

Compatibilism and Incompatibilism in Social Cognition*

John Turri^a

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

Abstract: Compatibilism is the view that determinism is compatible with acting freely and being morally responsible. Incompatibilism is the opposite view. It is often claimed that compatibilism or incompatibilism is a natural part of ordinary social cognition. That is, it is often claimed that patterns in our everyday social judgments reveal an implicit commitment to either compatibilism or incompatibilism. This paper reports five experiments designed to identify such patterns. The results support a nuanced hybrid account: the central tendencies in ordinary social cognition are compatibilism about moral responsibility, compatibilism about positive moral accountability (i.e. about deserving credit for good outcomes), neither compatibilism nor incompatibilism about negative moral accountability (i.e. about deserving blame for bad outcomes), compatibilism about choice for actions with positive outcomes, and incompatibilism about choice for actions with negative or neutral outcomes.

Keywords: determinism; responsibility; choice; blame; credit; folk metaphysics; agency; moral psychology

Word count: 5748

* This is the penultimate version of a paper to appear in *Cognitive Science*. Please cite the final, published version if possible.

Introduction

Suppose that someone is in a situation where she can do only one thing — it is literally impossible for her to do anything else. Unsurprisingly, she behaves in the only way available to her. In such a situation, does she choose to behave as she does? Is she morally responsible for her behavior? Can she be properly blamed or credited for how she behaves?

Compatibilism is the view that determinism is compatible with acting freely and being morally responsible. Philosophers and psychologists have long debated the merits of compatibilism and its denial, incompatibilism (for reviews, see McKenna 2009, O'Connor 2010, Vihvelin 2011, Clarke & Capes 2013; see also Dennett 1984). The theoretical debates surrounding these issues can become extremely complicated, but one aspect of the debate has remained firmly rooted in commonsense, and it is this that I focus on here. It is often claimed that compatibilism or incompatibilism is a natural part of ordinary social cognition (e.g. Hume 1748/1993; Reid 1785; Kane 1999, p. 217; Pereboom 2001, p. xvi; Roskies & Nichols 2008; Rose & Nichols 2013). That is, it is often claimed that our commonsense moral psychology is implicitly committed to one view or the other, as revealed by patterns in people's judgments about particular cases. The theoretical debate has often assumed that the burden of proof rests with the side that contradicts commonsense (for a review, see Nahmias, Morris, Nadelhoffer & Turner 2006). For example, if ordinary moral psychology assumes that compatibilism is true, then incompatibilists will need stronger arguments to persuade us that their position is correct.

Several recent experimental studies have begun examining the claim that people are natu-

rally or pre-theoretically committed to either compatibilism or incompatibilism (e.g. Nahmias, Morris, Nadelhoffer & Turner 2005; Woolfolk, Doris & Darley 2006; Nichols & Knobe 2007; Sarkissian, Chatterjee, De Brigard, Knobe, Nichols & Sirker 2010; Schulz, Cokely & Feltz 2011; Nahmias, Shepard & Reuter 2014; Cova & Kitano 2014; May 2014; Rose, Buckwalter & Nichols under review; Turri under review a). The results have been mixed, with some suggesting that people are natural compatibilists and some suggesting that they are natural incompatibilists (for reviews, see Sommers 2010 and Nichols 2011).

Hybrid views are also possible. For example, one might propose that determinism is consistent with moral responsibility, blame, or credit even if it is not compatible with acting freely (compare Clarke 2003, Fischer 1994, Fischer & Ravizza 1998). If this hybrid position is true, then there can be situations where an agent is determined to behave in only one way, she does not freely choose her behavior, but she is morally responsible, blameworthy, or creditable for it.

In this paper, I report five experiments that advance our understanding of these issues. The guiding question is whether patterns in our everyday social judgments reveal an implicit commitment to compatibilism, incompatibilism, or a hybrid account. Experiment 1 tests people's judgments about choice and free choice. Experiment 2 tests judgments about moral responsibility and compares them to judgments about choice. Experiments 3 and 4 test judgments about blame and replicate the earlier findings on choice and moral responsibility. Experiment 5 tests judgments about credit for desirable outcomes and compares them to judgments about blame for undesirable outcomes. Overall, the results suggest a nuanced hybrid account: the central tendencies in ordinary social cognition are compatibilism about moral responsibility, compatibilism about

positive moral accountability (i.e. deserving credit for good outcomes), neither compatibilism nor incompatibilism about negative moral accountability (i.e. deserving blame for bad outcomes), compatibilism about choice for actions with positive outcomes, and incompatibilism about choice for actions with negative or neutral outcomes.

Experiment 1

This experiment tests people's judgments about choice and free choice in contexts where only one behavior is possible. I focused on choice because recent work suggests that the ordinary concept of free will is tantamount to free choice (Monroe & Malle 2010; see also Monroe, Dillon & Malle 2014). This led me to hypothesize that in ordinary social cognition, people might not distinguish between choosing and choosing freely. In order to determine whether findings reflect people's views about agency in general or human agency in particular, I compared people's responses to cases involving human and non-human agents.

Method

Participants

Four hundred participants were tested (aged 18-70 years, mean age = 31 years; 156 female; 96% reporting English as a native language). Participants were U.S. residents, recruited and tested online using Amazon Mechanical Turk and Qualtrics, and compensated \$0.35 for approximately 2 minutes of their time. The same recruitment and compensation procedures were used for all ex-

periments reported in this paper. Repeat participation was prevented within and across experiments.

Materials and Procedure

Participants were randomly assigned to one of eight conditions in a 2 (agent: human, crow) \times 2 (status: impossible, possible) \times 2 (choice type: plain choice, free choice) between-subjects design. All participants read a simple story, responded to three test items, then completed a brief demographic questionnaire.

The story focused on a very simple event: a marble being rolled across a table. The agent factor manipulated whether a human or crow rolled the marble to the left. The status factor manipulated whether it was either “literally impossible” or “equally possible” for the agent to roll it another way. The choice-type factor manipulated one aspect of one test statement: whether the agent “chose” (plain choice) or “freely chose” (free choice). Here is the text of the story for the human condition (status manipulation in brackets):

In just a moment, a man will roll a marble across a table. Given the current conditions of the man’s brain, [the only/one] possibility is for him to roll it to the left. As a matter of brain chemistry, it is [literally impossible/equally possible] for him to roll it a different way. The man rolls it to the left.

The story for the crow condition was exactly the same except that “crow” replaced “man” throughout.

After reading the story, participants responded to three test items while the story remained

at the top of the screen. They first responded to two statements in a matrix table, the order of which was randomized:

1. The [man/crow] could have rolled it a different way. (could)
2. The [man/crow] [chose/freely chose] to roll it to the left. (choice)

Response to these items was collected on a standard 7-point Likert scale, 1 (“strongly disagree”) – 7 (“strongly agree”), left-to-right across the participant’s screen. Participants then proceeded to a new screen and completed a percentage task:

3. On a scale of 0% to 100%, how likely was it that the [man/crow] would roll it to the left? (percent)

Response to this item was collected in a text box directly below the question. After testing, participants completed a brief demographic questionnaire.

Results

A multivariate analysis of variance showed that response to the dependent variables was affected by neither agent nor choice type. Response to all three variables was affected by status. There were no interactions. (See Table 1.) Follow-up independent samples t-tests (collapsing across the agent and choice-type factors) showed that in the impossible condition, mean response was lower for the could statement, lower for the choice attribution, and higher for the percentage task. (See Fig. 1 and Table 2.) One sample t-tests showed that mean response to the could statement and choice attribution was below the neutral midpoint (= 4) in the impossible condition and above the midpoint in the possible condition. (See Table 3.) In the impossible condition, modal re-

sponse to both items was “strongly disagree.” In the possible condition, modal response was “strongly agree” to the could statement and “agree” to the choice attribution.

