(1) **Title**: Responsibility attribution about mechanical devices by children and adults.

(2) **Keywords**: Responsibility attribution, pivotality, criticality, development of causation, counterfactual thinking

(3) **Corresponding author**: Cristina Gordo Gordo

Address: Universidad de Jaén. Departamento de Psicología. Campus Universitario Las Lagunillas, S/N, Jaén, (Spain).

Telephone number: +3468062773
E-mail: cggordo@ujaen.es
Orcid: https://orcid.org/0000-0003-4934-5838

(4) **Co-author:** Jesica Gómez-Sánchez

Address: Facultad de Psicología, Campus Universitario de La Cartuja, S/N, 18071, Granada, (Spain)

Telephone number: +3468062773

E-mail: gomezjs@ugr.es
Orcid: https://orcid.org/0000-0002-0247-1012

(5) **Co-author**: Sergio Moreno-Ríos

Address: Facultad de Psicología, Campus Universitario de La Cartuja, S/N, 18071, Granada, (Spain)

Telephone number: +34958240676

E-mail: semoreno@ugr.es

Orcid: https://orcid.org/0000-0001-5553-207X

(6) This research was funded by the Spanish Government, Ministry of Economy and Competitiveness (PGC2018-095868-B-I00) and by the Junta de Andalucía - Conserjería de Universidad, Investigación e Innovación - Project (P21_00073).

(7) Disclosure: The authors have no relevant financial or non-financial relationships to disclose. They declare that they have no conflict of interest.

**Responsibility attribution about mechanical devices by children and adults**

## Abstract

We investigated the causal responsibility attributions of adults and children to mechanical devices in the framework of the criticality-pivotality model. It establishes that, to assign responsibility, people consider how important a target is to reaching a positive outcome (criticality) and how much the target contributed to the actual outcome (pivotality). We also tested theoretical predictions about relations between the development of counterfactual thinking and assessments of pivotality. In Experiment 1, we replicated previous findings in adults using our task. In Experiment 2, we administered this task and a brief counterfactual reasoning questionnaire to children aged between 8 and 13 years. Results showed that children also considered both criticality and pivotality when they attributed responsibility. However, older children were more sensitive than younger ones to pivotality. Also, we found a positive correlation between children's pivotality judgements and a measure of counterfactual thinking. Results are discussed regarding the model's relation to counterfactual thinking.



Keywords: Responsibility attribution, pivotality, criticality, development of causation, counterfactual thinking

# Introduction

Imagine the following situation: 11 members of a committee must vote for either candidate A or B. The members vote and candidate B loses 6-5. Now imagine a different situation where candidate B also loses but in this case 11-0. Sarah is a member of the committee who voted for candidate A in both situations. Is Sarah responsible for B's defeat in both cases? And if so, is Sarah equally responsible in both? Although most of us would agree that Sarah is responsible for the defeat in both cases, we intuitively attribute more responsibility to Sarah in the situation where candidate B loses 6-5 than in the one where B loses 11-0. This might be because only in the former situation would B have won if Sarah had voted for her/him. This example, offered by Chockler and Halpern (2004), illustrates the close relation between responsibility judgments, causality and the ability to think how things could have turned out differently (counterfactual thinking).

In the present work, we will test children's and adults' causal attribution to the components of a battery for its failure. We are particularly interested in testing developmental changes in children's causal attribution. We will explore them in the framework of a novel model of causal responsibility attribution: the *criticality-pivotality model* (henceforth, CP model) proposed by Lagnado and colleagues (2014, Lagnado & Gerstenber, 2015, 2017). We will start by reviewing some important models of causal attribution and the role of counterfactual thinking in them. We will then review previous results regarding the development of counterfactual thinking.

## Theories of causation

Many traditional theories of causation have tried to explain causality based on counterfactual models. The standard counterfactual model of causation (Lewis, 1973, 1986), for example, defines actual causality in terms of counterfactual dependence.

Thus, to establish that a cause “X” causes an outcome “Y”, there must be a counterfactual dependence between “X” and “Y”. That is, if “X” had not occurred, then “Y” would not have happened. In the voting example, only the situation where B loses 6-5 passes this counterfactual test: only here, if Sarah had voted for B, would the outcome have been different (B would have won). Conversely, in the 11-0 situation, even if Sarah had voted for B, the outcome would have been the same. Thus, in this case it is not possible to establish a causal relation between Sarah’s vote and B’s defeat, given that there is no counterfactual dependence between them. However, although Lewis’s (1973, 1986) model does not predict it, adults also intuitively established a causal relation in this situation. The voting example illustrates one of the main limitations of the standard counterfactual model of causation (Lewis, 1973, 1986): the over-determination problem. The model fails to explain causality in overdetermination cases, that is, in situations where multiple agents contribute to the same outcome.

The structural account of causation (e.g., Pearl, 2009; Halpern & Pearl, 2005; Woodward, 2003) has tried to overcome this and other problems of the standard account (see Collins et al., 2004; Moore, 2009 for a review of the standard counterfactual model’s limitations) by offering a broader and dynamic view of causality. In this account, causality is represented in terms of the functional relations between variables. Thus, it is evaluated in the framework of a specific causal event which specifies how the contributing causes relate with each other and, in turn, with the final outcome (for further details, see Halpern & Pearl, 2005). Like the standard account (Lewis, 1973, 1986), the structural account assumes that it is possible to attribute causality only when there is a counterfactual dependence between “X” and “Y”. However, the way in which this account conceives causality allows it to establish degrees of causality on the basis of how close the outcome was to counterfactually depending on a specific cause, thus solving the overdetermination problem of the standard account of causation (Lewis, 1973, 1986).

**A structural model of responsibility**

One way to study attribution of responsibility is to use a distributed system with multiple agents or components that interact to achieve a goal. Systems can be built in a very simple way, for example, so that the number of operations needed to achieve it is relevant, but not the intensity or quality of them. You can also define different operating rules in the system or the mode of interaction between agents. That is what Chockler and Halpern (2004) did. Taking the structural account as a framework, they provided an operational definition of responsibility that allows different degrees of responsibility to be established. Specifically, they proposed that the degree to which the agent X is responsible for the outcome Y depends on the distance between the actual world and the (hypothetical) causal world where X was pivotal for Y. That is, the hypothetical world where Y counterfactually depends on X. In particular, Chockler and Halpern (2004) define responsibility as:

i)

$$\text{Resp. }(Xi, Y) = 1/\,(N + 1)$$

where, *N* represents the minimal number of changes that would have to be made in the actual world to make Xi pivotal for Y. Returning to the voting example, Sarah's responsibility for B's defeat depends on the distance between the actual situation and the imagined world in which Sarah's vote is a determinant in changing the result. That is, the world where B's defeat counterfactually depends on Sarah's vote. Thus, in the situation where candidate A wins 11-0, five changes would be necessary in the actual world (five members would have had to vote for B) to make Sarah pivotal to B's defeat.

In this case then, Sarah would be assigned a Responsibility of 1/6 (= 1/ (5 + 1)). Conversely, in a situation where candidate A wins 7 - 4, Sarah's responsibility would be higher. Here, only one change is required in the actual situation to make Sarah pivotal to B's defeat: the vote of only one of the members who voted for A would have to change. In this case then, Sarah would be assigned a Responsibility of ½ (= 1/ (1 + 1)). Thus, the fewer changes required to make Sarah's vote pivotal for the defeat, the greater the degree of responsibility Sarah bears for B's defeat.

Several studies have tested Chockler and Halpern's (2004) definition of responsibility in a game context where the outcome depended on several players' performance (e.g., Gerstenberg & Lagnado, 2010; Zultan et al., 2012). Zultan et al. (2012), for example, developed the "dot-clicking game", a task where adult participants were asked to attribute blame to one player in a team of four. Players had to click on dots that appeared in different locations on a computer screen. There were two game rules (challenges) that differed in the number of players who had to succeed for the team to win. Both rules included conjunctive and disjunctive causal structures (causal functions; for details, see Steiner, 1972). In the conjunctive structure, all the players had to succeed for the team to win. Conversely, in disjunctive structures, just one of the players had to succeed for the team to win. So, for example, in one game rule, the team won if players P1, P2 and at least one out of P3 and P4 succeeded in their task (Structure: P1and P2, and [P3 or P4]). Thus, the performance of some players was related in a conjunctive way (e.g., P1-P2) and of others in a disjunctive one (e.g., P3-P4). The players' degree of responsibility was manipulated by varying the number of players who succeeded in their task, also including a baseline where all players failed. As Chockler and Halpern's (2004) definition predicted, responsibility attributions decreased with the number of changes required to make a specific player pivotal for the outcome. Thus, for example, participants held Player 1 more responsible when only she failed than when all the members failed (Zultan et al., 2012). However, the most interesting and unexpected result was that participants' attributions varied as a function of the causal structure in which a player was embedded. So on the baseline where all players failed, they blamed a player in a conjunctive structure (e.g., P1) more than a player in a disjunctive one (e.g., P3). The authors explained these results by arguing that they reflected the different ways in which a player could be made pivotal for an outcome. Note that although there was only one way to make P3 pivotal (changing P1 and P2 performance), there were multiple ways to make P1 pivotal (changing P2 and P3 or P2 and P4). So the more ways of being pivotal, the more blame attributed to the player.

**The criticality-pivotality model**

Based on the structural account of causation (e.g., Halpern & Pearl, 2005), its extension to responsibility judgements (Chockler & Halpern, 2004) and the results of previous empirical works (e.g., Gerstenberg & Lagnado, 2010, 2012; Zultan et al., 2012), Lagnado and colleagues (2014, Lagnado & Gerstenberg, 2015, 2017) proposed a formal model to account for individuals' causal responsibility judgements in contexts where several agents are contributing to the same outcome (overdetermination cases): the CP model. The core idea in the model is that responsibility judgements are the product of both prospective and retrospective factors. In particular, it assumes that when individuals attribute responsibility to an agent, they take into account two key considerations: 1) how important an agent's contribution is expected to be for reaching a joint positive outcome, and 2) how close the agent's action was to being pivotal for the outcome, that is, how close the outcome was to counterfactually depending on the agent's action. The former factor is evaluated before knowing the outcome and is called “criticality” in the model. The latter, however, is evaluated once the outcome is known and is termed pivotality. As the reader may have noticed, pivotality in the model constitutes the formal definition of responsibility offered by Chockler and Halpern (2004), and is therefore computed in the same way.

