Children and Adults Don't Think They Are Free: A Skeptical Look at Agent Causationism

Lukas S. Huber[1, *, §] Kevin Reuter[2, *, §] Trix Cacchione[3]

[1]*University of Bern*
[2]*University of Zurich*
[3]*University of Applied Sciences and Arts Northwestern Switzerland*
[*]*Joint first authors*
[§]*To whom correspondence should be addressed:*
`lukas.huber1@students.unibe.ch` *and* `kevin.reuter@uzh.ch`
This research was supported by the Swiss National Science Foundation (grant number: 100012_169484)

Abstract

Two strands of evidence supposedly exist in support of the claim that people think of themselves as agent causationists, that is, as agents who can start and prevent causal chains. First, results from developmental studies suggest that children between the ages of four and six undergo a transition towards thinking of themselves as unconditional free agents. Second, experimental studies indicate that adults think of themselves as agents who, having made some choice, could have done otherwise under exactly the same circumstances. In this paper, we present new evidence that tells against both strands of evidence. Based on empirical data we collected with children ages four to six (Study 1), we argue that six-year-old children only endorse freedom of choice if they are presented with at least two conflicting desires which are compatible with their own desires. This undermines any strong conclusion to the claim that children think of themselves as agent causationists. The results of Study 2 reveal that people (adults) indeed agree that they 'could have done otherwise' given the same circumstances, but only when this phrase is interpreted as a matter of *ability*. When people are asked whether *it is possible* that an agent does otherwise, holding the circumstances exactly the same, a majority of people think not. Given that belief in agent causationism is one of the main motivations for the metaphysical account of agent causation, our results also can be seen as evidence against agent causation.

Keywords: *Agent Causation; Conflicting Desires; Developmental Studies; Experimental Philosophy; Free Will.*

1 Introduction

Imagine you walk by a table. On the table is a plate of cookies. They look delicious, though not particularly healthy. You think it over and decide to take a cookie. You eat it, and it is

indeed delicious. A researcher appears and wishes to ask you some questions about free will and your decision to take a cookie. "You took a cookie," the researcher says. "But could you have done otherwise?"

Philosophers have noticed a crucial ambiguity in this type of question. The phrase 'could you have done otherwise' can be understood in two different ways (Hobbes, 1839/1646; Nichols, 2004; Turner and Nahmias, 2006): according to an *unconditional reading*, the question asks whether you could have done otherwise if your beliefs and desires were exactly the same. In contrast, the *conditional reading* allows for changes to your original state of mind, and hence asks whether you would have done otherwise if you'd had some other beliefs and desires. The conditional reading is compatible with determinism, while the unconditional is not.

Consequently, if we are interested to know how people think about free will, it will not be sufficient to ask them whether they believe that they could have done otherwise. Importantly, we also need to find out whether people interpret the phrase "could have done otherwise" conditionally or unconditionally. Recently, a number of empirical studies (Nichols, 2006; Nichols and Knobe, 2007; Sarkissian et al., 2010) have been interpreted as demonstrating that (at least) many of us entertain an unconditional notion of free will. It would seem that many people believe that they could have chosen not to eat the cookie, even if everything (else) had been exactly the same.

One popular way to make sense of this is agent causationism. This is the view that people *believe* they can prevent causal chains from happening as well as start new causal chains. Our primary concern in this essay, is thus not whether agent causation exists, but only whether people think of themselves as agent causationists. Applied to the case at hand, we are interested in whether people think they can intervene in the causal process leading from their desire to actually taking the cookie and, instead, guide themselves towards a different action.[1]

What are the results that suggest that people think of themselves as agent causationists? Two strands of evidence have been put forward in support of that claim. First, certain developmental studies have been interpreted to show that children between the ages of four and six undergo a transition towards thinking of themselves as unconditional free agents (Nichols, 2004; Kushnir, Gopnik, Chernyak, Seiver, and Wellman, 2015). Second, experimental studies indicate that adults think of themselves as agents who can do otherwise under exactly the same circumstances (Nichols, 2006; Nichols and Knobe, 2007; Sarkissian et al., 2010). In this paper, we present new research that tells against both these strands of evidence. Let us start with the first strand.

The literature on children's understanding of free will is still scarce. To our knowledge, only a few studies have examined conceptions of free will in four- to six-year-old children (Nichols, 2004; Kushnir et al., 2015; Wente et al., 2016, Chernyak, Kushnir, Sullivan, and

[1]While our focus is on whether people believe they can start and stop causal chains, these beliefs are often put forward as a motivating reason for arguing for agent causation (Campbell, 1967; O'Connor, 1995). Thus, the debate on agent causationism has a direct bearing on the metaphysical account of agent causation (see our General Discussion for further details.)

Wang, 2011). However, these studies seem to agree on the point that young children reason in accord with an unconditional conception of free will. Nichols (2004), for example, concludes that children believe in agent causation:

> "[T]he available evidence provides support for the claim that children embrace both claims of the agent-causal account. Apparently children think that an agent is a causal factor in the production of an action." (Nichols 2004, p. 488)

Along similar lines, Kushnir et al. (2015) propose that during development an early intuitive theory is replaced by a theory in which free will is conceptualized as some sort of causal force that mediates between desires and action:

> "In particular, we propose there might be an earlier intuitive causal theory in place by four, or even in late infancy, in which desires are the immediate and *necessary* cause of choices and actions and so are tightly linked to choice itself. Between four and six that intuitive theory may be replaced by a theory in which a more powerful sense of choice is a further causal factor, choice as a separate mental activity that can itself influence and modify not only actions but desires." (Kushnir et al. 2015, p. 98)

In this paper, however, we will put forward a different view. Based on new empirical data, collected with children aged four to six (Study 1), we argue that children do not necessarily believe in agent causation. Instead, our results show that children think they could have done otherwise only if a conflicting desire is made salient. This suggests that, to children, 'could have done otherwise' means 'would have done otherwise if I had a different desire'. Hence, they may indeed think of their freedom conditionally.

The second strand of evidence comes from questionnaire studies showing that adults think they could have done otherwise even if everything had been exactly the same (Nichols, 2006; Nichols and Knobe, 2007; Sarkissian et al., 2010). We do not dispute that data. In fact, our own data shows that people strongly agree that they can do otherwise in exactly the same situation. However, we will argue that a further ambiguity in the phrases 'could have done' and 'can do' has gone largely unnoticed. These phrases can be interpreted to mean either 'it is possible to do' or 'having the ability to do' (see, e.g., Kratzer, 1991). In order to show that people think of themselves as agent causationists, one needs to make sure that the phrase 'can do otherwise' is being interpreted (by study participants, for example) as a *possibility*, and not as an *ability*. The results of Study 2 reveal that the crucial question is, by and large, interpreted as a question about ability. When people are asked to state whether it is possible that an agent does otherwise when everything remains the same, a majority of people say that it is not, which suggests that adults think of their freedom conditionally.

Here is how we will proceed. In section 2, we discuss recent empirical work that seems to demonstrate that children think of themselves as agent causationists. We then present the results of Study 1, in which we tested children's intuitions across different conditions, thereby investigating the influence of conflicting desires on the ascription of freedom of

choice. Our empirical work on adults' intuitions on free will (Study 2) is presented and discussed in section 3. In section 4, we summarize our results, and discuss how our studies bear on the metaphysical account of agent causation.