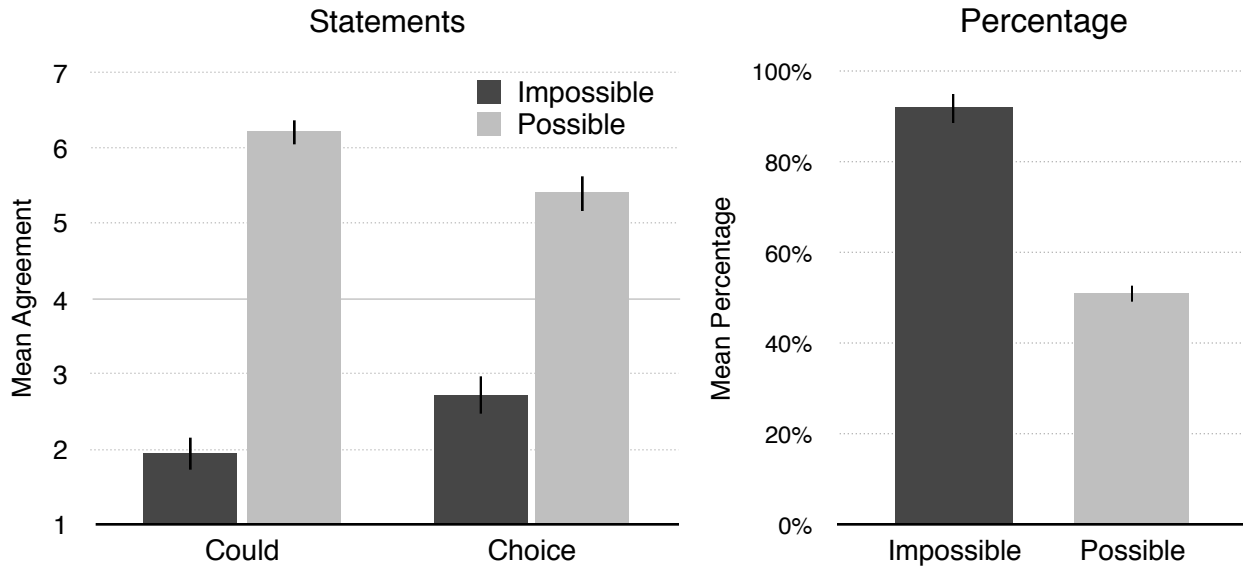


Fig. 1. Experiment 1. Panel A: mean response to the test statements about whether the agent could have done otherwise, and whether the agent chose to perform the action. Panel B: mean estimate of the percentage chance that the agent would perform the action. For Panel A, the scale ran 1 (SD) - 7 (SA) and the midpoint (= 4) is marked by a solid line. For Panel B, the scale ran 0-100%. Error bars represent bootstrapped 95% confidence intervals

Table 1. Experiment 1. Multivariate analysis of variance.

Factor	Dependent Variable											
	Could				Choice				Percent			
	F	df	p	η_p^2	F	df	p	η_p^2	F	df	p	η_p^2
Agent	1.92	1, 392	.166	.005	1.23	1, 392	.268	.003	0.04	1, 392	.851	<.001
Status	1080.45	1, 392	<.001	.734	240.43	1, 392	<.001	.380	482.36	1, 392	<.001	.552
Choice	1.06	1, 392	.304	.003	1.21	1, 392	.272	.003	0.01	1, 392	.938	<.001
A*S	1.03	1, 392	.312	.003	2.06	1, 392	.152	.005	0.01	1, 392	.970	<.001
A*C	0.73	1, 392	.394	.002	0.01	1, 392	.931	<.001	0.20	1, 392	.659	<.001
S*C	0.25	1, 392	.696	.001	0.98	1, 392	.323	.002	0.51	1, 392	.475	.001
A*S*C	1.42	1, 392	.235	.001	0.01	1, 392	.936	<.001	0.02	1, 392	.889	<.001

Table 2. Experiment 1. Independent samples t-tests.

Measure	Impossible		Possible		t	df	p	MD	95% CI	d
	M	SD	M	SD						
could	1.94	1.49	6.21	1.06	-33.06	367.55	<.001	-4.27	-4.52, -4.01	3.45
chose	2.71	1.84	5.40	1.60	-15.60	393.82	<.001	-2.69	-3.03, -2.35	1.57
percent	92.18	22.17	51.16	13.82	22.30	342.19	<.001	41.02	37.40, 44.64	2.41

Table 3. Experiment 1. One sample t-tests. Test value = 4.

Measure	Impossible						Possible					
	t	df	p	MD	95% CI	d	t	df	p	MD	95% CI	d
could	-19.72	203	<.001	-2.06	-2.26, -1.85	1.38	29.09	195	<.001	2.21	2.06, 2.36	2.08
chose	-10.01	203	<.001	-1.29	-1.55, -1.04	0.47	12.24	195	<.001	1.40	1.17, 1.62	0.88

Discussion

The results support incompatibilism about choice. When only one behavior was possible, people denied that the agent chose to behave as he did. In a closely matched control where more than one behavior was possible, people agreed that the agent chose to behave as he did. The same pattern was observed for both *choice* and *free choice*, suggesting that in ordinary social cognition, people do not readily distinguish between choosing and choosing freely. This in turn suggests an explanation for why people do not attribute free choice in situations where only one behavior is possible: people do not attribute free choice because they do not attribute choice. Moreover, the same pattern was observed for both a human and non-human agent (a crow), suggesting that the findings reflect people's conception of agency and choice in general, not just human agency and choice specifically.

Experiment 2

This experiment tests people's judgements about moral responsibility and compares them to judgments about choice. Having observed a pattern consistent with incompatibilism about choice in Experiment 1, the main question is whether we would see a similar pattern for moral responsibility.

Method

Participants

Two hundred and three new participants were tested (aged 18-77 years, mean age = 32 years; 70 female; 98% reporting English as a native language).

Materials and Procedure

Participants were randomly assigned to one of four conditions in a 2 (status: impossible, possible) \times 2 (attribution type: choice, responsibility) between-subjects design. The basic procedures were the same as in Experiment 1. The story focused on a woman evaluating an employee who has performed excellently. The status factor manipulated whether it was either "literally impossible" or "possible" for the woman to give the employee a good evaluation. The attribution factor manipulated one aspect of one test statement: whether the agent "chose not to" or "had a moral responsibility to" give the employee a good evaluation. Here is the text of the story and the test items (status manipulation in brackets):

A woman is evaluating her employee's performance. The employee performed excellently. Given the current condition of the woman's brain, it is [impossible/possible] for her to give the employee a good evaluation. As a matter of brain chemistry, it is [literally impossible/possible] that she can give the employee a good evaluation. She does not give the employee a good evaluation.

1. The woman could have given the employee a good evaluation. (could)
2. The woman [chose not/had a moral responsibility] to give the employee a good evaluation. (attribution)
3. On a scale of 0% to 100%, how likely was it that the woman would give the employee a good evaluation? (percent)

Results

A preliminary multivariate analysis of variance revealed main effects on all the dependent variables and a large interaction effect on attributions. (See Fig. 2 and Table 4.) The interaction effect on attributions is critical. When it was possible for the woman to give a good evaluation, people strongly agreed that the woman had a moral responsibility to do so, and they also strongly agreed that she chose not to do so. By contrast, when it was impossible for the woman to give a good evaluation, people still agreed that the woman had a moral responsibility to do so, but they denied that she chose not to do so. Table 5 reports pairwise comparisons of mean scores on the three dependent measures for the impossible and possible conditions, grouped by attribution type. Table 6 reports one sample t-tests for the could statement and the attribution in all condi-

tions.