Lagnado and colleagues (2014) used the “dot-clicking game” task (Zultan et al., 2012) to investigate the role of criticality in causal responsibility judgements. In particular, they tested two different models to account for criticality: the “expected pivotality model” and the “heuristic model”. The expected pivotality model was based on Rapoport’s (1987) definition of criticality. Here, a player’s criticality would be equal to the probability that a player’s contribution will make a difference to the outcome. That is, the probability that a player’s action will be pivotal for the outcome. So, this definition posits that a player’s perceived criticality will be the same in a conjunctive as in a disjunctive structure, given that in both cases it is uncertain how the other players will perform. Note that this prediction cannot explain why participants in Zultan et al.’s (2012) study blamed more players in conjunctive structures than in disjunctive ones. Nevertheless, it correctly predicts that an agent’s criticality will be reduced in a disjunctive structure as the size of the group increases. The heuristic model, however, assumes that a player is fully critical when her contribution is necessary for the team to win. So a player’s criticality in a conjunctive structure will be maximum, equal to 1. Conversely, in a disjunctive structure, a player’s criticality will be divided among all the players related in a disjunctive way. So in a disjunctive structure composed of four players, each player’s criticality would be equal to ¼. On the other hand, if only two players are related in a disjunctive way, as happened in Zultan et al.’s (2012) study, each player’s criticality would be equal to ½. Criticality in this model could then be represented as “the relative decrease in the probability of the team winning when the player fails” (Lagnado et al., 2014, p. 1056).

In order to explore which model best predicts people’s prospective judgements, Lagnado and colleagues (2014) presented participants with different challenges that differed in team size and causal structure (conjunctive, disjunctive and mixed – some players were related with other members of their team in a disjunctive fashion and with the other team’s players in a conjunctive way). The participants’ task was to rate how critical a player A was for winning. They provided their answers on a separate sliding scale whose endpoints were “not at all” and “very much”. Results showed that participants rated player A’s contribution as more critical when A formed part of a conjunctive structure than when she was embedded in a disjunctive structure. In this latter case, participants’ criticality ratings were determined by the number of players related in a disjunctive way. Thus, the more players in the disjunction, the lower the player’s perceived criticality. Therefore, the heuristic model not only accurately predicted participants’ criticality judgements but also did so better than the expected pivotality one. These results led Lagnado and Gerstenberg (2015, p. 222) to define criticality of an agent $A_i$’s contribution in a situation S “as the probability that their contribution ($x_i$) is necessary for the positive group outcome (y)”:

ii)

$$Criticality(A, S) = \frac{p(y|x_i) - p(y|\neg x_i)}{p(y|x_i)}$$

where *A* represents the target player and *S* is the causal structure of the situation. Note that the minimal criticality (0) happens with 0 as numerator: when the probability of the result (y) given the intervention of an agent $p(y|x_i)$ is the same as when the agent does not intervene $p(y|\neg x_i)$. The maximum criticality (1) happens when the probability of the result without the intervention of the agent is null $p(y|\neg x_i)$, given that in this case, the numerator and denominator are the same.

In the second part of this experiment, Lagnado and colleagues (2014) presented participants with the results of four different group challenges, which varied in Player A's perceived criticality and her pivotality for the outcome. That is, they varied in how important the player's contribution was to reaching a positive outcome (criticality) and the degree to which the outcome counterfactually depends on the target player's performance (pivotality). Thus, there were challenges where the player was fully critical and pivotal to the outcome (C = 1, P = 1). Thus, for example, Player A's success was necessary to reach a positive outcome, and only she failed in their task. Also, there were challenges where she was fully critical but not pivotal to the outcome (C = 1, P = ½). For example, Player A's success was necessary for a positive outcome, but she and another fully critical player (e.g., Player B) failed in their task. In another challenge used by the authors, the player was pivotal to the outcome but not fully critical (C = ½, P = 1). For example, Player A's contribution was not necessary to the team's success (e.g., if player B had succeeded), but it would have been sufficient if Player B had failed. The participants' task was to estimate how responsible Player A was for the team's outcome in each challenge. Participants provided their answers on a sliding scale like the one used for assessing criticality. Results showed that participants attribute more responsibility for the team's outcome to A in the challenges where the player's action was pivotal to the outcome than in those where changes were required in the actual situation to make Player A pivotal. In particular, the more changes that were required to make Player A pivotal, the less responsibility was attributed to her. Furthermore, they found that in the challenges where Player A was pivotal to the outcome, participants attributed more responsibility to A when her success was necessary for the team to win (conjunctive structure) than when her success was sufficient but not necessary for winning (disjunctive structure). Lagnado et al. (2014, 2015) defined causal responsibility as:

iii)

$$Responsibility\ (A, O, S) = \alpha + W \text{ x Criticality } (A,S) + (1- W) \text{ x pivotality } (A, O, S)$$

Where responsibility (A, O, S) is agent A's causal responsibility for the outcome O in a causal structure S; α "is the global intercept and *w* is a weighting parameter whose range is constrained from 0 to 1" (Lagnado et al., 2014, p. 1061). Thus, according to the CP model, individuals' causal responsibility judgements are the result of an agent's perceived criticality to bringing about a positive outcome and the degree of the agent's pivotality for the actual outcome. Previous research with adults found that people's blame attributions differed depending on whether the target of judgments was a human agent or a robot (Leo & Huh, 2020). Then, in order to reduce the effect of additional potential intervening factors (such as players' intentions, beliefs, and motivation), we decided to use mechanical devices rather than agents to study causal responsibility attribution in children.

Lagnado and Gerstenberg (2015) mentioned an unpublished study in which participants were presented with the same causal structures and the same outcome patterns as described above. However, instead of agents, they asked participants about several components of a machine. The participants' task was to estimate the extent to which a component of a machine caused the machine to pass or fail a trial. Results showed that adults also considered both the target's criticality and its pivotality for the outcome when they assign causal responsibility to mechanical devices. It is not clear, however, whether children make these two considerations when they assign causal responsibility. This is the goal of the present research: to explore whether children take into account both a target's criticality and its pivotality for an outcome when they attribute responsibility to mechanical devices. To shed light on this question, we created a task with a similar structure to those mentioned above. In the "Battery task", participants were asked to rate how critical a wire's contribution was for the battery to work (prospective judgements) and how responsible the target wire was for the outcome (retrospective judgements). In Experiment 1, we applied it to adults in order to check whether our task would replicate Lagnado and colleagues' (2014) previous findings. In Experiment 2, we applied the same task to children aged between 8 and 13 years. We also asked children to complete a brief counterfactual reasoning questionnaire in order to check whether, as the CP model posits (e.g., Lagnado et al., 2014), there was a relationship between children's counterfactual reasoning abilities and their ability to distinguish the cases where the target is pivotal for the outcome from those where it was not (sensitivity to pivotality).

**The development of counterfactual thinking**

As we have already mentioned, the CP model claims that the assessment of pivotality involves counterfactual thinking (Lagnado et al., 2014; Lagnado & Gerstenberg, 2015). That is, it requires envisaging how things could have turned out differently (Byrne, 2016). As Byrne (2016) pointed out, counterfactual thinking serves several purposes, such as explaining the past or preparing for the future. Thus, when people read a counterfactual conditional, for example, "if the wire had been plugged in, the light would have been on", they think of two different possibilities: the conjecture ("the wire was plugged in, and the light was on") and the real fact ("the wire was not plugged in, and the light was off"). In contrast, hypothetical conditionals such as "if the wire was plugged in, the light was on" are easier, because they only require people to consider the conjectured possibility (see a review of the evidence of this dual representation using inference tasks and comprehension tasks with priming and eye movement procedures in Byrne, 2016 and Espino & Byrne, 2021).

Depending on the complexity of the tasks, some research found basic abilities of counterfactual thinking in preschoolers (Beck et al., 2006; Guajardo et al., 2009; Roldán-Tapia et al., 2017; Nyhout & Ganea, 2019) while other studies did not find this ability until 6–8 years (Burns et al., 2012; O'Connor et al., 2012; Weisberg & Beck, 2012; Beck & Guthrie, 2011; Kominsky et al., 2021; McCormack et al., 2018; Nyhout et al., 2019; Rafetseder & Perner, 2018). Other researchers had found that competence increases throughout the school years until around the age of 12 to 14, when it reaches a mature state of counterfactual thinking. (Gómez-Sánchez et al., 2020; Rafetseder et al., 2013, 2021). These developmental differences have been attributed to different factors. For example, Rafetseder and colleagues (Rafetseder et al., 2010, 2013) noticed that children's correct responses to counterfactual expressions ("if the wire had been plugged in, the light would have been on") did not necessarily mean that they were thinking counterfactually in the same way as adults. Rather they might have just interpreted it as a basic conditional expression: "if the wire was plugged in, the light was on". Children could have concluded that "if the wire was not plugged in, the light had to be off". To distinguish the basic rule (when plugged in, the light is on, and when not plugged in, the light is off) from understanding the counterfactual fact about "what really happened", it was necessary to introduce an "overdetermined" condition to test a mature comprehension of counterfactuals: for example, informing them that the switch was off. In these cases, a basic rule does not allow someone to predict what happened, given that "even if the wire was plugged in, the light was still off". Nonetheless, while the inclusion of an "overdetermined " condition is required to demonstrate the development of counterfactual abilities in schoolchildren, with some simpler tasks, even preschoolers can make correct inferences (McCormack et al. 2018; Nyhout and Ganea, 2019; see in discussion).

Summarising, although it is not clear what aspect of counterfactual ability changes during the last years of school age, we could expect that the causal attribution of responsibility could be influenced by this change. More specifically, if pivotality assessment implies counterfactual evaluation, then we could expect some developmental changes in causal responsibility attribution as well as in children's sensitivity to pivotality. That is, changes in children's ability to distinguish between the cases where the target's action was and was not pivotal for the outcome.