2 Study 1: Children and Agent Causationism

Researchers usually opt to examine children's intuitions about free will by first letting them (or another agent) perform a certain action and subsequently asking them, "Could you (the agent) have done otherwise or did you (the agent) have to act that way?" (e.g. Lane, Ronfard, Francioli, and Harris, 2016; Kushnir et al., 2015; Wente et al., 2016; Chernyak et al., 2011; Nichols, 2004). Attributing the ability to do otherwise (to yourself or another agent) is a necessary condition for attributing agent causationism: without the ability to do otherwise, the question of whether the agent can start and prevent causal chains doesn't even arise.

Nichols (2004) found that young children distinguish between the ability of objects and that of agents to behave otherwise. For example, children were asked whether a ball that fell to the floor due to gravity and an agent who touched the floor with her hand due to a desire could have acted other than they did. All the children attributed the ability to do otherwise to agents but not to the ball, suggesting that children between the ages of four and six develop intuitions fitting the agent causation view. Against this interpretation, Nichols himself raises the following possibility: "the compatibilist might say that when the children claimed that the agent could have done otherwise, they were only claiming that the experimenter would have done otherwise under different conditions" (p.486). To address this concern, Nichols ran a second study, in which children were told that all the conditions were exactly the same. The results of the second study were somewhat less clear, but there was still a tendency for children to attribute greater ability to do otherwise to agents compared to objects. Nichols therefore concludes that "children regard agents as having the capacity to have done otherwise in a way that can't merely be reduced to a conditionalized analysis" (p.488). Some scholars have raised doubts that these experiments show that people's conception of choosing indeed fundamentally differs from their conception of physical causation (Turner and Nahmias, 2006). The found differences could rather be explained by a difference in the complexity-level of the two processes (human decision making vs. ball dropping). Indeed, Turner and Nahmias showed that when adults were confronted with two processes which are better matched for their complexity (human decision making vs. a lightning strike) and had to rate how these processes played out each time a hypothetical universe was recreated, they no longer assumed that the physical processes are different from the process of choosing in terms of their inevitability. Turner and Nahmias summarize their findings stating that "our results suggest that most people do not believe that all physical processes are deterministic while human choices are indeterministic, even if they might believe that certain simple physical events are deterministic" (2006, p. 605).

Kushnir et al. (2015) ultimately arrive at the same conclusion as Nichols, although the primary aim of their research was not to establish an unconditional reading of the ability to do otherwise, but rather to detail the transitional phase in children's reasoning that happens between the ages of four and six. In this research, experimenters asked four- and six-year-olds whether a certain agent (either a toy character or the child him- or herself) could choose to act against their desire. Across several presented stories, the agent was either about to perform an undesired action (action stories) or to inhibit a desired action (inhibition stories). Children were then asked whether the agent has to do x or whether s/he could choose not to do x. A 'have to' response would indicate that the children conceived of the agent's choice as constrained by her desire and that this desire led necessarily to the corresponding action. In contrast, a 'choose to' response would indicate an intuition according to which a person can choose to act against her desire.

The results showed that 6-year-olds give significantly more 'choose to' responses than 4-year-olds. While 6-year-olds performed above chance level in most cases, this was not true for 4-year-olds.[2] Crucially, Kushnir et al. (2015) also invited children to provide qualitative explanations for their responses. Looking at explanations of 'choose to' responses, they found that most children either said they didn't know or referred to alternative desires or alternative external conditions. Other explanations mentioned the agent's autonomy to act against her desire (17% when explaining another agent's freedom of choice, and 10% when explaining their own freedom of choice).

These latter explanations were interpreted as reflecting the beginning of an unconditional understanding of free will. In contrast, 'have to' responses were viewed as indicators for a concept of free will in which desires necessarily lead to corresponding actions (necessary-link concept). And the conditional explanations of 'choose to' responses were understood as reflecting some sort of developmental transitional stage in which precursor intuitions are giving way to intuitions about an autonomous free will. Although this picture seems perfectly consistent with the data, we propose another possible interpretation: What if the ascription of freedom of choice was in every case based on the availability of another desire? That is, children might start to 'pass the test' not because they start to think of themselves as endowed with absolute autonomy over their actions and desires, but rather because other desires—which are not implied by the story—start to become salient to them. Knowing that other desires are present could make children aware of the fact that, had they followed such *alternative desires*, they would have done otherwise.[3] If that were the case, we could no longer conclude that data from developmental studies support an unconditional interpretation of their freedom of choice.

So, even though both Nichols (2004) and Kushnir et al. (2015) conclude that children

[2]However, there were exceptions to this general pattern: In action stories, where children had to reason about another agent, the results did not significantly differ between the age groups. Additionally, in inhibition stories where children had to reason about themselves, neither the 6-year-olds nor the 4-year-olds performed significantly above chance level.

[3]Note that the term 'alternative desire' makes only sense from a post-choice point of view. Throughout the present paper, we use the term 'alternative desire' to refer to an agent's desire *which might have but did not* result in an action. The term can thus only be ascribed after a choice has been made.

aged five to six have an unconditional reading of the ability to do otherwise, we are skeptical about their conclusion. The children might have managed to answer the question simply because they were aware of the possibility of an alternative desire—a desire that might have guided their actions differently. If children in both studies were accessing an alternative desire, then children's answers are fully consistent with a conditional reading of the ability to do otherwise. To test this alternative hypothesis, we conducted an empirical study where we manipulated the availability of alternative desires systematically.

2.1 Manipulations and Hypotheses

In order to provide such a manipulation we adopted the design of Kushnir et al. (2015) and added similar cases in which two desires are implied by the story. If no significant differences are recorded between cases implying an alternative desire and cases not implying an alternative desire (H0), this would suggest that the availability of an alternative desire plays no a crucial role in children's ascriptions of freedom of choice. If, however, significant differences are shown (H1), we will have reason to surmise that the saliency of an alternative desire influences children's responses.[4] Additionally, such a result would provide support for the view that children hold a conditional understanding of freedom of choice, even though we cannot decisively reject the conclusion that children believe in agent causation based on such results.

In order to investigate the influence of the agent's perspective as well as the compatibility of desires, two additional manipulations were implemented: (1) Children had to answer questions not only about themselves, but also about another agent. Note that Kushnir et al. (2015) found that children endorse freedom of choice more often for another agent than for themselves. Given that the concepts involved in first-person ascriptions and third-person ascriptions are the same, however, we expect no differences between the two conditions (i.e., self/other agent). (2) The desires of the other agent were either compatible or incompatible with the child's own. Considering that an additional aspect (subjectivity of desires) has to be integrated, we expect there to be differences in response patterns, showing that incompatible cases are more difficult to understand. In order to investigate the developmental trajectory, we tested 4-, 5-, and 6-year-olds, generally expecting to replicate the increase in 'chose to' responses between four and six that were found by Kushnir et al. (2015).

[4]Note that we cannot exclude the possibility that even without implying a second desire, children make an implicit inference to such a desire. Consequently, we would not find any differences even though the availability of an alternative desire plays a crucial role. However, some studies have shown that it seems to be difficult for young children to reason about conflicting desires when one of those desires has to be inferred first (Cassidy et al., 2005).