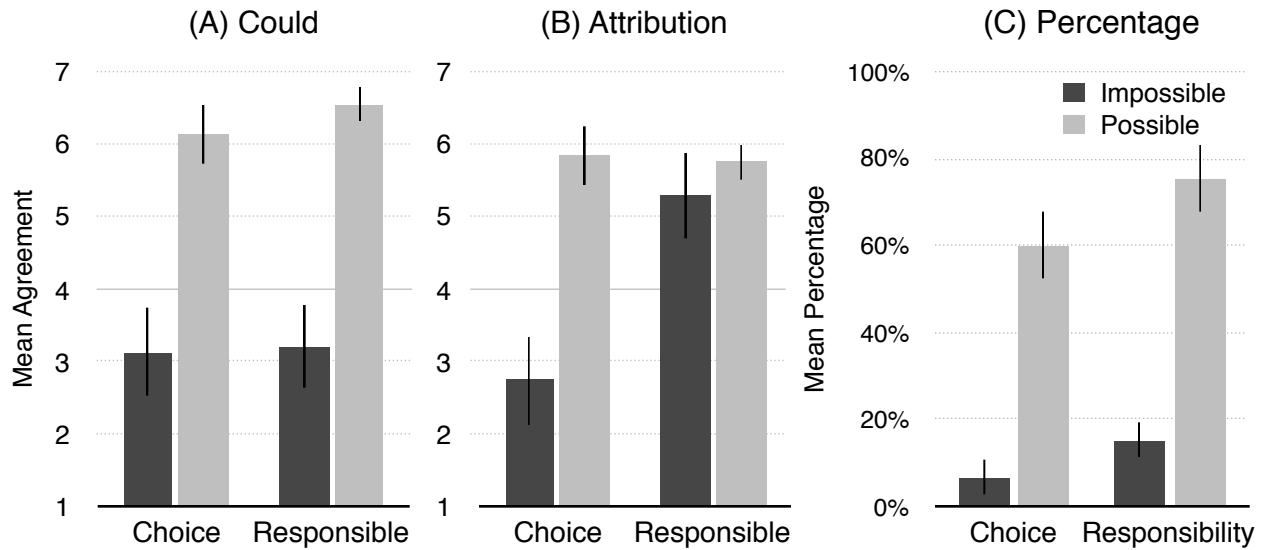


Fig. 2. Experiment 2. Panel A: mean response to the test statements about whether the agent could have performed the action. Panel B: mean agreement that the agent either chose not to perform the action, or had a moral responsibility to perform the action. Panel C: mean estimate of the percentage chance that the agent would perform the action. For Panels A and B, the scale ran 1 (SD) - 7 (SA) and the midpoint (= 4) is marked by a solid line. For Panel C, the scale ran 0-100%. Error bars represent bootstrapped 95% confidence intervals.

Table 4. Experiment 2. Multivariate analysis of variance.

Factor	Dependent Variable											
	Could				Attribution				Percent			
	F	df	p	η_p^2	F	df	p	η_p^2	F	df	p	η_p^2
Status	156.10	1, 199	<.001	.440	54.66	1, 199	<.001	.215	279.36	1, 199	<.001	.584
Attribution	1.00	1, 199	.320	.005	26.31	1, 199	<.001	.117	13.53	1, 199	<.001	.064
S*A	0.48	1, 199	.491	.002	30.50	1, 199	<.001	.133	0.84	1, 199	.360	.004

Table 5. Experiment 2. Independent samples t-tests.

Measure	Impossible		Possible		t	df	p	MD	95% CI	d
	M	SD	M	SD						
Chose										

Measure	Impossible		Possible		t	df	p	MD	95% CI	d
	M	SD	M	SD						
could	3.12	2.41	6.12	1.39	-7.69	80.50	<.001	-3.00	-3.78, -2.23	1.71
attribution	2.73	1.87	5.84	1.53	-9.18	96.01	<.001	-3.12	-3.79, -2.44	1.87
percent	6.37	15.14	59.70	29.16	-11.49	73.29	<.001	-53.29	-62.53, -44.05	2.68
Responsible										
could	3.20	2.18	6.55	0.78	-10.33	62.66	<.001	-3.35	-4.00, -2.70	2.61
attribution	5.29	1.85	5.75	1.60	-1.32	100	.190	-0.45	-1.13, 0.23	0.26
percent	15.70	29.47	75.16	19.16	-12.08	100	<.001	-59.47	-69.24, -49.71	2.42

Table 6. Experiment 2. One sample t-tests. Test value = 4.

Measure	Impossible						Possible					
	t	df	p	MD	95% CI	d	t	df	p	MD	95% CI	d
Chose												
could	-2.60	50	.012	-0.88	-1.56, -0.21	0.37	10.75	49	<.001	2.12	1.72, 2.52	1.53
attribution	-4.88	50	<.001	-1.28	-1.80, -0.75	0.68	8.50	49	<.001	1.84	1.41, 2.27	1.20
Responsible												
could	-2.63	50	.011	-0.80	-1.42, -0.19	0.37	23.26	50	<.001	2.55	2.33, 2.77	3.26
attribution	5.00	50	<.001	1.29	0.77, 1.81	0.70	7.80	50	<.001	1.75	1.30, 2.19	1.09

In the impossible condition, mean responsibility attribution was higher than choice attribution, $t(100) = 6.99$, $p < .001$. The size of the mean difference was very large, MD = 2.57 [1.84, 3.30], $d = 1.40$. In the impossible condition, mean attribution was above the neutral midpoint (= 4) for responsibility, $t(50) = 5.00$, $p < .001$, MD = 1.29 [0.77, 1.81], $d = 0.70$, and below the midpoint for choice, $t(50) = -4.88$, $p < .001$, MD = -1.28 [-1.80, -0.75], $d = 0.68$. Modal response was “strongly agree” for the responsibility attribution and “disagree” for the choice attribution.

The same basic pattern emerges if we include data from only those participants in the impossible conditions who answered “0%” on the percentage task. Among these participants, mean response to the could statement did not differ between the choice and responsibility conditions ($N = 37/27$, $M = 2.49/2.26$, $SD = 2.17, 1.97$, $t(62) = -0.43$, $p = .669$, n.s. Mean responsibility attribution was higher than mean choice attribution ($M = 5.30/2.32$, $SD = 1.96/1.68$), $t(62) = 6.51$, $p < .001$, $MD = 2.94 [2.06, 3.39]$, $d = 1.65$. Mean attribution was above the neutral midpoint for responsibility, $t(26) = 3.44$, $p = .002$, $MD = 1.30 [0.52, 2.07]$, $d = 0.66$, and below the midpoint for choice, $t(36) = -6.05$, $p < .001$, $MD = -1.68 [-2.24, -1.11]$, $d = 1.00$. Modal response was “strongly agree” for the responsibility attribution and “disagree” for the choice attribution.

Discussion

The results support compatibilism about moral responsibility. People judged that an agent had a moral responsibility to perform an action that was impossible for her to perform. The rate of responsibility attribution did not differ from that observed in a closely matched control where the action was possible. By contrast, when an action was impossible for the agent to perform, people denied that the agent chose to not perform the action. In a closely matched control where it was possible to perform the action, people agreed that the agent chose to not perform the action. These results show that the absence of alternative possibilities affects moral responsibility and choice differently. The results also replicate and extend the main finding from Experiment 1.

Experiment 3

This experiment tests people's judgments about blame in contexts where only one behavior is possible. The basic approach was the same as in the earlier experiments, but after participants were asked about blame they were asked follow-up questions about choice and responsibility too. I did this for two reasons. On the one hand, I wanted to see if the results from Experiments 1 and 2 would replicate within-subjects. On the other hand, I wanted to see which overall pattern would emerge when people judged blame, choice, and responsibility all in the same context.

Method

Participants

Two hundred new participants were tested (aged 18-75 years, mean age = 33 years; 84 female; 99% reporting English as a native language). Because participants in this experiment and the next two experiments answered more questions, I compensated them \$0.45 (compared to \$0.35 in Experiments 1 and 2).