## Experiment 1

In Experiment 1 the aim was to analyse adults' attribution with the Battery task to test whether it was sensitive and accurate enough to detect the effects found in previous studies (Lagnado et al., 2014; Zultan et al., 2012). Thus, it was applied following the same procedure as we intended to follow in Experiment 2 with children. In this task, participants had to think about how a plug-in battery works. Participants were asked about the responsibility of a particular wire in the functioning of a battery with four coloured wires. The wires were connected in such a way that it was possible to manipulate its criticality and pivotality for the outcome. The battery worked if the brown, blue and at least one of the two special wires, either yellow or orange, worked correctly (see Figure 1). Therefore, the brown and the blue wires are in a conjunctive structure: their functioning is necessary for the battery to work. Conversely, the yellow and orange wires are in a disjunctive structure, given that with only one of them functioning the battery could work. As can be seen, according to ii), the brown wire's criticality is maximum, given that if it does not work, the probability for the battery to run is null $p(y|\neg x_i)$. In contrast, the criticality of the yellow wire is ½ because, even if it does not work, the orange wire can work $p(y|\neg x_i)$= ½, and therefore, assuming whatever probability for $p(y|x_i)$ in the numerator and denominator, criticality is ½.

Similarly, when the outcome is known, participants can compute the pivotality of a wire. Thus, if the battery does not work and all the wires except the brown one work, it can be concluded that the brown wire was pivotal to the outcome (See Figure 2A). Thinking counterfactually, we realise that no change is necessary to make the brown wire pivotal. So its responsibility according to i) is maximum 1 (= 1/(0+1)). However, when not only the brown wire but also the blue one fails to work correctly (See Figure 2C), we have to mentally simulate and change the blue wire's functioning in order to make the brown wire pivotal for the outcome. In this case, N is equal to 1 and, therefore, the brown wire's pivotality is ½ (= 1/ (1 + 1)).

Participants were asked to make two sorts of judgement:1) prospective judgements, where they were asked to estimate how critical a wire's contribution was for the battery to work; and 2) after the outcome was known, retrospective judgements, where they were asked to rate how responsible a target wire was for the outcome. In prospective judgement trials, criticality was manipulated by asking participants about either a wire embedded in a conjunctive structure whose contribution was, therefore, necessary for the battery to work (the brown wire, C = 1), or asking them about a wire embedded in a disjunctive structure whose contribution was sufficient but not necessary (the yellow wire, C = ½) (see Figure 1).

In the retrospective judgement trials, we manipulated the target wire's criticality and its pivotality for the outcome. Criticality was manipulated in the same way as in the prospective trials. Pivotality, however, was manipulated by varying the number of wires that worked correctly in each trial. In all trials, we told participants that the battery had not worked, the reason being that negative outcomes trigger causal reasoning as well as counterfactual thinking in adults (Enzle & Schopflocher, 1978; Hilton, 1990) and children (e.g., German, 1999).

As did Lagnado and colleagues (2014), we expected that the heuristic model would predict participants' prospective judgments better than the pivotality model. So we expected participants to rate the wire whose contribution was necessary for a positive outcome (C = 1) as more critical than the wire whose contribution was sufficient but not necessary, C = ½ (Lagnado et al., 2014; Zultan et al., 2012). With regard to retrospective judgements, we expected that participants would hold the target wire more responsible in the constant condition trials (C= 1, P = 1) than in pivotality reduced (C = 1, P = ½) ones (Gerstenberg & Lagnado, 2010; Lagnado et al., 2014; Zultan et al., 2012). That is, we expected them to rate the fully critical wire (the brown wire, C = 1) as more responsible when its functioning was pivotal for the outcome (P = 1) than when it was not (P = ½). Similarly, we expected that in the cases where the target was pivotal for the outcome (P = 1), participants' responsibility attributions would vary in function of the target's criticality for the outcome. In particular, we expected participants to rate the fully critical wire (the brown wire, C = 1) as more responsible than the non-fully critical one (the yellow wire, C = ½) in the trials where both were pivotal for the outcome (Lagnado et al., 2014).

## Method

In these studies, we report all measures, manipulations and exclusions. Both of our studies (Experiments 1 and 2) were approved by the university's ethical review board (Comité de Ética en Investigación Humana de la Universidad de Granada: 1068/CEIH/2020). Sample size was determined before any data analysis.

**Participants**

A total of thirty-eight students (6 males and 32 females) from the University of Granada (age: *M* = 21.20; *SD* = 3.71) took part in Experiment 1. They participated in exchange for course credits. Twenty participants completed prospective judgments trials first, while the remaining eighteen participants started by completing retrospective judgements trials. For this study, participants read and filled out consent forms complying with the University Research Ethics Committee guidelines.

**Materials**

In the Battery task, a battery had four wires identified by different colours: brown, blue, yellow, and orange (see Figure 1). The battery worked only if the brown and blue wires worked along with one of the special wires (either yellow or orange). The yellow and orange wires were labelled as "special wires" because, unlike the first two wires, the battery can work even if one of them fails, as long as the other works. So, the causal structure of the battery's operation is mixed: It includes conjunctive (brown + blue + [yellow or orange]) and disjunctive (yellow or orange) conditions. Participants were asked to make prospective judgements and retrospective judgements about the battery's operation. Eight slideshow presentations were prepared. They can be found at DOI: 10.6084/m9.figshare.23503548. Each presentation included six prospective judgements trials and six retrospective judgements trials. The order of presentation of judgements was counterbalanced across presentations.

[FIGURE 1 ABOUT HERE]

In prospective judgements trials, we manipulated the wires' criticality value by asking participants about either a wire whose contribution was necessary for the battery to work (brown wire in conjunctive condition trials, C = 1) or a wire whose contribution was not necessary for running the battery (yellow wire in disjunctive condition trials, C = ½ ). More specifically, in four out of the six prospective judgements trials, we asked participants to rate the brown wire's criticality (C = 1) and in the remaining two trials to rate the yellow wire's criticality (C = ½). The participants' task was to estimate how decisive the target wire (either the brown or the yellow) was to running the battery. Although we kept the term "critical" throughout the instructions, we used the Spanish word "decisivo" ("decisive" in English) because this word is more familiar to children than critical and the two words are synonyms in Spanish. We then decided to retain the same term in the study with adults.

In the retrospective judgements trials, we manipulated both the target wire's criticality and pivotality values. Pivotality was manipulated by varying the number of wires that worked correctly, while criticality was manipulated in the same way as in the prospective judgements trials. In all trials, the outcome was negative: the battery did not work (see Figure 2). Participants completed trials for three different conditions (two trials per experimental condition). In the constant condition trials, the target wire was fully critical and pivotal to the outcome (C = 1, P = 1): all wires worked correctly with the exception of the target wire (brown wire) – see Figure 2A. In the reduced criticality condition trials, the target wire (yellow wire) was pivotal to the outcome but not fully critical (C = ½, P = 1): all wires worked correctly except for the special wires (yellow and orange). See Figure 2B. Finally, in the reduced pivotality condition trials, the target wire (brown wire) was fully critical but not pivotal to the battery's malfunctioning (C = 1, P = ½): only the special wires worked correctly. Neither the brown nor the blue wire worked (see Figure 2C). The order of presentation of the three different retrospective judgements was randomised in each slideshow presentation.

[FIGURE 2 ABOUT HERE]

In all the presentations, we included a "Balls task" before the experimental trials. The purpose of this task was to familiarise participants (especially children, in the next experiment) with the meaning of the scale. The Balls task included three trials. In each trial, participants viewed a transparent box that included balls of two different colours, with different proportions of each. For each trial, they were asked to estimate how possible it was that a blindfolded child would pull out of the box a ball with a specific colour. After this task, the experimental trials (either prospective or retrospective) appeared.

For both tasks, participants provided their answers in a booklet containing 15 non-numbered scales (ranging from 0 to 10) whose endpoints were labelled "not at all" and "very much". Each scale was preceded by the trial number (e.g., Balls 1 for the Balls task trials, Question 1A for prospective judgements trials and Question 1B for retrospective ones). This number as well as the trial question (either Balls task, prospective judgement or retrospective judgement) also appeared on the screen. For both tasks, participants provided their answers by circling one of the lines of the non-numbered scale.

**Procedure**

Participants were tested in a quiet room, in groups of three to five people. They were seated at a table in front of a laptop where the slideshow presentations were displayed. We started each experimental session by explaining the meaning of the scale. After that, and to ensure that participants understood the use of the scale (specially children in the next experiment), they completed the Balls task. In each trial, the experimenter displayed a slide with a box containing several balls of different colours. After displaying this, she also informed them verbally about how many balls of each colour were inside. Participants' task was to rate in their booklet how possible it was that a blindfolded child would pull out of the box a ball with a specific colour. Following this, the experimental task started.

At the beginning of the Battery task, all participants received general instructions that included the following scenario: "*Chinese researchers have developed a special plug-in battery. It is special because it operates for 72 hours per charge and it takes just 20 minutes to load. However, researchers continue testing because they do not get the battery to run. This battery has four wires: a brown wire, a blue wire and two special wires, one yellow and the other orange. The battery runs if the brown and blue wires work and at least one of the special wires, either the yellow or the orange. For that reason, the yellow and the orange wires are special: only one of them needs to work to run the battery as long as the brown and blue wires also work".* During this explanation, a slide representing the causal structure of the battery's functioning was displayed (see Figure 1). This slide did not include either the trial question or the trial number. We tested children's understanding of the battery's operation rules by asking them orally what would happen if, for example, only the blue and brown wires worked properly. Once the experimenter was sure that participants had understood the instructions (how many wires had to work to get the battery to run), they completed the experimental trials, writing their answers in the booklets.

Before completing the prospective judgements trials, we explained to participants what "critical" ("decisive" in Spanish) meant. This explanation was included to ensure that children (in the next experiment) understood the meaning of "critical". We also provided this explanation to adults. We used a well-known Spanish cookery show to illustrate what "critical" means. In this show, three judges decide who is the best cook among several candidates. We told participants that "critical" refers to "*the power that* *someone has to make that thing happen or not*." After that, we asked them to envisage that in this TV show, there was only one judge and we asked them how critical this judge's vote was to decide the winner. Participants provided verbal answers and we told them: "*in this case, the vote of Chicote (the judge) was fully critical because he had the power to decide who would be the winner*". Next, we asked participants whether Chicote's vote would be more or less critical in the actual situation (in which there were three judges) than in the hypothetical situation mentioned previously. If any of the participants claimed that Chicote's vote would be more critical in the actual situation than the hypothetical one, they received a further explanation. Subsequently, they completed the prospective judgments trials.

