in earlier experiments participants agreed that natural events "cause" human actions only because they lacked an alternative way to convey that the natural events were explanatorily relevant. The results rule out this concern. When given the op-

tion to describe human action as "caused" or "explained" by a natural event, the vast majority selected "caused." And participants selected "caused" at similar rates regardless of whether the action was performed by a human or a computer.

General Discussion

In ordinary social cognition, is human agency viewed as radically different from other things in the world, standing outside of the causal order (a transcendence account)? Is it viewed as a purely natural thing, subject to the same physical laws as other objects and completely open to scientific investigation (a naturalistic account)? Or is it viewed as something in between, fitting broadly within the causal order while still being exceptional in some respects (exceptionalist naturalism)?

The results from seven experiments advance our understanding of how human agency is ordinarily viewed by supporting a version of this last possibility, exceptionalist naturalism. Across a variety of contexts and actions, participants agreed that human agents can resist outcomes described as inevitable, guaranteed by physical processes, and causally determined. That is, participants judged that humans can still do, could still do, or could have done otherwise. Participants also viewed some non-human animal agents (dolphins, raccoons, crows, cats) similarly, whereas they viewed computers, robots, plants, and simple inanimate objects (tennis balls, solar panels) differently (Experiments 1-4). For instance, when a ball is causally determined to fall, participants judged that it could not avoid falling; but when a human or raccoon is causally de-

termined to fall, participants judged that they could still avoid falling (Experiment 3). At the same time, participants judged that human actions are caused by many things, especially mental states and brain activity, and in some cases social and cultural processes (Experiments 5 and 6). Participants judged that human actions are not merely *explained* but *caused* by natural events (Experiment 7). Overall, then, human agents are viewed as exceptional parts of the natural world.

The results rule out one initial hypothesis to explain the basic finding supporting exceptionalist naturalism. The hypothesis is that people are more likely to spontaneously consider the complexity of causal relations in scenarios involving humans. This could lead people to more easily imagine ways that an inevitable outcome would not occur when a human is involved. However, the similar findings for non-human animals rule out this hypothesis.

It is unlikely that the basic findings are due to peculiarities of the stimuli used here, because I used a diverse range of stimuli that varied the type of activity, entity, and wording of test items. The activities included placing a bet, making a donation, falling, and inspecting a stimulus. The entities included a human, dolphin, raccoon, crow, cat, frog, grasshopper, plant, solar panel, ball, and robot. Outcomes were described as "inevitable," "guaranteed" by physical processes, and "causally determined." Some test items were prospective and others retrospective; some test items were phrased subjunctively and others indicatively. Of course, exceptionalist naturalism does not imply that human agents are viewed as exceptional with respect to all scenarios, abilities, and modes of questioning. Nevertheless, the same basic pattern emerged across several simple and natural ways of framing scenarios and questions.

It might be wondered whether the basic findings are due to people rejecting one or more

36

assumptions of the scenarios described. For example, the findings might somehow be due to people rejecting that the outcome is causally determined. This hypothesis faces two problems. On the one hand, it cannot explain why people would do this for scenarios involving living agents but not mechanical and inanimate objects. If the hypothesis is amended so that people reject the assumption specifically for scenarios involving living agents but not others, then it becomes hard to distinguish the amended hypothesis from exceptionalist naturalism. That is, the amended hypothesis starts to sound like just another way of saying that agents are viewed as exceptional — more resistant to causal determination, more capable of avoiding inevitable outcomes. On the other hand, a pattern supporting exceptionalist naturalism emerged even when people overwhelmingly agreed that the outcome is causally determined (Experiments 3 and 4), and even among those who rated the probability of the outcome's occurrence at "100%" (Experiments 1 and 2). The fact that this same pattern occurs when participants themselves rate the outcome as 100% probable addresses the potential concern that the results are due simply to ordinary people ignoring or not understanding "causal determinism" jargon.

Some of the findings suggest that many people's naive understanding of psychological and physical processes is naturally indeterministic. (See Rose & Nichols 2014 for related evidence on adult judgments about decision-making specifically.) Even when an outcome is described as "causally determined" and "guaranteed" by physical processes, many people still think there is a chance that it will not happen. When asked to rate the probability of an inevitable or causally determined outcome, people often rated it between 85% and 90% (Experiments 1 and 2). This was true for the activities of humans and computers.

An interesting question is why we would be natural indeterminists. One possibility is that indeterminism is an adaptive assumption, partly constitutive of our natural aptitude for construing social partners and other animals, which, as suggested by recent work in theoretical biology (Brembs 2011), are inherently stochastic systems. This assumption is then extended, perhaps mistakenly, to more mundane physical systems that are, for all practical purposes, deterministic. In other words, the extension of indeterminism beyond animals could be an example of "explicitly adopting a social model" of non-social physical systems, as happens when people try to "bargain with nature" (Humphrey 1976, p. 313). Alternatively, it could be that a general assumption of indeterminism is adaptive because it prevents us from smugly assuming that we have everything figured out in advance. Even if determinism is true, in practice we rarely, if ever, have access to all the information relevant to predicting predetermined events. If the cost of falsely assuming that we have all the information tends be high enough, and we are bad enough at judging that we have all the information, then a simple way to prevent such assumptions would be to presuppose indeterminism in general. In short, presupposing indeterminism could be a false but effective solution to potentially problematic overconfidence.