Materials and Procedure

Participants were randomly assigned to one of two conditions (impossible, possible) in a between-subjects design. The basic procedures and story were the same as in Experiment 2. After reading the story, participants responded to two statements in a matrix table (order randomized):

1. The woman could have given the employee a good evaluation. (could)

2. The woman is blameworthy for not giving the employee a good evaluation. (blame)

Participants then advanced to a new screen and responded to two more statements in the same way:

3. The woman chose not to give the employee a good evaluation. (choice)
4. The woman had a moral responsibility to give the employee a good evaluation. (responsibility)

Participants then advanced to a new screen and completed the same percentage task as in Experiment 2:

5. On a scale of 0% to 100%, how likely was it that the woman would give the employee a good evaluation? (percent)

Results

The status manipulation was effective again and the basic pattern of response to the could statement and percentage task was the same as in Experiment 2. (See Fig. 3.) Independent samples t-tests showed that mean response was higher in the possible condition for all test items except for the responsibility attribution, which was the same across conditions. (See Table 7.) One sample t-tests showed that mean response to the could statement and choice attribution was below the neutral midpoint (= 4) in the impossible condition and above the midpoint in the possible condition. (See Table 8.) Mean blame attribution was above the midpoint in the possible condition but no different from the midpoint in the impossible condition. Mean responsibility attribution was above the midpoint in both conditions.

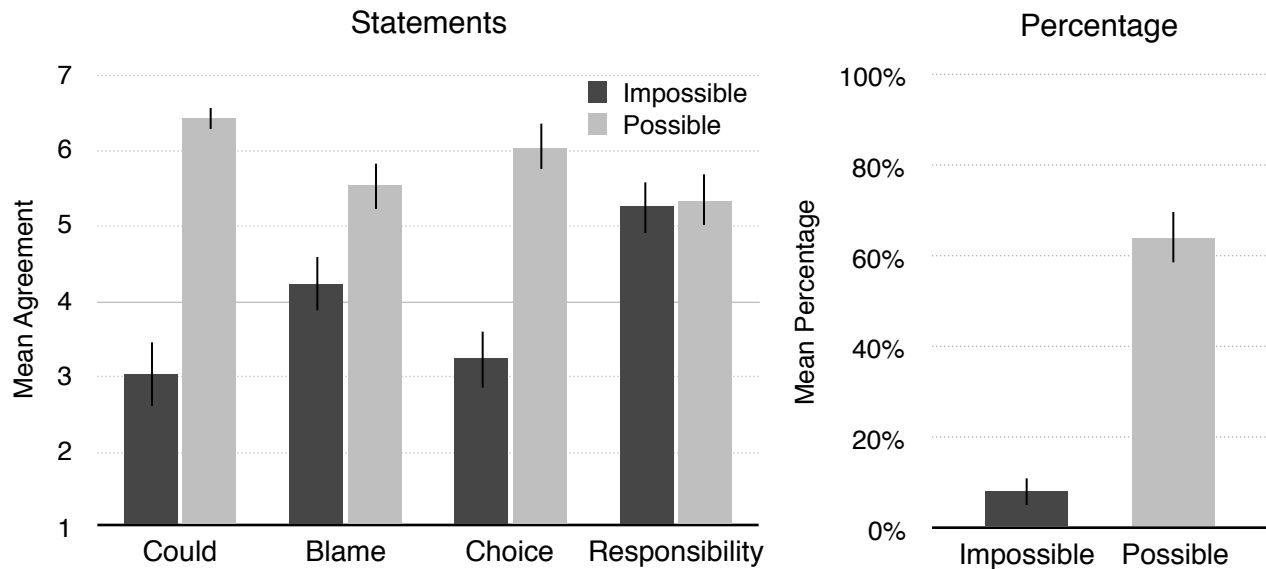


Fig. 3. Experiment 3. Panel A: mean response to the test statements about whether the agent could have performed the action, is blameworthy for not performing the action, chose not to perform the action, and had a moral responsibility to perform the action. Panel B: mean estimate of the percentage chance that the agent would perform the action. For Panel A, the scale ran 1 (SD) - 7 (SA) and the midpoint (= 4) is marked by a solid line. For Panel B, the scale ran 0-100%. Error bars represent bootstrapped 95% confidence intervals.

Table 7. Experiment 3. Independent samples t-tests.

Measure	Impossible		Possible		t	df	p	MD	95% CI	d
	M	SD	M	SD						
could	3.01	2.00	6.39	0.79	-15.74	192.23	<.001	-3.38	-3.81, -2.96	2.77
blame	4.21	1.83	5.51	1.55	-5.43	192.76	<.001	-1.30	-1.77, -0.83	0.78
chose	3.21	1.84	6.02	1.46	-11.96	187.91	<.001	-2.81	-3.27, -2.35	1.74
responsible	5.22	1.67	5.32	1.65	-0.43	198	.671	-0.10	-0.56, 0.36	0.06
percent	8.32	14.72	64.31	28.61	-17.40	148.01	<.001	-55.99	-62.34, -49.65	2.86

Table 8. Experiment 3. One sample t-tests. Test value = 4.

Measure	Impossible						Possible					
	t	df	p	MD	95% CI	d	t	df	p	MD	95% CI	d
could	-4.96	99	<.001	0.99	-1.39, -0.59	0.50	30.25	99	<.001	2.39	2.23, 2.55	3.03
blame	1.15	99	.253	0.21	-0.15, 0.57	0.11	9.76	99	<.001	1.51	1.20, 1.82	0.97
chose	-4.28	99	<.001	-0.79	-1.16, -0.42	0.43	13.87	99	<.001	2.02	1.73, 2.31	1.38

Measure	Impossible						Possible					
	t	df	p	MD	95% CI	d	t	df	p	MD	95% CI	d
responsible	7.29	99	<.001	1.22	0.89, 1.55	0.73	8.00	99	<.001	1.32	0.99, 1.65	0.80

Paired sample t-tests showed that in the impossible condition, mean blame attribution was higher than mean response to the could statement, $t(99) = 5.47$, $p < .001$, MD = 1.20 [0.77, 1.64], $d = 0.63$, and also higher than mean choice attribution, $t(99) = 5.03$, $p < .001$, MD = 1.00 [0.61, 1.40], $d = 0.55$. Mean response to the could statement did not differ from mean choice attribution, $t(99) = -1.11$, $p = .270$, n.s. Mean responsibility attribution was higher than mean blame attribution, $t(99) = 4.84$, $p < .001$, MD = 1.01 [0.60, 1.42], $d = 0.58$, higher than mean response to the could statement, $t(99) = 8.19$, $p < .001$, MD = 2.21 [1.68, 2.75], $d = 0.84$, and also higher than mean choice attribution, $t(99) = 8.63$, $p < .001$, MD = 2.01 [1.55, 2.47], $d = 1.08$.

Focusing on only those participants in the impossible condition who answered “0%” on the percentage task ($N = 53$), the same basic pattern emerges. One sample t-tests showed that mean response to the could statement was below the midpoint ($M = 2.40$, $SD = 1.79$), $t(52) = -6.52$, $p < .001$, MD = 1.60 [-2.10, 1.11], $d = 0.89$. Mean blame attribution did not differ from midpoint ($M = 4.11$, $SD = 1.93$), $t(52) = 0.43$, $p = .671$, n.s. Mean choice attribution was below the midpoint ($M = 2.79$, $SD = 1.83$), $t(52) = -4.80$, $p < .001$, MD = 1.21 [-1.71, =0.70], $d = 0.66$. Mean responsibility attribution was above the midpoint ($M = 5.28$, $SD = 1.82$), $t(52) = 5.13$, $p < .001$, MD = 1.28 [0.78, 1.79], $d = 0.70$. Paired sample t-tests showed that mean blame attribution was higher than mean choice attribution, $t(52) = 4.21$, $p < .001$, MD = 1.32 [0.69, 1.95], $d = 0.70$.

Mean responsibility attribution was higher than mean choice attribution, $t(52) = 7.53$, $p < .001$, $MD = 2.49$ [1.83, 3.15], $d = 1.19$, and also higher than mean blame attribution, $t(52) = 3.84$, $p < .001$, $MD = 1.17$ [0.56, 1.78], $d = 0.62$.