In the prospective judgements trials, the experimenter showed participants the same image of the four wires, but headed now by the trial question (referring to either the brown or the yellow wire) and its number (see Figure 1). Once the image was displayed, the experimenter asked participants to estimate and mark using the scales in their booklet how critical the target wire was to getting the battery to run.

In the retrospective judgements trials, participants were presented with the same image as in the prospective ones, but in addition, they included information about each wire's functioning: the wires that worked correctly and those that did not. Specifically, when a wire worked correctly, a green tick appeared above it. When a wire did not work, a red cross appeared above it (see Figure 2). In these trials, we told participants: "*Now I will show you the results of some operating tests carried out by the Chinese researchers*. *Specifically, I will tell you which wires worked in the operation of each test. Your task is to estimate how responsible one wire is for either the battery's functioning or its malfunctioning*." We did not tell participants beforehand that in the trials, the battery would not work. The reason for this was to keep participants in anticipation so that they would focus on the task. In each trial, a different slide was displayed on the screen. The experimenter informed participants, one by one, about which wires worked in each trial (this information also appeared on the slide, see as example, Figure 2). Once participants knew each wire's functioning (whether it worked or not), they were asked to rate how responsible the target wire was for the battery's malfunctioning. As in the prospective trials, each slide included the trial's question and its number (e.g., Question 1B). For both trials, participants provided their answers by circling one of the lines on the non-numbered scale.

## Results

Data for the experiments are available at DOI: 10.6084/m9.figshare.23503548

### Prospective judgements

A paired t-test was conducted to compare participants' ratings in conjunctive ($C = 1$) and disjunctive condition ($C = ½$) trials. Results are shown in Table 1. There were statistically significant differences in participants' ratings among trials, $t(38) = 12.73$, $p < .001$. In particular, participants rated as more critical the wire whose contribution was necessary for the battery to work ($C = 1$, the brown wire) than the wire whose contribution was sufficient but not necessary ($C = ½$, the yellow wire).

[TABLE 1 ABOUT HERE]

### Retrospective judgements

Again, a paired t-test was conducted to compare participants' ratings in the constant condition ($C = 1$, $P = 1$) and the pivotality reduced condition ($C = 1$, $P = ½$). The mean ratings are shown in Table 1. The analysis revealed statistically significant differences among trials, $t(38) = 8.90$, $p < .001$: participants held the fully critical wire (the brown wire) as more responsible in the trials where its functioning was pivotal to the outcome (P = 1) than in those where a change was required in the actual situation to make it pivotal (P = ½).

In addition, in order to test the effect of criticality in participants' retrospective judgements, we conducted an additional analysis comparing participants' ratings to constant condition (C =1, P = 1) and criticality reduced condition trials (C = ½ , P = 1). Data are shown in Table 1. Again, a paired t-test was conducted. There were statistically significant differences, $t(38) = 11.70$, $p < .001$. Although in both conditions the target wire was pivotal to the outcome, participants attributed more responsibility for the outcome to the fully critical wire (the brown wire) than to the non-fully critical wire (the yellow wire).

**Discussion of Experiment 1**

Lagnado and Gerstenberg (2015) reported an unpublished study in which they found that adults attribute responsibility to components of a device in the same way as they do with players in a game. In Experiment 1, we replicated their main findings with a new task based on the functioning of a battery designed to test children's attributions of responsibility. Like these authors, we found that participants rated as more critical the wire whose functioning was necessary for the battery to work (the brown wire) than the wire whose functioning was sufficient but not necessary (the yellow wire). These results could not be explained by the "expected pivotality model" (Rapoport, 1987), given that it predicts that a wire embedded in a conjunctive structure would be perceived as equally critical to a wire embedded in a disjunctive one. Also, in our study, the heuristic model seems to predict better than the expected pivotality model participants' prospective judgements. Furthermore, we found that participants' retrospective judgements varied in function of the target's criticality and pivotality for the outcome. In particular, the fully critical wire (the brown one) was rated as more responsible when its functioning was pivotal to the outcome than when it was not. Conversely, in the cases where the target wire was pivotal to the outcome, participants rated as more responsible the fully critical wire (the brown one) than the non-fully critical wire (the yellow one). The present results show that the new task is sensitive enough to detect changes in the responsibility judgments of adults. Also, the results support the CP model, replicating previous results obtained with different tasks, and show that the model accounts for the attribution of responsibility to components of mechanical devices.

## Experiment 2

In Experiment 2, we applied the Battery task to schoolchildren aged between 8 and 13 in order to test the CP model. We aimed to discover 1) whether children are able to discriminate the causal role, necessary or sufficient, of one component in a mechanism's operation, 2) whether they take into account this role as well as the component's pivotality for the outcome in their causal responsibility judgements, and also 3) to test an additional prediction of the CP model: the relation between assessment of pivotality and counterfactual thinking (Lagnado et al., 2014). To this end, we asked children to complete a brief counterfactual reasoning questionnaire and related their performance in it to their sensitivity to pivotality and criticality in both prospective and retrospective judgements. We expected that only retrospective judgments and not prospective ones would relate to counterfactual thinking. Counterfactual thinking requires people to mentally undo a known outcome, and outcomes are not given in prospective judgments. A particularly interesting kind of counterfactual is the semifactual, in which people also search for contrary antecedents, but they keep the same outcome. We computed three indices to correlate with counterfactual thinking measures. For prospective judgments, we computed the *Sensitivity to prospective criticality* (see Figure 1). This was defined as the difference in children's ratings between the conditions where the wire's contribution was necessary for the battery to work (conjunctive condition, C = 1) and those where it was not (disjunctive condition, C = ½). The smaller the difference between the ratings among conditions, the lower the child's sensitivity. We did not expect any association with the counterfactual measures. For retrospective judgements, we computed the *Sensitivity to Pivotality* and *Sensitivity to* Retrospective *Criticality indices* (see Figure 2). *Sensitivity to pivotality* was defined as the difference between the constant condition trials (C = 1, P = 1) and the pivotality reduced condition trials (C = 1, P = ½) in children's ratings. Thus, for example, a child's sensitivity to pivotality was "5" when she gave a rating of 10 (labelled as "very responsible") in the trials where the target wire was pivotal to the outcome (P = 1) and a rating of 5 (labelled as "partly responsible") in the trials where the wire was not pivotal (P = ½). Finally, the *Sensitivity to Retrospective Criticality* index reflects children's sensitivity to criticality after knowing the outcome, in retrospective judgements trials. It was defined as the difference between children's ratings in the constant (C = 1, P = 1) and the criticality reduced (C = ½, P = 1) conditions. The interest of this last index was to check whether the two indices behave in a similar way to the causal structure in prospective and retrospective judgments. No differences were initially predicted.

As we mentioned in the introduction, there is mixed evidence about when children think counterfactually in the same way as adults. However, some previous studies have found that when children are tested with tasks that include overdetermined conditions, their performance with the counterfactual questions, more specifically, with the semifactual ones (in the overdetermined conditions), improves during mid-childhood (Gómez-Sánchez et al., 2020; Moreno-Ríos & García-Madruga, 2002; Rafetseder et al., 2013, 2021). Our questionnaire included questions similar to those used in the studies that show developmental differences in schoolchildren (e.g., Gómez-Sánchez et al., 2020, 2021; Rafetseder & Perner, 2013). The items in the questionnaire consisted of stories with conditionals that require children to think not only how a different outcome could have occurred (counterfactual thinking) but also how the same outcome could have resulted from different situations (semifactual thinking). Semifactual conditionals are a kind of counterfactual conditional with a different initial representation in which the conclusion is fixed (Byrne, 2016; Moreno-Ríos, et al., 2008; Ruiz-Ballesteros & Moreno-Ríos, 2016; Santamaría et al., 2005). In the case of the Battery task, counterfactual thinking leads children to realise that if the wire had functioned, the battery would have worked (for example in pivotality = 1 conditions). Conversely, semifactual thinking would lead them to detect that "even if the wire had functioned, the battery would not have worked" (for example, in pivotality =1/2 condition). It would allow them to distinguish the cases where the target was pivotal to the battery's malfunctioning from those where it was not (sensitivity to pivotality). Based on this, we hypothesised that if retrospective judgements and, more specifically, the assessment of pivotality, rely on counterfactual thinking, then children with better counterfactual (and semifactual) reasoning abilities should be more sensitive in their judgments to the target's pivotality for the outcome. Thus, we predicted that older children would be more sensitive to pivotality than younger ones, given they have more competence in counterfactual reasoning (e.g., Gómez-Sánchez et al., 2020, 2021; Rafetseder et al., 2013). It does not mean, however, that we expected younger children to be insensitive to pivotality. As we have seen, even pre-schoolers make correct inferences from counterfactual conditionals with some particular tasks (Beck et al., 2006; Guajardo et al., 2009; Nyhout & Ganea, 2019; Roldán-Tapia et al., 2017). However, some results suggest that this kind of thinking improves during childhood, becoming as they develop, more similar to adults' counterfactual thinking (e.g., Gómez-Sánchez et al., 2020, 2021; Rafetseder et al., 2013). Thus, we expected that younger children would be sensitive to pivotality but to a lesser extent than older ones, given that only older ones use mature counterfactual reasoning strategies.

Regarding prospective judgments, a recent study has found that children's attributions are sensitive to causal structure. In this study, Koskuba et al. (2018) asked children of 4 to 7 years and adults to allocate rewards/penalties to the members of a team. They manipulated the players' performance and the causal structure of the game, which determined how the players' performance combined to win or lose (conjunctive or disjunctive). Interestingly, there were no developmental differences between 4 and 7 year-old children, but there were between children's and adults' allocations (Koskuba et al., 2018; Experiment 2). So despite the lack of specific research with middle-grade schoolchildren, we expected children's prospective judgments to be sensitive to a wire's (necessary or sufficient) causal role in the battery's functioning, with no developmental differences. Nevertheless, it is important to note that there are many differences between Koskuba et al.'s (2018) study and the present one. For example, in their study, children had to think about people's behaviour rather than a component of a mechanism. In addition, the participants' task was different: they were asked to allocate rewards/penalties to players while here, children were asked to attribute responsibility to a component of a mechanical device. As Koskuba et al. (2018, p. 241) pointed out, although responsibility and reward attributions are closely linked, each kind of attribution may involve specific factors. Also, the effect of the causal structure was tested between subjects while here it was tested within subjects. Thus, in the light of the large differences among studies and the lack of specific research, we kept an open mind.