2.2 Methods

2.2.1 Participants

Three age groups participated: 16 4-year-olds (M = 4.66 years; SD = .21), 16 5-year-olds (M = 5.47; SD = .29) and 16 6-year-olds (M = 6.24; SD = .26). All age groups were counterbalanced for sex. The children were recruited from five nursery schools in and around Bern (Switzerland).

2.2.2 Material

The material consisted of two drawings for the control questions and 14 photographs (10cm x 10cm) of different foods (seven generally liked and seven generally disliked foods). There were also four Playmobil toy characters (two female and two male) used for the control questions and the focal cases involving another agent (the exact phrasing of the focal questions, as well as some examples of the stimuli, can be accessed through this online repository: https://osf.io/gfut8).

2.2.3 Procedure

Children were interviewed individually in a separate room or quiet place in their nursery schools. After a short icebreaker talk the interview started with two warm-up questions, which also served as control questions. They were designed to test whether the child was able to reason adequately about questions concerning freedom of choice where no motivational constraints were involved. Two drawings were shown in succession (order counterbalanced): One drawing displayed a house located on an island, surrounded by ice-cold water, which could only be reached by a bridge. The other drawing showed a house located on land, with two routes leading to it. The child was then introduced to two toy characters (matched for gender), each living in one of the houses. The child was told that the toy characters were both on their way home after long exhausting days at work. The investigator took each toy character to the door of his or her house, taking the bridge to the island house and randomly choosing one of the routes to the other house. The child was then asked whether, in order to get home, the toy character *had to* take the specific route or the bridge (to the dry land or island house, respectively), or whether the character *could have chosen not to* take it (this was the same dichotomous answer format later used in the focal questions).

After answering these two control questions, the main body of the experiment started. It was structured as follows: First, the child was shown 14 food items (seven generally liked and seven generally disliked foods). Second, in order to adjust the focal questions, the child was invited to pick those two food items s/he liked most and those two s/he disliked the most (the order of choosing was counterbalanced). Third, the child was presented with six different cases featuring slightly different stories (order randomized). For each case the child answered a focal question (see Table 1 for two examples).

Table 1: Examples of two cases: c3 implies two desires, while c4 only implies a single desire. Since the agent is only allowed to take one food item, the two desires in case c3 are conflicting. Note that the food items were always matched to the taste (or distaste, in c5 and c6) of the children and shown by a picture. While in the two cases shown here (and in c5 and c6) the second part consists of a simple statement about the other agent, this was not the same in c1 and c2. In c1, children were asked which of the food items they wanted to take and in c2 *if* they wanted to take the only food item that was presented (all children answered this question affirmatively).

	c3	c4
1. Introduction	Imagine Simon is hungry and sees a cookie *and a candy*, which he both likes. His mum says it's ok to eat something sweet but he is only allowed to take one of the two.	Imagine Simon is hungry and sees a cookie which he likes. His mum says it's ok to eat something sweet.
2. Act	He takes a cookie.	He takes a cookie.
3. Focal Question	Did he *have to* take the cookie or could he have *chosen not* to take it, even though he likes it?	Did he *have to* take the cookie or could he have *chosen not* to take it, even though he likes it?

Table 2: Features of different cases. In some cases (c1, c3 and c5) the story implies two desires; in the other cases (c2, c4 and c6) only one. In c1 and c2, the children had to answer the focal question about themselves, in all other cases (c3, c4, c5, c6) about an other agent. In c4 and c5, the other agent held compatible desires, while in c5 and c6, the other agent held incompatible desires.

Case	**Second Desire**	**Agent**	**Congruence**
c1	yes	self	-
c2	no	self	-
c3	yes	other	yes
c4	no	other	yes
c5	yes	other	no
c6	no	other	no

In half of the cases two different food items were mentioned (c1, c3, and c5). The child, or the other agent respectively, was only allowed to take one of them. In the other cases, only one food item was mentioned (c2, c4, and c6). Thus, cases c1, c3, and c5 imply that the agent holds two desires which are conflicting because, for the given situation, only one can be fulfilled. From a post-choice perspective, there is an alternative desire involved in cases c1, c3 and c5, but not in cases c2, c4 and c6. While in c1 and c2, children had to answer the questions about themselves, in the remaining cases (c3 - c6) children had to

answer the questions about another person. This other agent was in each case one of the toy characters, already known from the control questions (one character for c3 and c4 and the other for c5 and c6). Additionally, the foods item in cases c1, c2, c3, and c4 referred to food items the child liked and those in cases c5 and c6 to those s/he disliked. The latter accounted for the agent having incompatible desires and thus required the children to integrate the aspect of the subjectivity of desires (see Table 2 for an overview of the different cases).

2.3 Results

To check whether children were able to adequately answer questions having the same format as the focal questions but not involving motivational constraints, we took a look at the two control questions. Almost all the children, 46 of 48 succeeded in answering those questions. Only two six-year-olds failed; this was due to insufficient familiarity with the language used in the case descriptions (German). They were excluded from further analysis (remaining $N = 46$).

Table 3: Number and percentages (in brackets) of 'choose to' responses, separated by age groups. Note that the complement to 100% of one cell reflects 'had to' answers (e.g. c1, 5-year-olds: 68.8% 'choose to' and 31.2% 'had to' responses). Asterisks mark cases in which the distribution of 'choose to' and 'had to' responses were found to be significantly different than expected under independency (Fisher's exact test).

Case	**4-year-olds (n = 16)**	**5-year-olds (n = 16)**	**6-year-olds (n = 14)**	**All Children (n = 46)**
c1*	8 (50.0%)	11 (68.8%)	13 (92.9%)	32 (69.6%)
c2	9 (56.3%)	8 (50.0%)	7 (50%)	24 (52.2%)
c3*	9 (56.3%)	10 (62.5%)	14 (100.0%)	33 (71.7%)
c4	8 (50.0%)	10 (62.5%)	7 (50.0%)	25 (54.3%)
c5	9 (56.3%)	8 (50.0%)	11 (78.6%)	28 (60.9%)
c6	9 (56.3%)	10 (62.5%)	11 (78.6%)	30 (65.2%)
All Cases	52 of 96 (54.2%)	57 of 96 (59.4%)	63 of 84 (75.0%)	172 of 276 (62.3%)

We first looked at the overall percentages of 'choose to' responses (see the last row in Table 3). Taking all cases together, 62.3% were answered by 'choose to'. Looking at the percentages within the different age groups, we observed an increase in 'choose to' responses with increasing age (and a decrease in 'had to' responses). This distribution of responses was found to be significantly different from the distribution expected under independency ($\chi 2(2) = 8.82$; $p = .012$). A look at the adjusted residuals showed that this was true only for the 4- ($prs = \pm 2.04$) and 6-year-olds ($prs = \pm 2.88$), but not the 5-year-olds ($prs = \pm 0.74$): while 4-year-olds gave fewer 'choose to' answers, 6-year-olds gave more 'choose to' answers than we would expect under independency. Moreover, the ratio of 'choose to' and 'had to' answers was not different from chance for the 5-year-olds. Summing up, this first overview suggests an increase in 'choose to' responses as a function of age and therefore, a successful replication of previous studies (Kushnir et al., 2015; Nichols, 2004).