Discussion

The overall pattern of blame attribution vindicated neither compatibilism nor incompatibilism about blame. Instead, the central tendency was one of ambivalence. Blame attribution was lower when the action was impossible than when it was possible, which suggests some incompatibilist sentiment about blame. Nevertheless, blame attribution was still higher than both choice attribution and the rate at which people agreed that the agent could perform the relevant action, which suggests some compatibilist sentiment about blame. The results replicated the earlier findings supporting incompatibilism about choice and compatibilism about moral responsibility. The results also generalized those earlier findings by showing that the same pattern emerges when choice and moral responsibility are rated within-subjects.

Experiment 4

This experiment tests whether the findings from Experiment 3 are robust across other actions and narrative contexts.

Method

Participants

One hundred new participants were tested (aged 18-68 years, mean age = 34 years; 48 female; 95% reporting English as a native language).

Materials and Procedure

Participants were randomly assigned to one of two conditions (impossible, possible) in a between-subjects design. The basic procedures were the same as in Experiment 3 but the cover story was completely new. The basic story featured a man walking his dog in the park. The man does not warn other people that the dog is violent and the dog bites someone. In the impossible condition, it is not possible for the man to warn others. In the possible condition, it is possible for the man to warn others. Here is the text of the story:

A man is walking his dog in a public park. The dog is very violent. Given the current condition of the man's brain, it is [impossible/possible] for him to warn people about the dog. As a matter of brain chemistry, it is [literally impossible/possible] that he can warn people. He does not warn anyone. The dog bites someone.

After reading the story, participants responded to two statements in a matrix table (order randomized):

1. The man could have warned others. (could)

2. The man is blameworthy for not warning others. (blame)

Participants then advanced to a new screen and responded to two more statements in the same way:

3. The man chose not to warn others. (choice)
4. The man had a moral responsibility to warn others. (responsibility)

Participants then advanced to a new screen and completed a percentage task similar to the one from earlier experiments:

5. On a scale of 0% to 100%, how likely was it that the man would warn others? (percent)

Results

The status manipulation was effective again and the basic pattern of response to the could statement and percentage task was the same as in Experiments 2 and 3. (See Fig. 4.) Independent samples t-tests showed that mean response was higher in the possible condition for all test items. (See Table 9.) One sample t-tests showed that mean response to the could statement and choice attribution was below the neutral midpoint (= 4) in the impossible condition and above the midpoint in the possible condition. (See Table 10.) Mean blame attribution was above the midpoint in the possible condition but no different from the midpoint in the impossible condition. Mean responsibility attribution was above the midpoint in both conditions. Overall, then, the pattern of response was very similar to that observed in Experiment 3, except that this time responsibility attribution was higher in the possible condition.

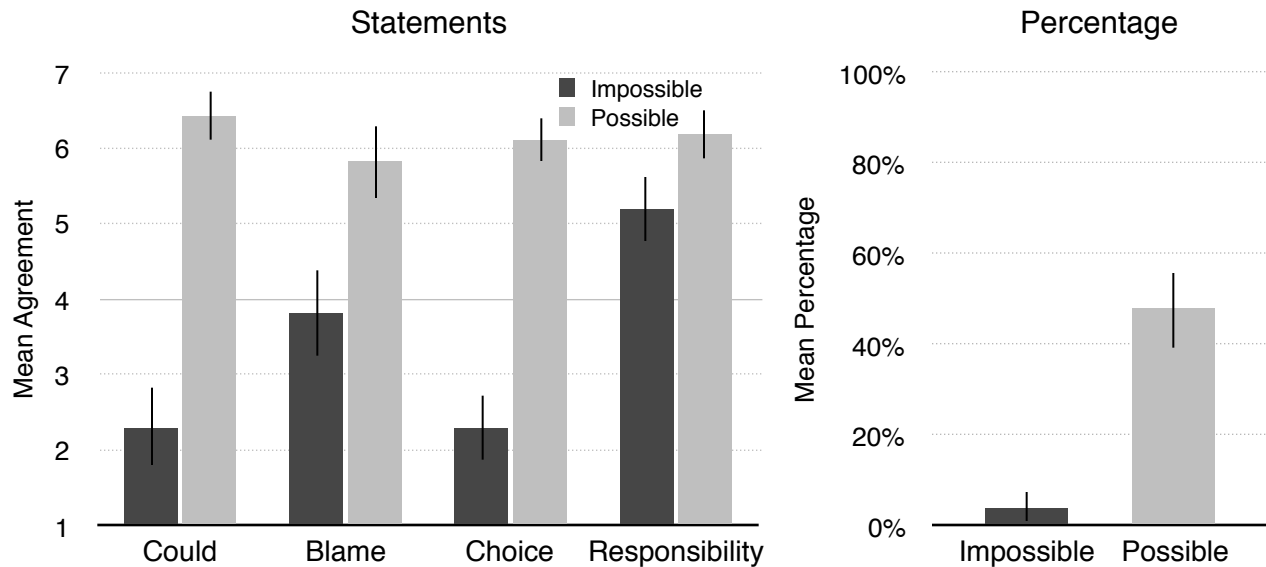


Fig. 4. Experiment 4. Panel A: mean response to the test statements about whether the agent could have performed the action, is blameworthy for not performing the action, chose not to perform the action, and had a moral responsibility to perform the action. Panel B: mean estimate of the percentage chance that the agent would perform the action. For Panel A, the scale ran 1 (SD) - 7 (SA) and the midpoint (= 4) is marked by a solid line. For Panel B, the scale ran 0-100%. Error bars represent bootstrapped 95% confidence intervals.

Table 9. Experiment 4. Independent samples t-tests.

Measure	Impossible		Possible		t	df	p	MD	95% CI	d
	M	SD	M	SD						
could	2.31	1.95	6.43	1.19	-12.77	83.15	<.001	-4.12	-4.76, -3.47	2.80
blame	3.82	2.19	5.82	1.65	-5.15	92.92	<.001	-1.99	-2.76, -1.22	1.07
chose	2.31	1.50	6.12	1.13	-14.36	92.70	<.001	-3.81	-4.34, -3.28	2.98
responsible	5.20	1.58	6.18	1.15	-3.59	91.48	.001	-0.99	-1.53, -0.44	0.75
percent	3.96	12.41	48.18	30.72	-9.37	62.76	<.001	-44.22	-53.66, -34.79	2.37

Table 10. Experiment 4. One sample t-tests. Test value = 4.

Measure	Impossible						Possible					
	t	df	p	MD	95% CI	d	t	df	p	MD	95% CI	d
could	-6.16	50	<.001	-1.69	-2.24, -1.14	0.87	14.28	48	<.001	2.43	2.09, 2.77	2.04
blame	-0.58	50	.567	-0.18	-0.79, 0.44	0.08	7.69	48	<.001	1.82	1.34, 2.29	1.10
chose	-8.01	50	<.001	-1.69	-2.11, -1.26	1.13	13.15	48	<.001	2.12	1.80, 2.45	1.88

Measure	Impossible						Possible					
	t	df	p	MD	95% CI	d	t	df	p	MD	95% CI	d
responsible	5.42	50	<.001	1.20	0.75, 1.64	0.76	13.31	48	<.001	2.18	1.85, 2.51	1.89

Paired sample t-tests showed that in the impossible condition, mean blame attribution was higher than mean response to the could statement, $t(50) = 4.65$, $p < .001$, $MD = 1.51$ [0.86, 2.16], $d = 0.73$, and also higher than mean choice attribution, $t(50) = 5.06$, $p < .001$, $MD = 1.51$ [2.10, 5.06], $d = 0.79$. Mean response to the could statement was the exact same as mean choice attribution, $t(50) = 0.00$, $p = 1$, n.s. Mean responsibility attribution was higher than mean blame attribution, $t(50) = 5.14$, $p < .001$, $MD = 1.37$ [0.84, 1.91], $d = 0.70$, higher than mean response to the could statement, $t(50) = 9.53$, $p < .001$, $MD = 2.88$ [2.28, 3.49], $d = 1.62$, and also higher than mean choice attribution, $t(50) = 10.16$, $p < .001$, $MD = 2.88$ [2.31, 3.45], $d = 1.87$.