## Method

### Participants

One hundred and sixty-nine schoolchildren (83 girls and 86 boys) aged between 8 and 13 years (*M* = 9.85; *SD* = 1.25, range = 8.08 to 13.17) completed the Battery task: seventy-nine children from the second level of Spanish primary school (2nd PS-Level; *M* = 8.81; *SD* = 0.70; range = 8.08 to 10.67) and ninety children from the third level of Spanish primary School (3rd PS-level; *M* = 10.76; *SD* = 0.87, range = 10.08 to 13.17). One hundred and sixty-one of them (*M* = 10.42; *SD* = 1.23; range = 8.08 to 13.17) also completed the counterfactual reasoning questionnaire: seventy-three from second PS-level (*M* = 8.85; *SD* = 0.70; range = 8.08 to 10.67) and eighty-eight from third PS-level (*M* = 10.76; *SD* = 0.87; range = 10.08 to 13.17). All of them were native speakers of Spanish and were enrolled in two different state schools in Granada, Spain. They participated only if their parents had given written consent for this study, complying with the ethical protocol of the University Ethics Committee.

### Materials

The materials used for the Battery task were the same as those described in Experiment 1. In addition, we asked participants to complete a brief counterfactual reasoning questionnaire. The questionnaire included two short stories and four questions labelled as: an inferential question, an epistemic status question, a comprehension question and a semifactual question. There were two different versions of the counterfactual reasoning questionnaire. The two versions differed only in the contextual information (the short stories). In both, the questions appeared in the same order as below.

In one of them, the first story said: "*María is on the phone with her husband Juan. María is in a shop because she wants to buy a new TV. She goes up to a shelf and sees a*

*TV for sale. She notes that the TV is unplugged. She says to Juan: "If somebody had pressed the remote control…" Just at that moment, the phone goes dead*." After that, the inferential and the epistemic status questions appeared.

1) In the inferential question, participants were asked: How do you think the conversation would have continued if the phone line had not gone dead? If somebody had pressed the remote control, then the TV…

   Then they had to choose between three answers: a) would have come on, b) would not have come on, and c) It's impossible to know.

2) In the Epistemic status question, they were asked: When Maria said "If somebody had pressed the remote control…", what do you think was really going on?

   The three options for response were: a) Somebody pressed the remote control, b) Nobody pressed the remote control, and c) There was no remote control.

The purpose of these first two questions was to evaluate whether children could identify what would have happened in a different situation (inferential question) and whether they were able to distinguish what really happened from what was hypothetically proposed (epistemic status question).

After that, a control comprehension question tested whether children understood the first story by asking them: Where was María when she was talking on the phone? And the three answer options were: a) In her home, b) In a shop, and c) On the beach.

The second short story was designed to assess children's semifactual reasoning abilities. That is, to test whether children realise that an outcome would have been the same even with a different antecedent(s). This second story said: *Imagine that you are in a park and you hear somebody say: "I know what really happened and, even if Ana* *had trained hard, she would not have won the competition."* After that, the semifactual question appeared:

3) What really happened? a) Ana won the competition, b) Ana did not win the competition, c) It's impossible to know.

**Procedure**

The procedure for the Battery task was the same as that described in Experiment 1. The counterfactual reasoning questionnaire was applied in groups of twenty to twenty-five children in their classroom. The experimenter asked the children to read the stories carefully and answer the questions that appeared below. The children were encouraged to ask any questions they might have while completing the questionnaire. When a child asked a question related to the content (the text in either the story or the question), he/she was asked to read it again and to ask any question relating to the wording. These doubts were clarified, taking care not to give any clues about the answer.

## Results

**Prospective causal judgements**

Children's responses to prospective causal judgements trials were submitted to a 2 (criticality's value: ½, 1) x 2 (PS-level: second, third) mixed ANOVA, with the first variable manipulated within participants and the second between groups[1]. Results are shown in Table 1. The analysis revealed a main effect for criticality value, $F(1, 167) = 183.07$, $p < .001$, $\eta 2 = .52$, and for PS-level, $F(1, 167) = 4.70$, $p = .032$, $\eta 2 = .03$). Children rated as more critical the wire whose contribution was necessary for the battery to work (the brown wire, C = 1) than the wire whose contribution was sufficient but not necessary (the yellow wire, C = ½). Furthermore, younger children's mean

[1] The "order" factor was used as a control in our experiment. Although we had no initial predictions, we analysed the effect of this factor in participants' prospective and retrospective judgments. A more detailed description of the "order" factor effect can be found at DOI: 10.6084/m9.figshare.23503548

ratings were higher than those of older ones. The interaction between criticality value and PS-level was not significant, $F(1, 167) = 0.59$, $p = .442$, $\eta 2 = .01$.

**Retrospective causal judgements**

In order to explore whether children were sensitive to pivotality, we performed an analysis including the children's responses to the constant (C = 1, P = 1) and the reduced pivotality (C = 1, P = ½) condition trials. Their responses were submitted to a 2 (pivotality value: ½, 1) x 2 (PS-level: second, third) mixed ANOVA, with the first variable manipulated within participants and the second between groups. The results are displayed in Table 1. Results showed a main effect for pivotality value, $F(1, 167) = 50.42$, $p < .001$, $\eta 2 = .23$: children held the fully critical wire (the brown wire) as more responsible in the trials where it was pivotal to the negative outcome (P = 1) than in those where a change was required in the actual situation to make it pivotal (P = ½). Also, although there were no global differences in children's mean ratings per PS-level, $F(1, 67) = 2.32$, $p = .573$, $\eta 2 = .01$, the interaction between pivotality value and PS-level was significant, $F(1, 167) = 7.66$, $p = .006$, $\eta 2 = .04$. Further analyses showed that although children of both second and third PS-levels were sensitive to pivotality, the effect was greater at the third PS-level ($F(1, 89) = 56.94$, $p < .001$, $\eta 2 = .39$) than at the second one ($F(1, 78) = 8.03$, $p = .006$, $\eta 2 = .09$).

In addition, in order to explore whether children's retrospective judgments were sensitive to the target's criticality value, we performed an analysis including their ratings in the constant condition (C = 1, P = 1) and the reduced criticality condition (C = ½, P = 1) trials. These ratings were submitted to a 2 (criticality value:  ½ or 1) x 2 (PS-level: second, third) mixed ANOVA. Results are displayed in Table 1. The analysis revealed a strong main effect for the criticality value factor, $F(1, 167) = 128.09$, $p < .001$, $\eta 2 = .43$. Although in both conditions the target wire was pivotal to the outcome

(P = 1), children held the fully critical wire (the brown wire, C = 1) to be more responsible than the non-fully critical wire (the yellow wire, C = ½). There were no differences in children's mean ratings between the two PS-levels, $F(1,167) = 0.58$, $p = .446$, $\eta 2 = .01$. Nevertheless, the interaction between criticality value and PS-level was significant, $F(1, 167) = 5.23$, $p = .023$, $\eta 2 = .03$. Further analyses showed that although at both levels children's responsibility judgements were influenced by the target's criticality value, the effect was greater at the third PS-level than at the second one, $F(1, 89) = 109.31$, $p < .001$, $\eta 2 = .55$ and $F(1,78) = 34.56$, $p < .001$, $\eta 2 = .30$, respectively.

**Counterfactual reasoning questionnaire and pivotality judgements**.

Children's responses to the counterfactual reasoning questionnaire were coded with scores of 1 for each correct response and 0 for incorrect ones. We excluded from the sample fifty-five participants for either not answering all the questions or for answering the control comprehension question incorrectly. Thus, the sample for this analysis consisted of one hundred and six children (51 from the second PS-level and 55 from the third PS-level). Firstly, we tested to check that the comprehension question was not answered correctly by chance in either of the two groups. The binomial test showed that both older and younger children answered the comprehension question above chance level (binomial test given a probability = .5 for second PS-level: 70%, $p = .001$; for third PS-level: 63%, $p = .025$). Subsequently, we analysed the counterfactual questions. In a similar previous study, the youngest children responded correctly by chance or even below to the counterfactual questions in some conditions (meaning that they confused the conjecture possibilities with the real fact; see in Gómez-Sánchez et al. 2020 Table 3). In that study, only older children distinguished clearly above chance level the real and the hypothetical possibilities in all the conditions. Even so, children's ability seems to be lower than that of adults. Our results showed the same developmental trend. In the three experimental questions (inferential, epistemic status and semifactual), only older children gave responses above chance (binomial test given a probability = .5 for inferential question: 76%, $p < .001$; for epistemic status question: 75%, $p <.001$; and for semifactual question: 71%, $p = .003$). Conversely, younger children performed at chance level (binomial test given a probability = .5 for inferential question: 59%, $p = .262$; for epistemic status question: 53%, $p = .780$; and for semifactual question: 57%, $p = .401$). Next, we computed a global index of counterfactual reasoning (CF index) based on the average of the correct responses to each question, with higher scores indicating a better counterfactual reasoning performance. An independent sample T-test was conducted using the CF index and grouped by the children's PS-level. Results showed that older children gave more correct responses than younger ones, $t(104) = 2.87$, $p =.005$, $d = 0.54$ ($M = 0.56$, $SD = 0.33$ for second PS-level and $M = 0.74$, $SD = 0.31$ for third PS-level).

Finally, in order to test whether there was a relation between children's counterfactual reasoning abilities and their sensitivity to criticality and pivotality, we computed three indices: an index of sensitivity to prospective criticality (SPC index), an index of sensitivity to criticality in the retrospective judgements trials (SRC index), and an index of sensitivity to pivotality (SP index) (See Figures 1 and 2). These indices reflected the differences mentioned above between children's ratings. After that, we analysed whether there was a relation between the three indices (SPC, SRC and SP) and the three counterfactual reasoning questions (inferential, epistemic status and semifactual) controlling for age. A positive correlation was found between the SP index and the semifactual question ($r = 0.18$; $p = .022$) and between the SRC index and the inferential question ($r = 0.15$; $p = .048$). The correlation between *SRC* and the CF index did not reach statistical significance ($r = 0.15$; $p = .057$). No other correlations were significant (see Table 2).