Next, when considering the responses of all children (the rightmost column in Table 3), we noticed that not every case contributed equally to this result. In order to test our predictions, we looked at each case individually. We found that the distribution of responses dif-

Table 4: Pairwise comparisons between age groups using Fisher's exact test

Case	4- vs. 5-year-olds	5- vs. 6-year-olds	4- vs. 6-year-olds
c1	$p = .473$	$p = .175$	$p = .017$
c3	$p = 1$	$p = .019$	$p = .007$

fered from the distribution expected under independency only in cases c1 and c3 (Fisher's exact test c1: $p = .045$; c3: $p = .012$).

In order to see when (between which ages) significant differences emerged, we computed pairwise comparisons for cases c1 and c3 (see Table 4). This revealed that the response pattern of 4- and 6-year-olds was significantly different (Fisher's exact test, c1: $p = .017$; c2: $p = .007$). Moreover, in case c3, this was true for the 5- and 6-year-olds (Fisher's exact test, $p = .019$). That is, a significant change in response patterns was found only in cases in which a second conflicting desire was mentioned explicitly. Considering the percentages, this tells us that for the significant pairwise comparisons, the older children responded significantly more with 'choose to' answers than the younger ones. In all other cases (including c5, where a second conflicting desire was mentioned, but questions had to be answered regarding another agent holding incompatible desires) there was no significant change in response patterns as a function of age.

Finally, to get a more concrete description of the relationship between age and 'choose to' answers, odds ratios were calculated using binomial logistic regression. Whereas in case c1 an increase of one year resulted in a 3.1 times higher probability of a 'choose to' response ($B = 1.15$, $p = .017$, OR = 3.14, 95% CI [1.23, 8.02]), it resulted in a 3.4 times higher probability in c3 ($B = 1.23$, $p = .015$, OR = 3.42, 95% CI [1.27, 9.17]).

2.4 Discussion

This study was designed to check whether the availability of an alternative desire has an impact on children's ascriptions of free choice. Indeed our results suggest that this is the case. While we observed a significant increase of 'choose to' responses in cases where a second conflicting desire was mentioned explicitly, this was not true for cases without an explicit mentioning of a second conflicting desire.[5]

Our results provide evidence for hypothesis (H1) and against hypothesis (H0), according to which there are no significant differences between cases implying an alternative desire and cases implying no such desire. On the assumption that children entertain an unconditional interpretation of 'could have done otherwise', we would expect to find that the availability of alternative desires does not influence children's responses: we would expect them to indicate that an agent can simply choose to act against his or her desire, whether or not an alternative desire is implied. But, it turns out, awareness of alternative desires matters. It is therefore no longer plausible to conclude that children entertain an unconditional understanding of the phrase 'could have done otherwise'.

[5]This did not hold for cases in which questions were asked about another agent holding incompatible desires. This suggests that the integration of incompatible desires sets an additional degree of difficulty and is not—at least not fully—achieved at the age of six.

Of course, it might still be the case that children think of themselves as agents with self-originating causal powers. This conjecture is most plausible if adults are indeed agent causationists; at some stage, agent causationist intuitions will become dominant and measurable. When it comes to adults, the debate is still not settled. However, most studies suggest adults to be agent causationists (Nichols, 2006; Nichols & Knobe, 2007; Sarkissian et al., 2010). Therefore, in a second study we investigated adults' intuitions about agent causation.

3 Study 2: Adults and Agent Causationism

As adults, we don't hold the intuition that each and every desire of ours impairs our freedom of choice. Of course, in some sense, our (at least partial) autonomy from desires is a trivial matter: If we hold two conflicting desires, both desires cannot be implemented simultaneously. For instance, when an adult actively chooses not to eat a cookie, even though she would really like to eat it, then her choice is usually motivated by a further desire—perhaps a desire to eat healthy. In other words, if the agent's desire to eat healthy is stronger than her desire to eat the cookie, it follows that her desire to eat the cookie did not constrain her choice. However, a stronger claim can be made according to which agents are not only free to resist acting upon weaker desires, but also capable of acting against their strongest desires and interests. The existence of such an ability would provide support for an unconditional reading of the ability to do otherwise, and, a fortiori, for the claim that we conceive of ourselves as agent causationists, who can prevent even our strongest desires from ruling our actions.

As mentioned above, the standard way to investigate adults' intuitions about agent causation features a 'can do otherwise' or 'could do otherwise' question. The exact meaning and ambiguity of the phrase has been debated for centuries. For Hobbes, the idea that an agent could do otherwise was a contradiction: it would be equivalent to saying that a cause is necessary and sufficient for a certain effect but this effect does not necessarily follow its cause (Hobbes, 1839/1646). Others argued that it is true but trivial that any given agent could have done otherwise if that agent had chosen to do otherwise—but that it's not possible for an agent to have chosen otherwise given the exact same circumstances (see for example Schlick, 1939). It has also been claimed that even if we accept that an agent could have done otherwise given the exact same circumstances, this would lead to the conclusion that such a choice is arbitrary and irrational (Kane, 1985, Double, 1990), given that we want our psychological circumstances (such as character traits, reasoning processes, and motives) to account for our choices.

The experimental-philosophical literature has made giant leaps in furthering our understanding of our intuitions concerning free will, moral responsibility, and determinism. A large chunk of this research has focused on whether laypeople are natural compatibilists or natural incompatibilists (Nichols and Knobe, 2007; Nahmias, Coates, and Kvaran, 2007; Feltz, Cokely, and Nadelhoffer, 2009; Murray and Nahmias, 2014; and many others). The

approach these papers take is different from ours in presupposing a deterministic universe to find out whether people believe free will and moral responsibility to be compatible with such a universe. In contrast, we investigate if laypeople believe they could have done otherwise in exactly the same circumstances. A negative answer would suggest that the folk believe that their choices are determined. In other words, determinism would follow from our studies and is not already presupposed.

Other experimental studies have indeed grappled with a similar problem that we try to tackle in this paper. Nichols (2006) asked participants to rate whether a conditional or an unconditional analysis of 'could have done otherwise' sounded more reasonable.[6] The results clearly favored an unconditional analysis. Nichols & Knobe (2007) asked participants whether they thought they lived in a universe in which our decisions were either completely caused by the past or not. Again, a majority of people seem to have indeterminist intuitions. Sarkissian et al. (2010) demonstrated the cultural universality of responses to the scenarios from Nichols & Knobe (2007). Knobe (2014) suggests that the current state of experimental-philosophical research favors a view according to which we view ourselves as beings who "transcend the whole causal order" (2014, p.70).

While these results are impressive, we believe an important aspect has so far been neglected. Even if the circumstances are held constant, and participants understand that aspect of the scenarios given, it is unclear whether the question at stake aims at a person's *ability* or at the *possibility* of implementing the ability to do otherwise. That is, even if participants claim that an agent could have done otherwise if everything had been exactly the same, it is not clear whether they mean that given these circumstances, the agent has the ability to follow a different desire, or that given the actual circumstances, it is possible to do otherwise.