Focusing on only those participants in the impossible condition who answered “0%” on the percentage task ($N = 41$), the same basic pattern emerges. One sample t-tests showed that mean response to the could statement was below the midpoint ($M = 2.17$, $SD = 1.87$), $t(40) = -6.27$, $p < .001$, $MD = -1.83$ [-2.42, -1.24], $d = 0.98$. Mean blame attribution did not differ from the midpoint ($M = 3.71$, $SD = 2.18$), $t(40) = 0.86$, $p = .396$, n.s. Mean choice attribution was below the midpoint ($M = 2.32$, $SD = 1.49$), $t(40) = -7.23$, $p < .001$, $MD = -1.68$ [-2.15, -1.21], $d = 1.13$. Mean responsibility attribution was above the midpoint ($M = 5.24$, $SD = 1.51$), $t(40) = 4.26$, $p < .001$, $MD = 1.24$ [0.77, 1.72], $d = 0.82$. Paired sample t-tests showed that mean blame attribution was higher than mean choice attribution, $t(40) = 4.45$, $p < .001$, $MD = 1.39$ [0.76, 2.02], $d = 0.72$.

Mean responsibility attribution was higher than mean choice attribution, $t(40) = 10.13$, $p < .001$, $MD = 2.93$ [2.34, 3.51], $d = 1.95$, and also higher than mean blame attribution, $t(40) = 5.15$, $p < .001$, $MD = 1.54$ [0.93, 2.14], $d = 0.80$.

Discussion

The results replicate the basic findings of the earlier experiments and demonstrate that they extend to other actions and narrative contexts. As in Experiment 3, the overall pattern was consistent with incompatibilism about choice, compatibilism about moral responsibility, and ambivalence about blame.

Experiment 5

This experiment tests people's judgments about credit for desirable outcomes and compares them to judgments about blame for undesirable outcomes. The main question is whether judgments about credit and blame will differ when it is impossible for the agent to act otherwise. This experiment also further tests the robustness of the earlier findings by testing judgments about another different action and narrative context.

Method

Participants

Two hundred and one new participants were tested (aged 18-82 years, mean age = 32 years; 82

female; 96% reporting English as a native language).

Materials and Procedure

Participants were randomly assigned to one of four conditions in a 2 (outcome: harmful, helpful) × 2 (status: impossible, possible) between-subjects design. The basic procedures were the same as in Experiments 3 and 4. The story focused on a man who notices a small child struggling in a shallow pond of water. The outcome factor manipulated whether the man does not or does save the child. When the man does not save the child, participants rated whether he deserves blame for not saving the child. When the man does save the child, participants rated whether he deserves credit for saving the child. The status factor manipulated whether it was impossible or possible for the man to have done otherwise. Here is the text of the story for the helpful condition (status manipulations in brackets):

(Helpful) A man is walking through the park when he notices a small child struggling in a shallow pond. No one else is around. The child is alone and in danger. Given the current condition of the man's brain, it is [impossible/possible] for him to not rescue the child. As a matter of brain chemistry, it is [literally impossible/possible] for him to not rescue the child. He does rescue the child.

The story for the harmful condition was exactly the same except that it is impossible/possible for the man to rescue the child and he does not rescue the child. Participants then responded to five test items (outcome manipulation in brackets):

1. The man could have [not] rescued the child. (could)

2. The man deserves credit [blame] for [not] rescuing the child. (accountability)
3. The man chose to [not] rescue the child. (choice)
4. The man had a moral responsibility to rescue the child. (responsibility)
5. On a scale of 0% to 100%, how likely was it that the man would rescue the child.
(percent)

As in Experiments 3 and 4, items 1 and 2 appeared on the first page of the survey (order randomly rotated), items 3 and 4 appeared on the second page (order randomly rotated), and item 5 appeared on the third page. The story remained at the top throughout.

Results

The status manipulation was effective: people agreed that the agent could act otherwise when acting otherwise was possible, and they disagreed when acting otherwise was impossible. Response to the percentage task was very sensible and closely resembled results from the earlier experiments. A preliminary multivariate analysis of variance revealed main effects on all the dependent variables and three interaction effects, including a large interaction effect on accountability attributions (blame or credit). (See Fig. 5 and Table 11.) Table 12 reports pairwise comparisons of mean scores on the five dependent measures for the impossible and possible conditions, grouped by outcome type. Table 13 reports one sample t-tests for the four test statements in all the conditions.

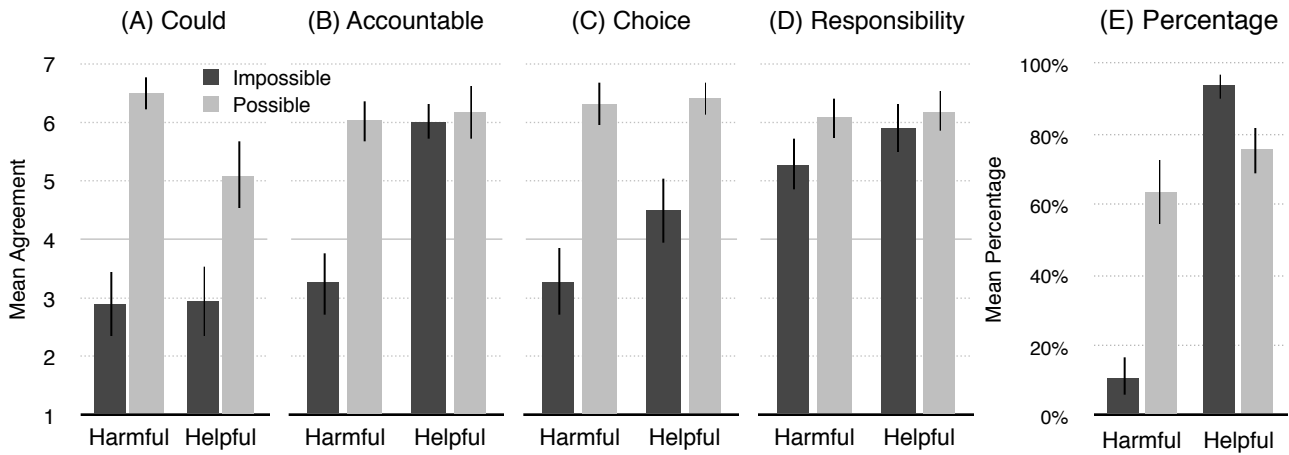


Fig. 5. Experiment 5. Mean response to the test statements about whether the agent could have behaved otherwise (Panel A), is accountable (i.e. deserves blame [harmful] or credit [helpful]) for his behavior (Panel B), chose to behave as he did (Panel C), and had a moral responsibility to save the child (Panel D). Panel E shows the mean estimate of the percentage chance that the agent would save the child. For Panels A-D, scales ran 1 (SD) - 7 (SA) and the midpoint (= 4) is marked by a solid line. For Panel E, the scale ran 0-100%. Error bars represent bootstrapped 95% confidence intervals.

Table 11. Experiment 5. Multivariate analysis of variance.

Measure	Factor											
	Outcome				Status				Outcome*Status			
	F	df	p	η_p^2	F	df	p	η_p^2	F	df	p	η_p^2
could	6.40	1, 197	.012	.031	114.98	1, 197	<.001	.368	7.20	1, 197	.008	.035
accountable	46.32	1, 197	<.001	.190	47.60	1, 197	<.001	.195	36.73	1, 197	<.001	.157
choice	8.02	1, 197	.005	.039	115.35	1, 197	<.001	.369	5.76	1, 197	.017	.028
responsible	3.78	1, 197	.053	.019	7.59	1, 197	.006	.037	1.77	1, 197	.184	.009
percent	203.64	1, 197	<.001	.508	26.99	1, 197	<.001	.121	115.31	1, 197	<.001	.369

Table 12. Experiment 5. Independent samples t-tests.