I found no evidence that human agency is viewed as essentially different from non-human animal agency, whereas other research has found differences which might suggest otherwise (Gray, Gray & Wegner 2007). Future research could investigate whether differences emerge with other stimuli or tasks. On the one hand, perhaps generic judgments about human and non-human animals will differ regarding the ability to resist causally determined outcomes. That is, when the question is posed more generally than I posed it here, differences might emerge in the perception

of human and non-human animals. Relatedly, people might perceive humans as more capable of resisting harmful outcomes. On the other hand, perhaps such differences will emerge for more complex or protracted tasks. For example, people might perceive humans as more capable of acting otherwise when the task involves repeated or multiple steps, such as digging a deep hole or building a dam, rather than falling or turning towards a stimulus.

Even if humans are viewed differently from all non-human animals in some respects, the fact that participants in the present experiments viewed them similarly should not be surprising. After all, belief in the transmigration and reincarnation of persons occurs in human cultures around the world (Obeyesekere 2002; Bjorling 2013; Nagaraj, Nanjegowda & Purushothama 2013). People in many cultural traditions have claimed that when a human dies, the person can be reincarnated as a member of another species and eventually as a human again. These beliefs would probably not be so widespread if people naturally viewed humans and non-humans as fundamentally different.

References

Bering, J. M. (2006). The folk psychology of souls. Behavioral and Brain Sciences, 29, 453–498.

Bjorling, J. (2013). Reincarnation: a bibliography. New York: Routledge.

Bloom, P. (2005). Descartes' baby: how the science of child development explains what makes us human. New York: Basic Books.

Bloom, P. (2012). Free will does not exist. So what? The Chronicle of Higher Education. Retrieved November 28, 2014, from http://chronicle.com/article/Paul-Bloom/131170/

Brembs, B. (2011). Towards a scientific concept of free will as a biological trait: spontaneous actions and decision-making in invertebrates. Proceedings of the Royal Society B: Biological Sciences, 278(1707), 930–939. http://doi.org/10.1111/j.1467-9280.2008.02045.x

Gabrieli, J. D. E., Ghosh, S. S., & Whitfield-Gabrieli, S. (2015). Prediction as a Humanitarian and Pragmatic Contribution from Human Cognitive Neuroscience. Neuron, 85(1), 11–26. doi:10.1016/j.neuron.2014.10.047

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

Haggard, P. (2005). Conscious intention and motor cognition. Trends in Cognitive Sciences, 9(6), 290–295. doi:10.1016/j.tics.2005.04.012

Humphrey, N. (1976). The social function of intellect. In P. P. G. Bateson & R. A. Hinde (Eds.), Growing points in ethology (pp. 303–317). Cambridge: Cambridge University Press.

Keller, I., & Heckhausen, H. (1990). Readiness potentials preceding spontaneous motor acts:

voluntary vs. involuntary control. Electroencephalography and Clinical Neurophysiology, 76(4), 351–361.

Knobe, J. (2014). Free will and the scientific vision. In E. Machery & E. O'neill (Eds.), Current controversies in experimental philosophy (pp. 69–85). New York: Routledge.

Kuhlmeier, V. A., Bloom, P., & Wynn, K. (2004). Do 5-month-old infants see humans as material objects? Cognition, 94(1), 95–103. doi:10.1016/j.cognition.2004.02.007

Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). Brain, 106(3), 623–642.

McCloskey, M., Washburn, A., & Felch, L. (1983). Intuitive physics: The straight-down belief and its origin. Journal of Experimental Psychology: Learning, Memory, and Cognition, 9(4), 636. doi:10.1037/0278-7393.9.4.636

Nagaraj, A. M., Nanjegowda, R., & Purushothama, S. M. (2013). The mystery of reincarnation. Indian Journal of Psychiatry, 55(6), 171. doi:10.4103/0019-5545.105519

Nahmias, E., & Thompson, M. (2014). A naturalistic vision of free will. In E. Machery & E. O'neill (Eds.), Current controversies in experimental philosophy (pp. 86–103). Routledge.

Nahmias, E., Shepard, J., & Reuter, S. (2014). It's OK if 'my brain made me do it': People's intuitions about free will and neuroscientific prediction. Cognition, 133(2), 502–516. doi: 10.1016/j.cognition.2014.07.009

Nichols, S., & Knobe, J. (2007). Moral responsibility and determinism: the cognitive science of folk intuitions. Nous, 41(4), 663–685. doi:10.1111/j.1468-0068.2007.00666.x

Obeyesekere, G. (2002). Imagining karma. Berkeley: University of California Press.

Rose, D., & Nichols, S. (2013). The lesson of bypassing. Review of Philosophy and Psychology, 4, 599–619.

Rose, D., Buckwalter, W., & Nichols, S. (2015). Neuroscientific prediction and the intrusion of intuitive metaphysics. Cognitive Science, n/a–n/a. http://doi.org/10.1111/cogs.12310

Sarkissian, H., Chatterjee, A., De Brigard, F., Knobe, J., Nichols, S., & Sirker, S. (2010). Is belief in free will a cultural universal?, 25(3), 346–358.

Saxe, R., Tzelnic, T., & Carey, S. (2006). Five-month-old infants know humans are solid, like inanimate objects. Cognition, 101(1), B1–B8. doi:10.1016/j.cognition.2005.10.005

Sellars, W. (1963). Science, perception and reality. Atascadero, CA: Ridgeview Publishing Company.

Soon, C. S., He, A. H., Bode, S., & Haynes, J.-D. (2013). Predicting free choices for abstract intentions. Proceedings of the National Academy of Sciences, 110(15), 6217–6222. doi: 10.1073/pnas.1212218110/-/DCSupplemental/pnas.201212218SI.pdf