In order to illustrate our point, consider the following case, which does not involve desires. A pro surfer called Jimmy faces an ocean devoid of waves. Can he catch a wave? Well, he is a pro surfer and has—as a matter of fact—the ability to catch waves, so it seems perfectly fine to state that he can catch a wave. But on the other hand, there is no possibility of him catching a wave on this particular day because the enabling condition—there being waves—is not met by the actual circumstances. To show the empirical adequacy of this distinction, we conducted a small-scale study, involving 102 participants, to confirm that response patterns indeed differed when the question type was manipulated. Given the scenario described above, participants disagreed with the claim "It is possible that Jimmy catches a wave" (M= -1.39,SD = 1.78), but they tended to agree that "Jimmy has the ability to catch a wave" (M= 0.2, SD= 2.55) (Ratings were obtained using a 7-point Likert Scale anchored at '-3' (Totally Disagree), '0' (Neutral), and '+3' (Totally Agree)).[7]

These results suggest that people indeed interpret questions about possibility and ability differently. Questions about ability ask whether there exists at least one set of circum-

[6]More specifically, Nichols asked participants to consider that Bill lied about his income when filling out his tax form. Subsequently, he inquired how right or wrong it sounded that "Bill could have decided to be honest at 10:30, 4/13/2005, but only if some things [even if nothing] had been different before the moment of his decision."

[7]For details about the methods of this study please go to the online repository: https://osf.io/gfut8/

stances under which an agent can perform a certain action. That set of circumstances is largely independent of the actual circumstances. Questions about possibility, on the other hand, ask whether, given the actual circumstances, an agent is in the position to perform a certain action.

When investigating intuitions about agent causation, we are interested in whether people think an agent could have done otherwise given a fixed set of circumstances. Therefore, we should ask participants whether it is possible for an agent to do x, but not whether he or she has the ability to do x. 'Can' questions are often interpreted to align themselves with an ability reading, also known in the literature as circumstantial possibility (Kratzer, 1991). However, 'can' questions are ambiguous between these meanings. It remains unclear what exactly adults mean when they state that an agent could have done otherwise. In order to resolve this ambiguity, we conducted a second empirical study, in which we investigated the effect of different modal formats on adults' intuitions about freedom of choice.[8]

3.1 Manipulations and Hypotheses

We designed a questionnaire study, in which we systematically manipulated the wording of certain statements. Three groups of participants read a short story in which an individual decides to take an action. They were then asked to rate a statement (a 'can', an 'ability', or a 'possibility' statement) about the agent's decision to take that action. We hoped thereby to assess whether people think of 'can' as indicating the ability to do otherwise in general or the capacity to perform a different action under the very same circumstances. That is, we intended to clarify whether adults interpret 'could have done otherwise' conditionally or unconditionally, and thus whether adults tend to hold agent causationist intuitions or not.

If 'can' is interpreted as indicating ability, we would expect agreement ratings not to differ significantly between 'ability' and 'can' statements. However, if 'can' is interpreted as indicating possibility, we would expect agreement ratings not to differ significantly between 'possibility' and 'can' statements. Furthermore, if people are agent causationists, we would expect high agreement ratings for both the 'ability' and the 'possibility' statements. That is, ratings should indicate both that the agent has the general ability to do otherwise and that, given the current circumstances, it is actually possible for her to do otherwise. Therefore, independent of whether the 'can' statement is interpreted as a statement about ability or a statement about possibility, ratings should be significantly above the midpoint for this statement as well. If people do not have agent causationist intuitions, we would expect average ratings for the 'possibility' statement to be significantly below the midpoint.

[8]While we are not aware of any studies in the free will literature that have questioned the effects of different modal formats, some authors have recently started to investigate different modal interpretations when it comes to the ought-implies-can principle (see, e.g., Turri, 2017; Kürthy, Del Prete, and Barlassina, ms; Willemsen and Wiegmann, ms)

3.2 Methods

3.2.1 Participants

218 participants were recruited online via Prolific and randomly assigned to one of three conditions (ability, possibility, can). An a priori power analysis with an effect size of 0.6 and a p-value of 0.05 yielded a total sample size of 183 to reach a power level of 0.8. Because we expected some dropouts and did not know how many participants would fail the attention check (described below), we set the number of participants to 200 (usually, slightly more participants are then recruited given the way Prolific operates). There was a dropout rate of 5.5% (n = 11). Additional participants were excluded either because they did not indicate English as their first language (n = 10) or because they failed to pass the attention check (n = 42). The remaining sample (n = 155, M_{Age} = 32.98, SD = 12.43) consisted of 65% females and 35% males. The experimental design, predictions, and statistical tests were preregistered with the Open Science Framework: https://osf.io/mcpbr.

3.2.2 Questionnaire

After participants were asked for consent they were instructed to read a short story and then rated a provided statement. The short story featured an agent who desires to move her belongings from one desk to another:

> Sarah works in an open office space and moves her belongings from one desk to another. She is almost finished. On her old desk are only her printer and a pile of books. Sarah desires to move the pile of books and she desires to take the printer to her new desk. Her desire to take the pile of books is greater than her desire to take the printer. So she chooses to take the pile of books and moves them to her new desk.
>
> Now, imagine we can turn back time to the point where Sarah makes her decision. Everything is exactly the same as before: Sarah has the same two desires and no other desire: She wants to take the pile of books and she wants to take the printer. Sarah's desire to take the pile of books is stronger than her desire to take the printer.[9]

After reading this short vignette, participants were presented with one of the following three claims and asked to rate how much they agreed with it:

- *Ability Condition*: Sarah has the ability to choose to move the printer first.
- *Possibility Condition*: It is possible that Sarah chooses to move the printer first.

[9] It would have been interesting to use similar vignettes in Study 1 & Study 2 that would allow for comparisons between responses in both studies. However, the questionnaire in Study 1 was designed to match the structure and wording of Kushnir et al., 2015 as close as possible. In contrast, in Study 2, our aim was to use a scenario that would allow for the different modal questions, and also to describe a situation that would be very unlikely to involve any emotional agitation.

- *Can Condition*: Sarah can choose to move the printer first.

Ratings were obtained using a 7-point Likert Scale anchored at -3 (Totally Disagree), 0 (Neutral), and +3 (Totally Agree). After answering the focal question, participants were asked to qualitatively explain their answers in a short sentence. In order to check whether the short story was read carefully, participants were also asked to indicate how many desires were involved in Sarah's decision. If participants indicated that either more or fewer than two desires were involved, they were excluded from the analysis.

3.3 Results

Participants in the *Ability* Condition agreed most strongly with the rated statement (M = 2.69, SD = .79). In the *Can* Condition, participants generally agreed with the statement (M = 1.81) but showed greater variability in their ratings (SD = 1.71). Crucially, participants in the *Possibility* Condition tended to disagree with the statement (M = -0.56, SD = 2.04). Figure 1 shows the mean results of the three conditions; Figure 2 depicts the distribution of responses grouped by different conditions.