[TABLE 2 ABOUT HERE]

## General Discussion

In two experiments, we investigated adults' and children's causal responsibility attributions to mechanical devices in the framework of the CP model (Lagnado et al., 2014; Lagnado & Gerstenberg, 2015). Although the model was tested in a context of responsibility attribution to human agents, Lagnado and Gerstenberg (2015, pp 229-230) reported on an unpublished study in which they found similar results using components of a device instead of agents as the target of judgments. In Experiment 1, using a novel task designed to test children, we replicated those reported findings with adults. In Experiment 2, we applied this task to children aged between 8 and 13 years. The results obtained in this experiment constitute the first evidence, as far as we know, of the CP model's (Lagnado et al., 2014; Lagnado & Gerstenberg, 2015) suitability to explain children's causal responsibility attributions. Thus, we found that children were able to distinguish the causal role of a component (necessary or sufficient) in a mechanism's operation. This happened not only when participants were asked to judge in advance the component's importance for a positive outcome (prospective judgements) but also when they were asked to attribute responsibility to a target component after knowing the outcome (retrospective judgements). Besides the causal role, children's responsibility attributions were influenced by the component's pivotality for the outcome. Thus, they held the target as more responsible when the outcome depended counterfactually on the target functioning than when it did not.

**Pivotality and Counterfactual thinking**

Another of our goals in this work was to explore whether the assessment of pivotality relies on counterfactual thinking, as the CP model posits (Lagnado et al., 2014; Lagnado & Gerstenberg, 2015). To test whether these changes in counterfactual thinking during school years are reflected in pivotality assessments, in Experiment 2, we asked children to complete a brief counterfactual reasoning questionnaire based on these propositional tasks (e.g., Gómez-Sánchez et al, 2020; Rafetseder et al., 2013). In accordance with previous studies (Gómez-Sánchez et al., 2020, 2021; Moreno-Ríos & García-Madruga, 2002; Rafetseder et al., 2010, 2013, 2021), we found that children's performance in the counterfactual task increased during mid-childhood. Older children obtained higher ratings than younger ones in the counterfactual reasoning questionnaire. Interestingly, the same developmental changes were found in children's sensitivity to pivotality (See Table 1): older children differentiated better than younger ones in the cases where the wire was pivotal to the outcome from those where it was not. In line with this result, after controlling the effect of age, we found a positive relationship between children's performance in the semifactual question and their sensitivity to pivotality (See Table 2). These results constitute the first clear evidence of the relation between counterfactual thinking and pivotality assessment. More specifically, they reveal a relationship between pivotality assessment and semifactual thinking. It is important to note that although all the three questions concern counterfactual thinking, they do so in different ways. The semifactual question, for example, requires an understanding that the outcome is fixed even if the antecedent is changed mentally. This operation of mentally undoing actions while the real outcome is fixed is employed in the computation of pivotality, as proposed by the CP model (Lagnado et al., 2014). The same operation was considered nuclear for determining the "mature counterfactual reasoning strategy", reached in later childhood. According to Leahy and collaborators (Leahy et al., 2014, see also Rafetseder et al., 2010), this strategy is shown when the problem requires preserving as much of the story and the outcome as possible while accommodating the antecedent. Then, it is not surprising that the semifactual question correlates with sensitivity to pivotality given that both use the same mental operations. These results strongly support the idea that pivotality assessment requires counterfactual thinking (e.g., Lagnado et al., 2014; Lagnado & Gerstenberg, 2015).

**Sensitivity to Criticality prospective and retrospective**

Although the CP model (e.g., Lagnado et al., 2014) only establishes a relation between counterfactual thinking and pivotality assessment, we also explored the relation between this kind of thinking and children's sensitivity to criticality in both judgements. We expected that if there is any role of counterfactual thinking in the sensitivity to criticality, that must be when the outcome is known. This only happens in retrospective judgments (SRC) but not in the prospective ones (SPC). The results corroborate this. On the one hand, results showed that the SRC index correlated with the general inferential question. Answering correctly to this question requires using contextual information to cancel the automatic modus ponens inference. For example, after knowing the following, "If she had pressed the remote control, the TV would have turned on", and "She pressed the remote control", we can conclude "the TV turned on". However, given the contextual information that the TV set was unplugged, this automatic inference must be cancelled, and conclude "the TV did not turn on". In our task, being sensitive to SRC requires realizing that, although in both conditions the target wire is pivotal to the outcome (P =1), only in the criticality reduced condition (C

= ½ - P = 1) is there another wire that could be mentally changed to get a different outcome. Here, the battery could have worked even if the yellow wire had not worked. Then, the counterfactual question and the responsibility judgment questions require considering the contextual information (either the battery functioning rules or the information about the (un)plugged state of the TV in the story).

On the other hand, as was expected, children's SPC was not related to either their PS-level or any of our counterfactual reasoning measures (inferential, epistemic, semifactual or general CF index). Therefore, the results are consistent with the idea that the assessment of the retrospective criticality does not require the same mental operations as the assessment of the prospective one (see Lagnado et al., 2014; Lagnado & Gerstenberg, 2015). In that sense, when people envisage the initial configuration of a causal structure, they can think forward with prefactuals. It allows participants to anticipate that if the brown wire does not work, the battery will not work (conjunctive causal structure) and also to anticipate that if the yellow wire does not work, the battery might still work: it is possible that the orange wire works, and then the battery will work (disjunctive causal structure). In both cases, prefactual (if then) thinking of conjectured possibilities can be the basis for computation of probabilities. However, retrospective criticality assessment requires backward thinking by using counterfactual thinking to compute a target's pivotality for an outcome.

The findings of developmental differences in attribution of responsibility in the retrospective judgements (and its components SP and SRC) but not in the prospective judgments (and its component SPC) are consistent with the differences found in propositional reasoning studies with differences in retrospective-counterfactual but not in prospective-prefactual conditionals. There are several reasons for the greater complexity of counterfactuals. Among them, a) counterfactual thinking involves mental simulations to think of the conjectured possibilities but, unlike with prefactuals, people need to maintain in mind the outcome (presupposed facts) to be evaluated relative to the events that they replace (Beck, 2016). Also, b) the difficulty in representing two things instead of just one (Byrne, 2016), and c) the need to follow the epistemic status of the cases (conjectured vs real), has been proposed as the differential feature between counterfactuals and “if then” conditions (Byrne, 2016; Espino & Byrne, 2021; Moreno-Ríos, et al. 2008).

**Development of counterfactual thinking and the overdetermined condition**

Although competence in thinking counterfactually improves throughout childhood, it is not clear which are the main factors that change. Some counterfactual tasks are solved by young children, but others are not solved until the age of 13 or 14. Differences found in various studies could be due to multiple factors. Although most of them employ quite similar stimuli (such as dynamic events and narratives), some results support explanations based on the ability to distinguish a real from a hypothetical situation (Gómez-Sánchez, et al., 2020) and the cognitive demands this entails (Carey et al., 2020; Gómez-Sánchez et al., 2020; Rafetseder et al., 2013, 2021), such as executive functions (Beck et al., 2009; Beck & Riggs, 2014). Kominsky et al. (2021), using videos where two balls collide, considered the possibility that the discrepancies found in previous studies were due to children envisaging different counterfactual possibilities from adults. The results supported the idea that children can think counterfactually but do so in a way that differs from adults. Their proposal is consistent with the results obtained by Gómez-Sánchez et al. (2021) using a propositional task. These authors compared schoolchildren and adults, and found that children, unlike adults, construct the counterfactual possibility more readily by recovering concrete rather than abstract alternatives.

To be sure that children understand counterfactuals and not “if then” basic conditionals, Rafetseder et al. (2013) proposed the need for including “overdetermined conditions” in the tasks (also see, Moreno-Ríos & García-Madruga, 2002; Gómez-Veiga, et al., 2010; Espino et al., 2020, Moreno-Ríos et al., 2008). They showed that inferences with counterfactual conditionals lead to the same conclusion as if-then conditionals because people tend to interpret basic conditionals as biconditional. Thus, “if the brown wire worked, then the battery worked” is also usually understood as “if the brown wire did not work, then the battery did not work”. Nevertheless, the two interpretations can be discriminated merely by including an “overdetermined condition”. Rafetseder et al. (2013), unlike previous studies with preschoolers, used scenarios that fixed the conclusion (e.g., “the battery always works”). Some other authors (Gómez-Sánchez et al., 2020; Ruiz-Ballesteros & Moreno-Ríos, 2017; Santamaría et al., 2005) used a kind of counterfactual conditional known as a “semifactual conditional” that generates the same effect: the conclusion is fixed but using a conditional expression: “Even if the brown wire had worked, the battery would not have worked”; and the opposite: the battery could have worked even if the yellow wire had not worked. In both cases the interpretation of the semifactual conditional (the wire worked but the battery did not work) is different from the basic conditional (the wire worked and the battery worked or the wire did not work and the battery did not work either). In these studies, schoolchildren's performance on counterfactual tasks improved over time, eventually reaching adulthood with mature counterfactual reasoning strategies at the end of their school years.

It seems that simply solving the overdetermined condition is insufficient to explain developmental differences in counterfactual reasoning in schoolchildren. Nyhout and Ganea, (2019) showed that preschoolers could engage in counterfactual thinking in simple fiscal-causal tasks with two possibilities. To test it, they used the blicket detector task (see also McCormack et al. 2018 for comparable results). They created an overdetermined condition: Two objects could independently make the blicket detector "turn on a light" when they were sited on it. Four- and five-year-olds could infer that even if an object was removed and the other remained, the light was still on. As a result, including the overdetermined condition in a task seems to be necessary but not sufficient to produce developmental differences in counterfactual abilities in older schoolchildren. There are many differences between the mentioned task and the tasks used here. For example, in the battery and propositional tasks, children must mentally operate in conceptual worlds with many alternatives. The factors and their weights are not yet known. Future work should address these questions.

Going back to the present study, in the brief counterfactual reasoning questionnaire, we included 1) counterfactuals with the overdetermined condition due to contextual information and 2) semifactuals (the "pure" overdetermined condition). The two measures were related to both sensitivity to pivotality and to sensitivity to retrospective criticality (SRC and SP), respectively. So it might be that the different developmental patterns found in children's prospective and retrospective judgements were due to the kind of thinking involved in each judgement. Note that only when children had information about the outcome (retrospective judgements), did their sensitivity increase with age. These conditions are particularly difficult for children because they require computing overdetermined conditions with a fixed outcome, even if a different wire were functioning (Leahy et al. 2014).

