Bachelor's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Social and Educational Sciences
Department of Psychology

10019

Revisiting The Temporality Effect of Counterfactual Thinking

Replication and extension of Miller and Gunasegaram (1990) and Segura, Fernandez-Berrocal, and Byrne (2002)

Bachelor's thesis in Psychology
Supervisor: Subramanya Prasad Chandrashekar
May 2022

NTNU
Norwegian University of
Science and Technology

10019

Revisiting The Temporality Effect of Counterfactual Thinking

Replication and extension of Miller and Gunasegaram (1990) and Segura, Fernandez-Berrocal, and Byrne (2002)

Bachelor's thesis in Psychology
Supervisor: Subramanya Prasad Chandrashekar
May 2022

Norwegian University of Science and Technology
Faculty of Social and Educational Sciences
Department of Psychology

NTNU
Norwegian University of
Science and Technology

PREFACE

This bachelor thesis marks the end of three educational and exciting years at the Norwegian University of Science and Technology in Trondheim. Through my years studying psychology I have been introduced to a lot of interesting subjects and have developed academically and experienced growth within myself. I took part in the project supervised by Subramanya Prasad Chandrashekar "Temporality effect and counterfactual thinking", as a part of a research group of students. The literature used in the thesis have been retrieved from Google Scholar, Oria (i.e., NTNU`s digital library), APA PsycNet and ScienceDirect. I would like to thank my supervisor, Subramanya Prasad Chandrashekar, for his patience and exceptional approach of teaching and guiding the research group throughout this process. Prasad has provided the research group with lectures, shared relevant articles, aided in development of pre-registrations and survey, and collected data for two of the studies. Additionally, he has supported the research group throughout this process, by being available both by email and in person to answer questions, discuss problems, provide conceptual clarity, and offering comments for guidance on the thesis. He also assisted me by providing the idea for the extension in this project. Further I want to thank the research group for valuable input and support throughout this process. The academic discussions and uplifting conversations have been deeply appreciated. The group have collectively completed pre-registration and translation of the information sheet and survey for the replication of Miller and Gunasegaram (1990). Finally, I want to thank friends and family for always being supporting and encouraging and giving me advice and always believing in me. With all this being said, I hereby declare this work as my own.

Word count: approx. 6442

ABSTRACT

When people imagine alternatives to reality, the order of which the events occur in a sequence influence the way they think an outcome could have turned out differently. The temporality effect is the tendency to mutate the more recent event in sequences of independent events. We conducted two pre-registered close replications of Miller and Gunasegaram (1990; $N = 211$) and Segura, Fernandez-Berrocal, and Byrne (2002; $N = 664$) and included an extension ($N = 327$) to test if the temporality effect also occurs when the statistical possibilities of winning outcomes were manipulated. The three experiments were based on simple scenarios in which two or four individuals participated in games where they either tossed a coin, drew cards from a deck, or drew marbles from a sack consisting of 30 red and 70 blue marbles. To win the game they had to individually come up with the same result (i.e., both tossing heads), and in all conditions the players lost the game. Participants were asked which of the events were easier to imagine changing to obtain a winning outcome, and in two of the experiments, participants were also asked who they think will experience more guilt, and who they predict will blame the other more for their failure to win. The two replications found support for the confirmatory predictions that people tend to mutate the more recent event, while the extension found that the statistical probability of possible outcomes affected this tendency.

The ability to imagine alternatives to reality is a pervasive aspect of our mental lives, that occur when people compare an actual situation to an imaginary alternative (Miller & Gunasegaram, 1990; Segura & McCloy, 2003; Byrne, 2016). Counterfactual alternatives to reality are often mentally constructed through "if only" or "what if" scenarios and serve several purposes in navigating the social and natural world, including explaining and understanding the past, learning from experiences, and preparing for the future (Byrne, 2016). Some aspects of reality are easy to imagine otherwise, like; if I had left two minutes earlier, I would have caught the bus, whereas others are more resistant to counterfactual alternatives (Miller & Gunasegaram, 1990). Research has also shown that when people imagine how an outcome could have happened differently, they tend to undo the more recent event when the events are independent of each other (Miller & Gunasegaram, 1990).

The temporality effect is known as a tendency to view the last event of a sequence of independent events as more mutable (i.e., easier to mentally undo) than the first event (Miller & Gunasegaram, 1990; Byrne et al., 2000; Segura et al., 2002). In a football game that resulted in a 2-1 win, each of the two goals of the winning team provides an equal contribution to the end result. Thus, the goal scored in the beginning of the match should be viewed as an equally important contribution to the outcome as the winning goal scored just before end time. However, people still tend to view the last goal as having a greater impact on the outcome, and the person scoring the decisive goal is even considered matchwinner. This shows to how we make judgements about events based on the temporal order they occur in, and that some events are given more causal strength than others (Miller & Gunasegaram, 1990; Byrne et al., 2000; Spellman, 1997; Henne, Kulesza, Perez & Houcek, 2021).