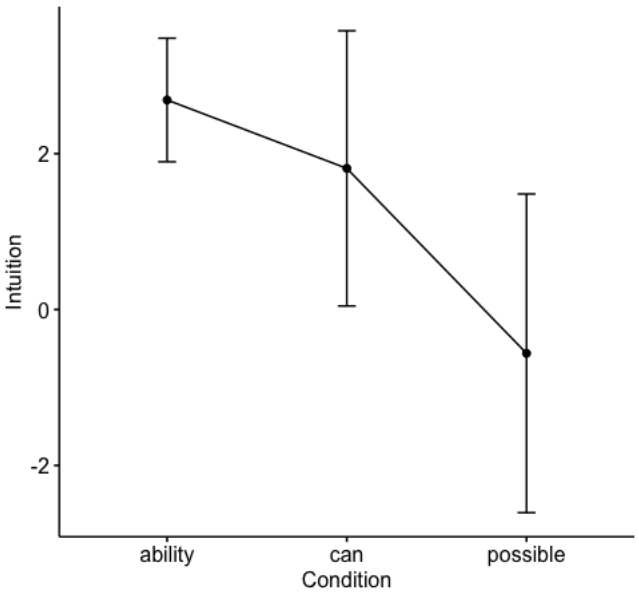

Figure 1: Means of different conditions. Error bars represent standard deviations.

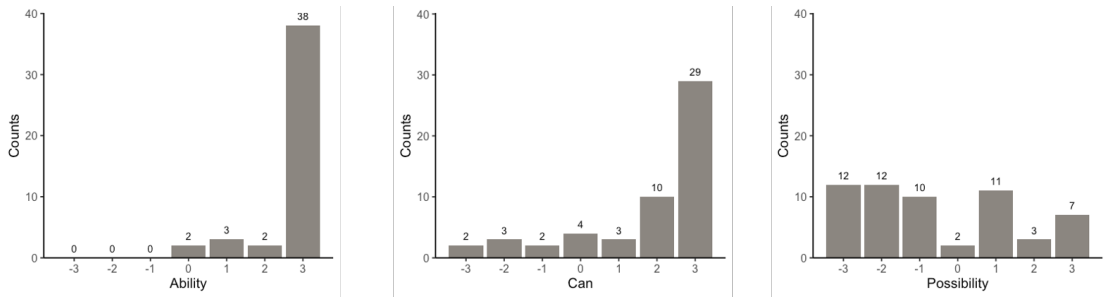

Figure 2: Distribution of responses across conditions.

Because a Levene's test indicated unequal variances ($F(2, 151) = 13.804$, $p < .001$) and a Shapiro-Wilk test showed a significant departure from normality ($W = .946$, $p < .01$) we used Kruskal–Wallis one-way analysis of variance to check whether the means of the conditions differed significantly. Employing this non-parametric test we found a significant effect of treatment (X^2 (2) = 65.37; $p < .0001$). To examine how different mean differences contributed to this result we used Dunn's test for pairwise post-hoc comparisons. While the *Ability* and *Can* Conditions differed significantly from the *Possible* Condition (both $p < .01$), this was not the case for the comparison between the *Ability* and the *Can* Condition ($p = .2$).

As a next step we investigated whether agreement ratings for the three conditions were significantly different from the midpoint of 0. Because data was not normally distributed within conditions (Shapiro-Wilk's test yielded *Ability*: $W = .447$, $p < .001$; *Can*: $W = .712$, $p < .001$; *Possibility*: $W = .885$, $p = .039$), we used the Wilcoxon signed rank test to evaluate whether the observed differences from the neutral midpoint of 0 were significant. This analysis showed that the means in the *Ability* and *Can* Conditions were significantly above 0 (both $p < .01$), whereas the mean of the *Possibility* Condition was significantly below 0 ($p = .042$). In other words, participants in the *Ability* and *Can* Conditions agreed with the rated statement, while participants in the *Possibility* Condition disagreed.

3.4 Discussion and Possible Objections

Having agent-causationist intuitions would inevitably include the belief that an agent has the ability to do otherwise as well as the belief that, given the current circumstances, it is possible that the agent does otherwise. However, our results show that there is a significant difference between ratings for 'ability' vs. 'possibility' statements. While people agreed that the agent has the ability to do otherwise, they tended to disagree that it is possible that the agent does otherwise given the actual circumstances. Thus, our results suggest that most adults don't hold agent causationist intuitions. Admittedly, the large variance of responses indicate that people may disagree on the question of whether it is possible to do otherwise in such situations. Unfortunately, our study was not designed to investigate the source of the large distribution of responses.

Furthermore, the absence of a significant difference between ratings for the 'can' vs. the 'ability' statement combined with the significant difference found between ratings for 'the 'can' vs. the 'possibility' statement suggest that 'can' questions are interpreted as questions about ability and not about possibility. This means that positive ratings for 'can' statements do not provide information about whether people think an agent could have done otherwise given the actual circumstances. They instead target the general capacity of an agent to choose another desire independent of the actual circumstances. Therefore, 'can' questions seem unsuitable for determining whether people hold agent causationist intuitions or not.

In order to avoid this conclusion, one might object that the 'possibility' statement was not interpreted by participants as we intended. Instead, the 'possibility' statement might

have been interpreted as a statement about the likelihood of the action in question, i.e., when participants stated that it is not possible that Sarah chooses to move the printer first, they actually meant that it is not very likely that Sarah chooses to move the printer first. Looking at the qualitative responses we collected, this alternative explanation seems rather implausible. Out of 34 responses disagreeing with the 'possibility' statement (responses of '-3', '-2', and '-1'), only a single person explained her response in reference to likelihood. All other responses do not allow for such an alternative interpretation.

One might also argue that whereas the 'ability' and 'can' statements have a natural agential reading, this is not the case for the 'possibility' statement, which participants might have interpreted in terms of physical possibility at the level of events and state of affairs.[10] Such a physical reading might be triggered because the agent's name 'Sarah' was not in the subject NP position as in the other statements. Furthermore, statements phrased in the form 'it is possible that' might be most frequently used when talking about events and less so when talking about agents. Accordingly, disagreeing with the 'possibility' statement would no longer speak against agent causationism. Again, even though some of the qualitative explanations are neutral in regards to how participants interpreted the main statement, the majority of responses do not support such an alternative reading (we have uploaded the complete data file in the online repository: https://osf.io/gfut8, including participants' responses). Additionally, we take the design of our Study 2 to be analogous to the preliminary study about Jimmy the surfer, for which we have shown that participants have the tendency to affirm the 'possibility' statement only if favorable circumstances hold.

4 General Discussion

While the metaphysical account of agent causation has only few followers, most scholars agree that we *think* of ourselves as agent causationists. The popularity of that latter view is not surprising. While free will might well be an illusion (Wegner, 2002), the illusion itself provides initial support for the idea that we think of ourselves as agent causationists: If we *feel* or *perceive* that we are able to start new causal chains, then perhaps we also *conceive* of ourselves as beings with such (mystical) powers. As such, previous experimental studies—like those we discussed above—that suggest that children and adults think of themselves as agent causationists, struck a chord with many scholars

However, perceiving and conceiving are two different pairs of mental shoes. We perceive the Müller-Lyer lines as being different in length, but conceive of them as having the same length. Thus, we need additional, independent evidence that shows that people *conceive* of themselves as agent causationists. We started this paper by highlighting two such strands of evidence. We then conducted two studies to investigate possible flaws in those strands. In this General Discussion, we first assess our results in light of the current debate. We will then discuss the impact of our results for the metaphysical account of agent

[10] We would like to thank (omitted for anonymous reviewing) for suggesting this alternative reading to us.

causation.