**Multiple potential factors in development.**

Finally, although during childhood individuals increase their abilities to reason about causes, they show some difficulties in thinking about situations with multiple potential causes. In Siegler's (1978) classic balance problems, as in those described by Inhelder and Piaget (1958), only older children, and in some cases only adolescents, were able to solve some problems with devices. This kind of problem requires thinking about actual outcomes, considering previous evidence and anticipating what the outcome could have been in other cases. Later studies have shown that even preschoolers, in adequate conditions, can solve problems that were thought to require formal operational thinking (e.g., Andrews et al., 2009; Ebersbach, 2009; Koskuba et al., 2018; Schlottman & Wilkening, 2011). Also, some research has shown that the last years of school are critical in its development (Peteranderl & Edelsbrunner, 2020). Children in the present study integrated criticality and pivotality values in their responsibility judgements. Besides acquiring this ability, during this developmental stage children also develop the ability to correctly interpret counterfactual questions about the physical behaviour of basic objects (Nyhout & Ganea, 2019). This is the case with understanding indeterminacy and confounded comparisons. The finding that young children show some abilities in simple tasks does not mean that their thinking is the same as that of adolescents, as was shown in classical tasks (Kuhn, 2012; Muentener & Bonawitz, 2017). We have used a task adjusted in complexity to find differences between school years to explore how counterfactual thinking is related to the attribution of causal responsibility and, more specifically, to retrospective judgements. Results seem to confirm this relation, supporting the CP model. Unfortunately, there are many questions that our study cannot answer; for example, which aspect of counterfactual thinking is related to the causal attribution of responsibility and what is the nature of that relation. There are many things that change during this developmental period, such as knowledge, reading comprehension and executive function. Some of these could be related to the improvement in mature counterfactual reasoning. It is possible that there are more sensitive tasks that could explain the differences at earlier ages. Also, we intentionally chose to study causal attribution in non-human agents to simplify the study. An interesting test of the model could be developmental testing with human agents. With the well-known changes in the theory of the mind in childhood, it is possible that even without intending (Gweon et al. 2012), children's attribution of causality to human agents could be mediated by their ability to attribute knowledge and intention to others. As we have seen, the CP model provides a good framework to analyse the development of children's thinking about mechanical devices. It makes specific predictions about the contributing effect of both criticality and pivotality, as well as the relation between retrospective responsibility judgments and reasoning with counterfactuals (and particularly with semifactuals). The model could also predict a fact that we have confirmed: when a judgement is required to assess the target's pivotality for an outcome, older children are more sensitive than younger ones, but they are similarly sensitive when no pivotality assessment is required, as in prospective judgements.

**Declaration of interest statement**

The authors have no relevant financial or non-financial relationships to disclose. They declare that they have no conflict of interest.

## References


Andrews, G., Halford, G. S., Murphy, K., & Knox, K. (2009). Integration of weight and distance information in young children: The role of relational complexity. *Cognitive_Development, 24*(1), 49–60. http://doi.org/10.1016/j.cogdev.2008.07.005

Beck, S. R. (2016). Why what is counterfactual really matters: A response to Weisberg and Gopnik (2013). *Cognitive Science, 40*, 253–256. https://doi.org/10.1111/cogs.12235

Beck, S. R., & Guthrie, C. (2011). Almost thinking counterfactually: Children's understanding of close counterfactuals. *Child Development, 82*(4), 1189–1198. https://doi.org/10.1111/j.1467-8624.2011.01590.x

Beck, S. R., & Riggs, K. J. (2014). Developing thoughts about what might have been. *Child Development Perspectives, 8*(3), 175– 179. https://doi.org/10.1111/cdep.12082

Beck, S. R., Riggs, K. J., & Gorniak, S. L. (2009). Relating developments in children's counterfactual thinking and executive functions. *Thinking and Reasoning, 15*(4), 337–354. https://doi.org/10.1080/13546780903135904

Beck, S. R., Robinson, E. J., Carroll, D. J., & Apperly, I. A. (2006). Children's thinking about counterfactuals and future hypotheticals as possibilities. *Child Development*, *77*(2), 413–426. https://doi.org/10.1111/j.1467-8624.2006.00879.x

Burns, P., Riggs, K. J., & Beck, S. R. (2012). Executive control and the experience of regret. *Journal of Experimental Child Psychology, 111*(3), 501-515. https://doi.org/10.1016/j.jecp.2011.10.003

Byrne, R. M. J. (2016). Counterfactual thought. *Annual Review of Psychology, 67,* 135–157. https://doi.org/10.1146/annurev-psych-122414-033249

Carey, S., Leahy, B., Redshaw, J., & Suddendorf, T. (2020). Could it be so? the cognitive science of possibility. *Trends in Cognitive Sciences*, *24*(1), 3–4. https://doi.org/10.1016/j.tics.2019.11.007

Chockler, H., & Halpern, J. Y. (2004). Responsibility and blame: A structural-model approach. *Journal of Artificial Intelligence Research, 22*(1), 93–115. https://doi.org/10.1613/jair.1391

Collins, J. D., Hall, E. J., & Paul, L. A. (2004). Causation and counterfactuals. Cambridge, MA: MIT Press Cambridge.

Ebersbach, M. (2009). Achieving a new dimension: Children integrate three dimensions stimulus in volume estimations. *Developmental Psychology, 45*(3), 877–883. http://doi.org/10.1037/a0014616

Enzle, M.E., & Schopflocher, D. (1978). Instigation of attribution processes by attributional questions. *Personality and Social Psychology Bulletin, 4*(4)*,* 595–599. https://doi.org/10.1177/014616727800400420

Espino, O., & Byrne, R. M. J. (2021). How people keep track of what is real and what is imagined: The epistemic status of counterfactual alternatives to reality. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 47*(4), 547–570. https://doi.org/10.1037/xlm0000965

Espino, O., Byrne, R. M. J., & Johnson-Laird, P. N. (2020). Possibilities and the parallel meanings of factual and counterfactual conditionals. *Memory & Cognition, 48*(7), 1263–1280. https://doi.org/10.3758/s13421-020-01040-6

German, T. P. (1999). Children's causal reasoning: Counterfactual thinking occurs for "negative" outcomes only. *Developmental Science, 2*(4), 442–447. https://doi.org/10.1111/1467-7687.00088

Gerstenberg, T., & Lagnado, D. A. (2010). Spreading the blame: The allocation of responsibility amongst multiple agents. *Cognition, 115*(1), 166–171. https://doi.org/10.1016/j.cognition.2009.12.011

Gerstenberg, T., & Lagnado, D. A. (2012). When contributions make a difference: Explaining order effects in responsibility attribution. *Psychonomic Bulletin & Review, 19*(4), 729–736. https://doi.org/10.3758/s13423-012-0256-4

Gómez-Sánchez, J., Moreno-Ríos, S. & Frosch, C. (2021). Alternatives or syntactic negation? Adults' and children's preferences for constructing counterfactual possibilities. *Current Psychology.* https://doi.org/10.1007/s12144-021-02456-2

Gómez-Sánchez, J., Ruiz-Ballesteros, J. A., & Moreno-Ríos, S. (2020). How children and adults keep track of real information when thinking counterfactually. *PLOS ONE*, *15*(12), e0242967. https://doi.org/10.1371/journal.pone.0242967

Gómez-Veiga, I., García-Madruga, J. A., & Moreno-Ríos, S. (2010). The interpretation of indicative and subjunctive concessives. *Acta Psychologica, 134(2),* 245-252. https://doi.org/10.1016/j.actpsy.2010.02.005

Guajardo, N. R., Parker, J., & Turley-Ames, K. (2009). Associations among false belief understanding, counterfactual reasoning, and executive function. *British Journal of Developmental Psychology*, *27*(3), 681–702. https://doi.org/10.1348/026151008X357886

Gweon, H., Dodell-Feder, D., Bedny, M., & Saxe, R. (2012). Theory of mind perfomance in children correlates with functional especialization of bran region for thinking abouth thougts. *Child Development, 83*(6), 1853-1868. https://doi.org/10.1348/026151008X357886

Halpern, J. Y., & Pearl, J. (2005). Causes and explanations: A structural-model approach. Part I: Causes. *The British Journal for the Philosophy of Science, 56*(4), 843–887. https://doi.org/10.1093/bjps/axi147

Hilton, D. J. (1990). Conversational processes and causal explanation. *Psychological Bulletin, 107*(1), 65–81. https://doi.org/10.1037/0033-2909.107.1.65

Inhelder, B., & Piaget, J. (1958). The Growth of Logical Thinking from Childhood to Adolescence. New York: Basic Books. http://dx.doi.org/10.1037/10034-000

Kominsky, J. F., Gerstenberg, T., Pelz, M., Sheskin, M., Singmann, H., Schulz, L., & Keil, F. C. (2021). The trajectory of counterfactual simulation in development. *Developmental Psychology, 57*(2), 253–268. https://doi.org/10.1037/dev0001140

Koskuba, K., Gerstenberg, T., Gordon, H., Lagnado, D., & Schlottmann, A. (2018). What's fair? How children assign reward to members of teams with differing causal structures. *Cognition, 177*, 234-248. https://doi.org/10.1016/j.cognition.2018.03.016

Kuhn, D. (2012). The development of causal reasoning. WIREs Cognitive Science, 3(3), 327-335. https://doi.org/10.1002/wcs.1160

Lagnado, D. A. & Gerstenberg, T. (2015). A difference-making framework for intuitive judgments of responsibility. In *Oxford Studies in Agency and Responsibility* (pp. 213-241). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199694853.001.0001

Lagnado, D. A., & Gerstenberg, T. (2017). Causation in legal and moral reasoning. In M. R. Waldmann (Ed.), *The Oxford handbook of causal reasoning* (pp. 565–601). Oxford University Press.

Lagnado, D. A., Gerstenberg, T., & Zultan, R. (2014). Causal responsability and counterfactuals. *Cognitive Science, 37*(6), 1036-1073. https://doi.org/10.1111/cogs.12054

Leahy, B., Rafetseder, E., & Perner, J. (2014). Basic conditional reasoning: How children mimic counterfactual reasoning. *Studia Logica, 102*, 793–810. https://doi.org/10.1007/s11225-013-9510-7

Leo, X., & Huh, Y.E. (2020). Who gets the blame for service failure? Attribution of responsibility toward robot versus human service providers and service firms. *Computer in Human Behavior, 113,* 106520. https://doi.org/10.1016/j.chb.2020.106520

Lewis, D. (1973). Causation. *The Journal of Philosophy, 70*(17), 556–567. http://dx.doi.org/10.2307/2025310

Lewis, D. (1986). Causal explanation. *Philosophical Papers, 2*, 214–240. http://dx.doi.org/10.1093/0195036468.001.0001

McCormack, T., Ho, M., Gribben, C., O'Connor, E., & Hoerl, C. (2018). The development of counterfactual reasoning about doubly-determined events. Cognitive Development, 45, 1–9. https://doi.org/10.1016/j.cogdev.2017.10.001.