Measure	Impossible		Possible		t	df	p	MD	95% CI	d
	M	SD	M	SD						
Harmful										
could	2.90	2.00	6.50	0.95	-11.48	70.11	<.001	-3.60	-4.23, -2.97	2.74
accountable	3.24	1.95	6.02	1.29	-8.40	84.72	<.001	-2.78	-3.49, -2.12	1.83
choice	3.28	2.06	6.32	1.19	-9.04	78.24	<.001	-3.04	-3.71, -2.37	2.04
responsible	5.28	1.58	6.08	1.26	-2.80	93.39	.006	-0.80	-1.37, -0.23	0.58
percent	10.64	18.47	63.60	33.36	-9.82	76.45	<.001	-52.96	-63.70, -42.22	2.25
Helpful										
could	2.94	2.22	5.10	2.15	-4.96	99	<.001	-2.16	-3.02, -1.30	1.00
accountable	6.00	1.15	6.18	1.57	-0.66	99	.512	-0.18	-0.72, 0.36	0.13
choice	4.49	1.99	6.42	1.05	-6.10	76.14	<.001	-1.93	-2.56, -1.30	1.40
responsible	5.92	1.48	6.20	1.20	-1.04	99	.302	-0.28	-0.81, 0.25	0.21
percent	93.76	13.79	75.34	24.16	4.69	77.53	<.001	18.42	10.61, 26.24	1.07

Table 13. Experiment 5. One sample t-tests. Test value = 4.

Measure	Impossible						Possible					
	t	df	p	MD	95% CI	d	t	df	p	MD	95% CI	d
Harmful												
could	-3.88	49	<.001	-1.10	-1.67, -0.53	0.55	18.55	49	<.001	2.50	2.23, 2.77	2.63
accountable	-2.75	49	.008	-0.76	-1.32, -0.20	0.39	11.11	49	<.001	2.02	1.65, 2.39	1.57
choice	-2.47	49	.017	-0.72	-1.31, -0.13	0.35	13.84	49	<.001	2.32	1.98, 2.66	1.95
responsible	5.73	49	<.001	1.28	0.83, 1.73	0.81	11.68	49	<.001	2.08	1.72, 2.44	1.65
Helpful												
could	-3.40	50	.001	-1.06	-1.68, -0.43	0.48	3.62	49	.001	1.10	0.49, 1.71	0.51
accountable	12.43	50	<.001	2.00	1.68, 2.32	1.74	9.79	49	<.001	2.18	1.73, 2.63	1.39

Measure	Impossible						Possible					
	t	df	p	MD	95% CI	d	t	df	p	MD	95% CI	d
choice	1.76	50	.085	0.49	-0.07, 1.05	0.25	16.27	49	<.001	2.42	2.12, 2.72	2.30
responsible	9.27	50	<.001	1.92	1.50, 2.34	1.30	13.02	49	<.001	2.20	1.86, 2.54	1.83

The interaction effect on accountability attributions is critical. When the man rescued the child, people strongly agreed that the man deserved credit for doing so, whether it was possible or impossible for him to do otherwise. By contrast, when the man did not rescue the child, people disagreed that the man deserved blame when it was impossible for him to do otherwise, whereas they agreed that he deserved blame when it was possible for him to do otherwise. In other words, in only one condition — when it was impossible to save the child — did people deny accountability. In the other three conditions, people attributed accountability.

There was a similar but smaller interaction effect on choice attributions. When acting otherwise was possible, people strongly agreed that the man chose to behave as he did. However, when acting otherwise was impossible, people tended to disagree that the man chose to not rescue the child, whereas they exhibited a marginal tendency to agree that the man chose to rescue the child. Nevertheless, whereas there was no difference in accountability between the impossible and possible helpful condition, there was a large difference in choice attributions between those two conditions.

In all four conditions, people agreed that the man had a moral responsibility to rescue the child.

I also analyzed the data from only those participants who were either in the impossible

harmful conditions and answered “0%” on the percentage task (N = 30), or in the impossible helpful condition and answered “100%” on the percentage task (N = 34). Among these participants, mean response to the could statement was low and did not differ between conditions (M = 2.10/2.35, SD = 1.81/2.00), $t(62) = -0.53$, $p = .599$, n.s. Mean blame attribution was below the neutral midpoint (= 4), $t(29) = -3.88$, $p < .001$, MD = -1.40 [-2.14, -0.66], $d = 0.71$, and mean credit attribution was above the midpoint, $t(33) = 9.25$, $p < .001$, MD = 1.97 [1.54, 2.40], $d = 1.59$. Mean choice attribution was below the midpoint in the harmful condition (M = 2.37, SD = 1.92), $t(29) = -4.66$, $p < .001$, MD = -1.63 [-2.35, -0.92], $d = 0.85$, but it did not differ from the midpoint in the helpful condition (M = 3.88, SD = 2.07), $t(33) = -0.33$, $p = .743$, n.s. Mean responsibility attribution was above the midpoint in the harmful condition (M = 5.13, SD = 1.72), $t(29) = 3.62$, $p = .001$, MD = 1.13 [0.49, 1.77], $d = 0.66$, and the helpful condition (M = 5.94, SD = 1.43), $t(33) = 7.89$, $p < .001$, MD = 1.94 [1.44, 2.44], $d = 1.36$.

Discussion

The results again replicate and further generalize the earlier findings supporting compatibilism about moral responsibility. At the same time, the results support a more nuanced interpretation of the earlier findings on the relationship between choice and the ability to behave otherwise, on the one hand, and blame and the ability to behave otherwise, on the other hand. In earlier experiments, when the agent could not have behaved otherwise, participants denied that the agent chose to behave as he or she did. In this experiment, when the agent could not have behaved otherwise and the outcome was harmful, participants also denied that the agent chose his behavior.

However, when the outcome was helpful, participants did not deny that the agent chose his behavior. Instead, they exhibited a marginal tendency to agree that the agent did choose. Similarly, in earlier experiments, when the agent could not have behaved otherwise, participants did not hold the agent accountable by saying he was blameworthy for a harmful outcome. In this experiment, when the agent could not have behaved otherwise and the outcome was harmful, participants again did not say that he deserved blame. However, when the outcome was helpful, participants strongly agreed that the agent deserved credit.

General Discussion

The results from five experiments support several conclusions about compatibilism and incompatibilism in ordinary social cognition. First, and most clearly, people are compatibilists about moral responsibility (Experiments 2-5). Across a wide range of actions and evaluative contexts, people attribute moral responsibilities to agents even while acknowledging that it is impossible for the agent to fulfill them. For example, when a woman is evaluating an employee who performed excellently, people judge that the woman has a moral responsibility to give the employee a good evaluation. People attribute this responsibility at equal rates whether or not it is possible for the woman to give the employee a good evaluation (Experiment 2).

Second, people are compatibilists about *positive accountability* (i.e. deserving credit for good outcomes), but they do not seem to be compatibilists about *negative accountability* (i.e. deserving blame for bad outcomes). For example, when an agent rescues a child and could not have

done otherwise, people judge that he deserves credit for saving the child. By contrast, when an agent does not rescue a child and could not have done so, people disagree that he deserves blame for not saving the child (Experiment 5).

Third, people seem to lean toward compatibilism about choice for helpful actions, but they are incompatibilists about choice for harmful and neutral actions. For example, when an agent rescues a child and could not have done otherwise, people tend to judge that he chose to rescue the child. By contrast, when an agent does not rescue a child and could not have done so, people disagreed that he chose to not rescue the child (Experiment 5). Similarly, when an agent performs a neutral action, such as rolling a marble to the left across a table, participants judged that the agent chose to roll it left when rolling it another way was possible, but they denied that he chose to roll it left when that was the only option (Experiment 1). I found no evidence that people view choice differently when it is made by humans compared to non-human animals, which is consistent with recent work that found no difference in how people view agency in human and non-human animals (Turri under review b). I also found no evidence that people distinguish between choice and free choice, which suggests that, at least for harmful and neutral actions, people are incompatibilists about free choice because they are incompatibilists about choice.