4.1 The 'New' Empirical Situation

Developmental research on children aged four to six suggests that around the age of five, children gain the conceptual sophistication required to state that they did not have to act a certain way but could have chosen otherwise. This achievement leaves open a crucial question: How are we to interpret children's cognitive development? As we claimed in the introduction, (at least) two interpretations are possible. According to the unconditional interpretation, children believe that they could have done otherwise even if the situation were exactly the same. This unconditional reading is very much in line with the agent causation model, which states that agents can interfere in the causal process that leads from desire to action. Both Nichols (2004) and Kushnir et al. (2015) are sympathetic to the agent causation model. They argue that the empirical evidence they have collected from their research is best explained by this model. The agent causation model is, however, not the only game in town. According to the conditional interpretation, children believe that they would have done otherwise if they'd had a different desire. Thus, when children contemplate whether they could have done otherwise, they answer affirmatively, because they believe that had they entertained an alternative desire, they would have acted differently.

Our own results indicate that by the age of six, almost every child succeeds at the given task as long as two conditions are met: (i) at least two conflicting desires have to be mentioned explicitly, and (ii) the two conflicting desires have to be compatible with the child's own desires. These results do not square easily with the idea that children's ability to understand freedom of choice is best explained via the agent causation model. This model does not predict that explicit reference to alternative desires facilitates children's conceptual transformation, nor that children's responses depend on the compatibility of the stated desire with the child's own preferences. After all, if children believe that they can do otherwise by intervening in the causal flow from desire to action, alternative desires should not play a crucial role in that process. As we have shown, however, the availability of alternative desires seems vital in prompting an affirmative response to the question "Could you have done otherwise?". This suggests that children believe they could have done otherwise only if an alternative desire had been stronger than the actual desire was. If true, this would provide support for the conditional interpretation.

Admittedly, the data we obtained do not rule out the possibility that children harbor agent-causationist intuitions on free will. Perhaps the two conditions that seem to be required for six-year-olds to consistently answer such questions affirmatively (to wit: the availability of explicitly stated alternatives and the compatibility of the desires under consideration with the child's own desires) foster agent-causationist thinking. For instance, it is indeed possible that, in the absence of an explicitly stated second conflicting desire, a young child is not sufficiently motivated to think of herself as an agent who can resist her desires. Nonetheless, our results put considerable pressure on those arguing that developmental data tells in favor of the agent causation model. At a minimum, our results open

up the possibility that children reason conditionally about free will under motivational constraints.

Several empirical studies on adults' intuitions on free will suggest that adults conceive of themselves as agent causationists (Nichols, 2006; Nichols and Knobe, 2007; Sarkissian et al., 2010). In fact, if we were to focus only on those parts of our second study in which we asked participants to tell us whether Sarah is able to (or can) choose to follow an alternative desire in a situation in which 'everything is exactly the same as before', an overwhelming majority of participants agree that Sarah can do so. Taken in isolation, this outcome favors the agent causation model. However, people's responses change dramatically if the question is asked differently. Most people disagree with the suggestion that *it is possible* that Sarah chooses an alternative desire. How can these contrasting results be explained? As our study on ability and possibility shows, people interpret ability claims to mean something akin to 'possible if the circumstances are favorable': a professional surfer is able to catch waves even if there are no waves, because if the situation were favorable (i.e. if some waves emerged), it would be possible for him to catch them.

A similar understanding of ability and possibility seems to hold for free will intuitions. If the situation favors an alternative desire, then we can follow that alternative desire. Consequently, people conceive of themselves as able to follow alternative desires. However, and crucially, people do not think it is possible for an agent to have followed an alternative desire if that desire has (in fact) proven weaker than another. If, in a hypothetical situation, everything is held exactly the same as in a (hypothetical or real) past situation in which an agent ate a cookie, then people do not think that it is possible for the agent to have chosen not to eat the cookie.

Our studies show that the ability question is clearly tied up with a conditional understanding of free will. Most (if not all) people believe they could have acted differently if the situation had been different. Unsurprisingly then, these results show support for a conditional understanding of free will. However, philosophers who argue for the agent causation model need more than just the ability to do otherwise — they are after an unconditional understanding of free will, which our studies suggest can be tracked by asking whether *it is possible* to do otherwise in the exact same circumstances. By and large, people do not think so. In other words, our studies suggest that people do not think we are the sorts of agents who have causal powers to prevent our strongest desires from happening.

Interestingly, questions about whether people *can* behave differently are largely interpreted as questions about ability, not possibility. The agreement ratings for the *Can* condition were slightly lower but not significantly different from those for the *Ability* condition. Note that we do not claim that this result can be generalized across a wide range of scenarios. Whether 'can' is read as 'ability' or 'possibility' likely depends on a number of factors, for instance, whether a certain 'can' question is more frequently raised in situations in which actual possibility is at stake or, alternatively, in situations relating to a person's ability and skills. In the scenarios we investigated, the 'can' formulation is more often interpreted as a question about ability, not possibility. This has important consequences.

When researchers design experiments (whether real experiments or thought experiments) featuring such scenarios, they should be careful to ask the right questions. Thus, we recommend that researchers use the 'possibility' question instead of the 'can' question or the 'ability' question.[11]

4.2 Folk Agent Causationism & Agent Causation

At the beginning of this paper, we highlighted the importance of distinguishing between folk agent causationism—the view that people think of themselves as agents who can start and prevent causal chains independently from the causal chains of events—and agent causation—the view that agents, regardless of how they think about their agency, can start and prevent such causal chains. It is fully consistent to endorse the former and reject the latter view. In fact, probably the majority of scholars hold that folk agent causationism is true, while agent causation is false. This position is often made plausible by the different perspectives that we can take in regards to agents. From a third-person perspective there is little evidence that agents are causally relevant beyond the physical processes that determine the workings of agents. Hence, agent causation is wrong. In contrast, from a first-person perspective it seems to us that we are not at the mercy of those physical processes but actually determine our own faith. Hence, agent causationism is true.

While there is no deductive link between agent causationism and agent causation, proponents (and some opponents) of agent causation state that folk agent causationism provides a strong motivation and a reason for agent causation. O'Connor, for instance, states that "the agency theory is appealing because it captures the way we experience our own activity. It does not seem to me (at least ordinarily) that I am caused to act by the reasons which favor doing so;" (O'Connor, 1995, 1996).

We think it is not just the "experience of of our own activity", but the putative *actual belief* in free agent activity, that provides the strongest motivation for agent causation. Campbell nicely connects the experience-claim with the belief-claim:

> "Let us ask, why do human beings so obstinately persist in *believing* that there is an indissoluble core of purely self-originated activity which even heredity and environment are powerless to affect? There can be little doubt, I think, of the answer in general terms. They do so, at bottom, because they *feel* certain of the existence of such activity from their immediate practical experience of themselves." (1967, p.41, our italics)

If Campbell were right, then most scholars actually believe—in the unguarded moments of life that comprise almost all our cognitive activity—that they can start new causal chains, but reject this ability only from a third-person perspective: "So far as we confine ourselves

[11] Such a recommendation might not hold for developmental studies. In our Study 1, we used a 'can' question to get at children's intuitions about free will. Arguably—though we lack the data to support this—it is easier for young children to understand a sentence like 'Can you do this,' as opposed to 'Is it possible for you to do this?'.

to external observation, I agree that this notion must seem to us pure nonsense." (Campbell, 1967, p.48).