Moore, M. S. (2009). Causation and responsibility: An essay in law, morals, and metaphysics. Oxford: Oxford University Press.

Moreno-Rios, S., & García-Madruga, J. A. (2002). El desarrollo del razonamiento sobre lo que podría haber ocurrido: condicionales indicativos y subjuntivos. *Infancia y Aprendizaje*, *25*(3), 485–498. http://dx.doi.org/10.1174/021037002762064055

Moreno-Ríos, S., García-Madruga & Byrne, R. (2008). Inferences from Semifactual 'Even if' Conditionals. Acta Psicológica. *Acta Psychologica, 128*, 197-209.

https://doi.org/10.1016/j.actpsy.2007.12.008

Muentener, P., & Bonawitz, E. (2017). *The development of causal reasoning.* In M. R. Waldmann (Ed.), *Oxford library of psychology. The Oxford handbook of causal reasoning* (p. 677–698). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199399550.013.40

Nyhout, A., & Ganea, P. A. (2019). Mature counterfactual reasoning in 4- and 5-year-olds. *Cognition, 183*, 57-66. https://doi.org/10.1016/j.cognition.2018.10.027

Nyhout, A., Henke, L., & Ganea, P. A. (2019). Children's counterfactual reasoning about causally overdetermined events. *Child Development, 90*(2), 610-622. https://doi.org/10.1111/cdev.12913

O'Connor, E., McCormack, T., & Feeney, A. (2012). Do children who experience regret make better decisions? A developmental study of the behavioral consequences of regret. *Child Development, 85*(5), 1995-2010. http://dx.doi.org/10.1111/cdev.12253

Pearl, J. (2009). Causality: Models, reasoning and inference (2nd Edition). New York: Cambridge University Press. https://doi.org/10.1017/CBO9780511803161

Peteranderl, S., & Edelsbrunner, P. A. (2020). The predictive value of children's understanding of indeterminacy and confounding for later mastery of the control-of-variables strategy. *Frontiers in Psychology*, *11*, PMID: 33335493. https://doi.org/10.3389/fpsyg.2020.531565

Santamaría, C., Espino, O., & Byrne, R. M. J. (2005). Counterfactual and Semifactual Conditionals Prime Alternative Possibilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(5), 1149–1154. https://doi.org/10.1037/0278-7393.31.5.1149

Schlottmann, A., & Wilkening, F. (2011). Judgment and decision making in young children. In M. K. Dhami, A. Schlottmann, & M. R. Waldmann (Eds.). Judgement and decision making as a skill: Learning, development, evolution (pp. 55–84). Cambridge, UK: Cambridge University Press.

Siegler, R. S. (1978). Children's Thinking: What Develops? Hillsdale, NJ: Lawrence Erlbaum.

Steiner, I. D. (1972). Group process and productivity. New York: Academic Press.

Rafetseder, E., Cristi-Vargas, R., & Perner, J. (2010). Counterfactual reasoning: Developing a sense of "nearest possible world." *Child Development*, *81*(1), 376–389. https://doi.org/10.1111/j.1467-8624.2009.01401.x

Rafetseder, E., O'Brien, C., Leahy, B., & Perner, J. (2021). Extended difficulties with counterfactuals persist in reasoning with false beliefs: Evidence for teleology-in-perspective. *Journal of Experimental Child Psychology*, *204*, 105058. https://doi.org/10.1016/j.jecp.2020.105058

Rafetseder, E., Schwitalla, M., & Perner, J. (2013). Counterfactual reasoning: From childhood to adulthood. *Journal of Experimental Child Psychology*, *114*(3), 389–404. https://doi.org/10.1016/j.jecp.2012.10.010

Rafetseder, E., & Perner, J. (2018). Belief and counterfactuality: A teleological theory of belief attribution. *Zeitschrift für Psychologie, 226*(2), 110-121. http://dx.doi.org/10.1027/2151-2604/a000327

Rapoport, A. (1987). Research paradigms and expected utility models for the provision of step-level public goods. *Psychological Review, 94*(1), 74-83. https://doi.org/10.1037/0033-295X.94.1.74

Roldán-Tapia, M. D., Moreno-Ríos, S., & Cánovas-López, R. (2017). Thinking about social and nonsocial alternative possibilities in premature preschoolers. *Journal of Clinical and Experimental Neuropsychology*, *39*(8), 725–737. https://doi.org/10.1080/13803395.2016.1257703

Ruíz-Ballesteros, J.A. & Moreno-Ríos, S. (2016). Inherent directionality of “even if”? counterfactual conditionals. *Journal of Cognitive Psychology*, *28*(4),505-511. https://doi.org/10.1080/20445911.2016.1142996

Ruiz-Ballesteros, J. A., & Moreno-Ríos, S. (2017). Concessive and semifactual interpretations during reasoning with multiple conditionals. *The Quarterly Journal of Experimental Psychology, 70*(7), 1140-1150. https://doi.org/10.1080/17470218.2016.1172098

Weisberg, D. P., & Beck, S. R. (2012). The development of children's regret and relief. *Cognition and Emotion, 26*(5), 820-835. https://doi.org/10.1080/02699931.2011.621933

Woodward, J. (2003). Making things happen: A theory of causal explanation. Oxford: Oxford University Press. https://doi.org/ 10.1093/0195155270.001.0001

Zultan, R., Gerstenberg, T., & Lagnado, D. A. (2012). Finding fault: Causality and counterfactuals in group attributions. *Cognition, 125*(3), 429–440. https://doi.org/10.1016/j.cognition.2012.07.014

**Table 1.** Mean ratings (SDs) for prospective and retrospective judgements (from 0 to 10) per condition regarding Criticality (C) and Pivotality (P) values, for Adults (Experiment 1) and Children (Experiment 2).

| | **Adults** | **Children** | | |
|---|---|---|---|---|
| | | **8 to 10 years** | **10 to 13 years** | **8 to 13 years** |
| **Prospective Judgments** | | | | |
| **C = 1** | 8.38 (2.14) | 8.28 (2.28) | 8.71 (1.77) | 8.48 (2.06) |
| **C = ½** | 2.69 (1.91) | 4.56 (2.66) | 5.39 (3.20) | 4.95 (2.95) |
| **Retrospective Judgments** | | | | |
| **P = 1 [C=1]** | 9.68 (0.73) | 8.41 (2.04) | 7.72 (2.46) | 8.09 (2.27) |
| **P = 1 [C = ½]** | 5.53 (2.09) | 5.59 (2.33) | 5.84 (2.43) | 5.71 (2.38) |
| **P = ½ [C=1]** | 6.88 (1.84) | 6.52 (2.19) | 6.89 (2.42) | 6.69 (2.30) |

**Table 2**. Correlations between the two indices (SC and SP), the three counterfactual questions and the CF index controlling for age.

| | **Inferential** | **Epistemic Status** | **Semifactual** | **CF index** |
|---|---|---|---|---|
| **SPC index** | *r* = .08, *p* = .282 | *r* = .07, *p* = .355 | *r* = -.01, *p* = .339 | *r* = .06, *p* =.391 |
| **SRC index** | *r* = .15*, *p* = .048 | *r* = .04, *p* = .613 | *r* = .12, *p* = .126 | *r* = .15, *p* = .057 |
| **SP index** | *r* = .09, *p* = .217 | *r* = -.07, *p* = .379 | *r* = 0.18*, *p* = .022 | *r* = .09 *p* = .213 |

*Note:* * p < .05. **p < .01.

**Figure 1**. Screenshot of the two criticality value conditions in the prospective judgements by asking participants about the leftmost wire. The criticality in panel A is 1 (brown wire) and in panel B ½ (yellow wire). The sensitivity to criticality is computed as the difference between the two conditions.

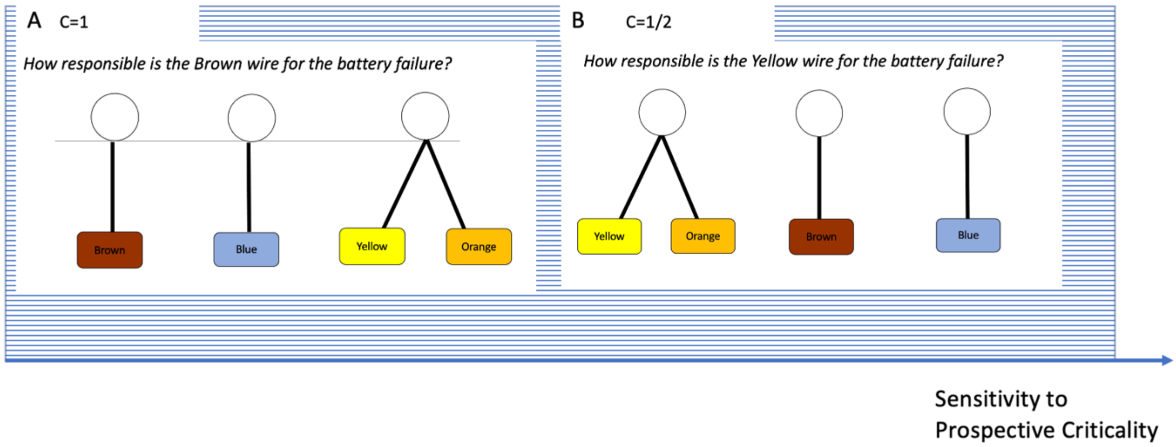

**Figure 2**. Screenshot of the three conditions in the retrospective judgments by asking participants about the leftmost wire. Panels A and B have the same pivotality value (1) but different Criticality values (1 and ½). The difference between them is the sensitivity to Criticality after outcome knowledge (*Retrospective Criticality*). Panels A and C have the same criticality value (1) but different pivotality values (1 and ½). The difference between them is the sensitivity to Pivotality.