Taken together, the present findings advance understanding of important social judgments and contribute to longstanding theoretical debates in philosophy, theology, and psychology about the relationship between determinism, free will, and morality. The findings paint a fascinatingly nuanced mix of compatibilism and incompatibilism about several closely related but distinct concepts. If ordinary social cognition exhibits this subtle mix of compatibilist and incompatibilist

sentiment, then it could explain why the the theoretical debate between compatibilists and incompatibilists has lasted so long. If determinism is compatible with some of the key categories connected to moral evaluation but not others, then perhaps each side of the debate has grasped part of the truth while failing to recognize that its grasp is only partial. This is consistent with recent work on the psychometrics of free-will attributions and moral judgments, which suggests that people exhibit a mix of compatibilist and incompatibilist tendencies (Deery, Davis & Carey 2014).

Some previous work on related issues has drawn conclusions based on how people tend to judge a single scenario, raising the possibility that the results are an artifact of how people react to one particular type of act or situation. The present research avoided this drawback by replicating key findings across multiple difference scenarios. Previous work has also lacked checks to assess how participants understood critical aspects of the scenario, in particular whether participants judged that the relevant action or outcome was going to happen or could still be avoided. The present research avoided this drawback by having participants judge the relevant aspects of the scenario.

Future work could explore the present findings in several ways. For example, although I observed similar results for several different types of action in contexts where people judged only one outcome to be possible, perhaps we will observe different patterns for other types of action or other forms of behavioral determination. To take another example, perhaps the mixed results about moral accountability or choice could be explained on the grounds that some of the findings reflect performance error or interference from pragmatic considerations. There are many

ways to elaborate the basic idea, but one possibility is that the finding on negative accountability reflects an unwillingness to explicitly blame someone, even though people do judge that the agent is blameworthy. For instance, people might be unwilling to explicitly blame someone without more information about the context because they do not want to appear vindictive, especially when the outcome is serious (e.g. a child drowning). In other words, it could be that people are compatibilists about negative moral accountability, but pragmatic considerations interfere with the expression of their judgment. Prior work has studied judgments about agency and determinism in children (e.g. Nichols 2004; Schulz & Sommerville 2006) and across cultures (e.g. Sarkissian et al. 2010), so a third possibility for future work is to study whether the present findings are robust developmentally and cross-culturally.

Acknowledgments — For helpful comments and feedback, I thank Wesley Buckwalter, Rick Cooper, Robert Kane, Al Mele, Thomas Nadelhoffer, Eddy Nahmias, David Rose, Angelo Turri, and Manuel Vargas. Thanks also to an audience at the 2015 meeting of the Society for Philosophy and Psychology. This research was supported by the Social Sciences and Humanities Research Council of Canada and an Early Researcher Award from the Ontario Ministry of Economic Development and Innovation.

References

- Clarke, R. (2003). *Libertarian theories of free will*. Oxford: Oxford University Press.
- Clarke, R., & Capes, J. (2013). Incompatibilist (nondeterministic) theories of free will. (E. N.

- Zalta, Ed.)Stanford Encyclopedia of Philosophy. Retrieved December 2014, from <http://plato.stanford.edu/entries/incompatibilism-theories/>
- Cova, F., & Kitano, Y. (2014). Experimental philosophy and the compatibility of free will and determinism: a survey. *Annals of the Japan Association for Philosophy of Science*, 22, 17–37.
- Deery, O., Davis, T., & Carey, J. (2014). The Free-Will Intuitions Scale and the question of natural compatibilism. *Philosophical Psychology*, 1–26. doi:10.1080/09515089.2014.893868
- Dennett, D. (1984). *Elbow room: the varieties of free will worth wanting*. Cambridge, Mass: MIT Press.
- Fischer, J. M. (1994). *The metaphysics of free will*. Oxford: Blackwell.
- Fischer, J. M., & Ravizza, M. (1998). *Responsibility and control: a theory of moral responsibility*. Cambridge: Cambridge University Press.
- Hume, D. (1993). *An enquiry concerning human understanding*. (E. Steinberg, Ed.) (2nd ed.). Indianapolis: Hackett.
- Kane, R. (1999). Responsibility, luck, and chance: reflections on free will and indeterminism. *Journal of Philosophy*, 96(5), 217–240.
- May, J. (2014). On the very concept of free will. *Synthese*, 191(12), 2849–2866. doi:10.1007/s11229-014-0426-1
- McKenna, M. (2009). Compatibilism. (E. N. Zalta, Ed.)Stanford Encyclopedia of Philosophy. Retrieved December 2014, from <http://plato.stanford.edu/entries/compatibilism/>
- Monroe, A. E., & Malle, B. F. (2010). From uncaused will to conscious choice: the need to study,

- not speculate about, people's folk concept of free will. *Review of Philosophy and Psychology*, 1(2), 211–224. doi:10.1007/s13164-009-0010-7
- Monroe, A. E., Dillon, K. D., & Malle, B. F. (2014). Bringing free will down to Earth: people's psychological concept of free will and its role in moral judgment. *Consciousness and Cognition*, 27, 100–108. doi:10.1016/j.concog.2014.04.011
- Nahmias, E., Morris, S., Nadelhoffer, T., & Turner, J. (2005). Surveying Freedom: Folk Intuitions about free will and moral responsibility. *Philosophical Psychology*, 18(5), 561–584. doi:10.1080/09515080500264180
- 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. (2004). The folk psychology of free will: fits and starts. *Mind & Language*, 19(5), 473–502.
- Nichols, S. (2011). Experimental Philosophy and the Problem of Free Will. *Science*, 331(6023), 1401–1403. doi:10.1126/science.1192931
- 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
- O'Connor, T. (2010). Free Will. (E. N. Zalta, Ed.) *Stanford Encyclopedia of Philosophy*. Retrieved December 2014, from <http://plato.stanford.edu/entries/freewill/>
- Pereboom, D. (2001). *Living without free will*. Cambridge: Cambridge University press.
- Reid, T. (1785). *Essays on the active powers of man*. Edinburgh: Bell, Robinson & Robinson.

- Rose, D., & Nichols, S. (2013). The Lesson of Bypassing. *Review of Philosophy and Psychology*, 4, 599–619.
- Rose, D., Buckwalter, W., & Nichols, S. (under review). Neuroscientific prediction and the intrusion of intuitive metaphysics.
- Roskies, A., & Nichols, S. (2008). Bringing moral responsibility down to earth. *Journal of Philosophy*, 105(7), 371–388.
- 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.
- Schulz, E., Cokely, E. T., & Feltz, A. (2011). Consciousness and Cognition. *Consciousness and Cognition*, 20(4), 1722–1731. doi:10.1016/j.concog.2011.04.007
- Schulz, L. E., & Sommerville, J. (2006). God does not play dice: Causal determinism and preschoolers' causal inferences. *Child Development*, 77(2), 427–442.
- Sommers, T. (2010). Experimental Philosophy and Free Will. *Philosophy Compass*, 5(2), 199–212. doi:10.1111/j.1747-9991.2009.00273.x
- Turri, J. (under review a). Compatibilism can be natural. University of Waterloo.
- Turri, J. (under review b). Exceptionalist naturalism: human agency and the causal order. University of Waterloo.
- Vihvelin, K. (2011). Arguments for incompatibilism. (E. N. Zalta, Ed.) *Stanford Encyclopedia of Philosophy*. Retrieved December 2014, from <http://plato.stanford.edu/entries/incompatibilism-arguments/>
- Woolfolk, R. L., Doris, J. M., & Darley, J. M. (2006). Identification, situational constraint, and

social cognition: Studies in the attribution of moral responsibility. *Cognition*, 100, 283–301.