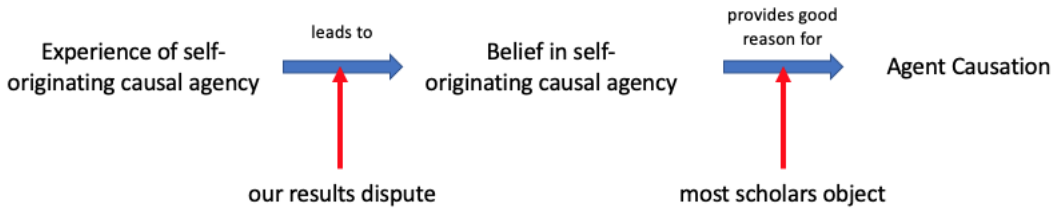

Figure 3: The results of Study 2 threaten a successful inference from the experience to the belief in causal agency.

Our results suggest that this view is mistaken. While perhaps most people may feel the existence of self-originated activity, it would be wrong to infer that those people also believe in such activity. In Figure 3, we depict the relation between the claims at stake. Many scholars object to the claim that belief in self-originating causal agency provides a good reason for agent causation. Independently of whether they are successful, our results dispute that people even entertain the belief in self-originating causal agency. Consequently,our results also provide evidence against the metaphysical account of agent causation, in the sense that if we can show that one of the main motivations for agent causation falls apart, advocating agent causation becomes that much harder.[12]

5 Conclusion

The present studies offer a detailed look at how different features influence our intuitions about choice under motivational constraints. The data of Study 1 suggest that children might reason conditionally about free will: six-year-olds succeed in consistently answering affirmatively that they could have done otherwise only if at least two conflicting desires are implied, which are compatible with their own desires. We also demonstrated (Study 2) that adults are likely not to conceive of themselves as agent causationists. When participants were questioned about the possibility of having done otherwise (rather than their ability to have done otherwise), they denied, in the main, that any such possibility exists.

[12]While the results of both studies suggest that people by and large do not conceive of themselves as agent causationists, our results do not allow us to draw a decisive conclusion against folk agent causationism in all its varieties (for different ways to cash out agent causation, see, e.g., Clarke, 1993; O'Connor, 1996). In fact, folk agent causationism can provide a compelling response. Two different versions of agent causationism need to be kept apart. First-order agent causationism is the view that agents can directly intervene between a desire—like a desire to eat a cookie—and a subsequent action. Second-order agent causationism allows agents to select which of their first-order desires (e.g., a desire to eat a cookie vs. a desire to eat healthy food) is effective in causing an action. The results of our empirical studies are consistent also with second-order agent causationism. In future studies, we aim to investigate the role of second-order desires in free will intuitions to find out whether a more sophisticated version of folk agent causationism stands a better chance of accounting for people's intuitions.

Acknowledgement

The authors would like to thank Luca Barlassina, Beat Huber-Eicher, Joshua Jäger, Miklos Kurthy, Romano De Maddalena, Shaun Nichols, Louis Oberli, Fabio Del Prete, Pascale Willemsen, and an anonymous reviewer for very helpful discussions and comments on earlier versions of this manuscript. An earlier version of this paper was presented at the Experimental Philosophy Conference in Bern 2019. We thank the participants for the valuable feedback.

References

Campbell, C. A. (1967). *In defence of free will, with other philosophical essays*. London: Allen Unwin.

Cassidy, K. W., Cosetti, M., Jones, R., Kelton, E., Meier Rafal, V., Richman, L., & Stanhaus, H. (2005). Preschool children's understanding of conflicting desires. *Journal of Cognition and Development*, *6*(3), 427–454.

Chernyak, N., Kushnir, T., Sullivan, K., & Wang, Q. (2011). A comparison of Nepalese and American children's concepts of free will. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 33, *33*).

Clarke, R. (1993). Toward a credible agent-causal account of free will. *Noûs*, *27*(2), 191–203.

Double, R. (1990). *The non-reality of free will*. Oxford University Press.

Feltz, A., Cokely, E. T., & Nadelhoffer, T. (2009). Natural compatibilism versus natural incompatibilism: Back to the drawing board. *Mind & Language*, *24*(1), 1–23.

Hobbes, T. (1839/1646). The questions concerning liberty, necessity, and chance. In S. W. Molesworth (Ed.), *English works of Thomas Hobbes, vol. V*. London: Routledge.

Kane, R. (1985). *Free will and values: Adaptive mechanisms and strategies of prey and predators*. SUNY Press.

Knobe, J. (2014). Free will and the scientific vision. In. O. Machery E. (Ed.), *Current controversies in experimental philosophy*. London: Routledge.

Kratzer, A. (1991). Modality. In A. von Stechow D. Wunderlich (Ed.), *Semantics: Aninternational handbook of contemporary research*. Berlin: Walter de Gruyter.

Kürthy, M., Del Prete, F., & Barlassina, L. (ms). 'Must' implies ‚can'.

Kushnir, T., Gopnik, A., Chernyak, N., Seiver, E., & Wellman, H. M. (2015). Developing intuitions about free will between ages four and six. *Cognition*, *138*, 79–101.

Lane, J. D., Ronfard, S., Francioli, S. P., & Harris, P. L. (2016). Children's imagination and belief: Prone to flights of fancy or grounded in reality? *Cognition*, *152*, 127–140.

Murray, D., & Nahmias, E. (2014). Explaining away incompatibilist intuitions. *Philosophy and Phenomenological Research*, *88*(2), 434–467.

Nahmias, E., Coates, D. J., & Kvaran, T. (2007). Free will, moral responsibility, and mechanism: Experiments on folk intuitions. *Midwest studies in Philosophy*, *31*, 214–242.

Nichols, S. (2004). The folk psychology of free will: Fits and starts. *Mind & Language*, *19*(5), 473–502.

Nichols, S. (2006). Free will and the folk: Responses to commentators. *Journal of Cognition and Culture*, *6*(1-2), 305–320.

Nichols, S., & Knobe, J. (2007). Moral responsibility and determinism: The cognitive science of folk intuitions. *Nous*, *41*(4), 663–685.

O'Connor, T. (1995). Agent causation. In T. O'Connor (Ed.), *Agents, causes, and events: Essays on indeterminism and free will*. Oxford: Oxford University Press.

O'Connor, T. (1996). Why agent causation? *Philosophical Topics*, *24*(2), 143–158.

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

Schlick, M. (1939). *Problems of ethics*. New York: Dover Publications.

Turner, J., & Nahmias, E. (2006). Are the folk agent-causationists? *Mind & Language*, *21*(5), 597–609.

Turri, J. (2017). How "ought" exceeds but implies "can": Description and encouragement in moral judgment. *Cognition*, *168*, 267–275.

Wegner, D. M. (2002). *The illusion of conscious will*. MIT press.

Wente, A. O., Bridgers, S., Zhao, X., Seiver, E., Zhu, L., & Gopnik, A. (2016). How universal are free will beliefs? cultural differences in chinese and us 4-and 6-year-olds. *Child development*, *87*(3), 666–676.

Willemsen, P., & Wiegmann, A. (ms). I must although I can't!? Suggestions for a two-level theory of 'ought implies can'. doi:10.31234/osf.io/hyq9u