The purpose of the present study is to revisit the temporality effect by replicating some of the first demonstrations of the effect in sequences of two and four independent events (Miller & Gunasegaram, 1990; Segura et al., 2002). We will also include an extension with the aim of

further investigating the robustness of the temporality effect, by introducing a new scenario where the statistical probabilities of the counterfactual possibilities are manipulated.

Counterfactual thinking

When imagining how a situation could have turned out differently, there are a number of counterfactuals that could account for an altered outcome (Byrne, 2016). If we look at the football scenario presented earlier, there are several ways in which the team could have lost. We have already visited the idea of the first or second goal not being scored, but one can also imagine neither of them scoring, the opposing team scoring, or that football simply don't exist. The research on counterfactual thinking has mainly focused on what aspects of reality people tend to undo, as people don't seem to treat all counterfactual alternatives equally (Byrne et al., 2000; Byrne, 2016; Kahneman & Miller, 1986; Henne, Kulesza, Perez & Houcek, 2021). When imagining alternatives to reality, people look for effective counterfactuals, meaning alternative possibilities that change the outcome. The cognitive mechanisms involved in counterfactual thinking have also shown to be guided by how easy it is to imagine a counterfactual to it (Byrne et al., 2000), and that the availability could differ based on the causal structure of sequences (Henne et al., 2021). For instance, in a causal sequence, where one event cause one or more preceding events, people seem to view the first event as more mutable, known as the causal primacy effect (Wells, Taylor, & Turtle, 1987), whereas they tend to undo the more recent event in sequences of independent events (Kahneman & Miller, 1986; Miller and Gunasegaram, 1990).

Temporal order effect in counterfactual thinking

Kahneman & Miller (1986) were the first to explore the temporal order effect, where they found that when presented with an ordered sequence of letters (i.e., *xf*) and asked to quickly change one, participants tended to mutate the second letter more than the first. Miller and

Gunasegaram (1990) followed this finding by presenting a coin-toss scenario where two individuals named Jones and Cooper were given the opportunity to win a substantial amount of money if they were able to individually toss a coin where both coins came up the same (i.e., both heads or both tails). Jones went first and tossed heads, and Cooper went second and tossed tails. Thus, neither individual won anything. When people are asked which event they imagine happening differently for Cooper and Jones to win the game, they tend to report Cooper tossing heads. This tendency reflects the person scoring the decisive goal in the football scenario being recognized as matchwinner, and the effect has been demonstrated in several studies attempting to further elucidate the cognitive processes that guide mental representation of counterfactuals in temporal sequences (e.g., Miller & Gunasegaram, 1990; Byrne et al., 2000; Segura et al., 2002; Walsh & Byrne, 2004). Most research exploring the temporality effect suggest that the observed effect of mutating the more recent event is because of a presupposition or immutability of the first event (Miller & Gunasegaram, 1990; Byrne et al., 2000; Walsh & Byrne, 2004; Segura et al., 2002). However, the theoretical mechanisms that account for the effect have been challenged and developed continuously attempting to account for all the mechanisms that guide the effect of mutating the more recent event.

Walsh and Byrne (2004) came up with six principles that explain how people mentally represent alternatives to an outcome, and how mutability of an event differs depending on the accessibility of alternatives to it. The three initial principles of the model account for representational assumptions that are present when considering alternative counterfactuals, where the two first are that people understand scenarios by considering the true possibilities of an outcome, and not necessarily the full set of counterfactual possibilities (Walsh & Byrne, 2004; Byrne et al., 2000). True possibilities represent the facts of the actual situation, which in the coin-toss scenario would be Jones tossing heads, Cooper tossing Tails, and the fact that they lost. The representation of the facts of the situation is kept in mind and the information is used

to further guide the counterfactual possibilities. In the coin-toss scenario there are several ways the outcome could have turned out differently; Jones tossing tails, Cooper tossing heads and they lose, both tossing tails and they win, or both tossing heads and they win. However, the second principle suggest that people do not keep in mind all counterfactual possibilities, as only a subset of them are considered in mental modeling of possibilities. As mentioned, the cognitive process of creating counterfactuals is guided by mental representation of the effective counterfactuals of an outcome. This assumption makes up the third principle, as the counterfactual possibilities that people consider are guided by the winning possibilities, and thereby the counterfactual possibility of Jones tossing tails and Cooper tossing heads where they lose, is not kept in mind as this does not alter the outcome.

The two following principles of the model represent strategies that mediate the way people manipulate the representational assumptions (Walsh & Byrne, 2004). Principle number four assumes that people mutate the aspects that are crucial in order to be consistent with the winning conditions. The fifth principle represent the assumption of an immutability or presupposition of the first event in sequences of independent events, where the first event serves as an anchor that is held constant if a match to the winning possibilities is identified. In the coin toss scenario, the first event of Jones tossing heads is matched to the winning possibilities, and because the winning condition of both tossing heads is present, people will attempt to match the subsequent event of Cooper tossing tails to obtain a winning outcome. If there is no match between the first event and the winning possibilities however, the first event is mutated to match the winning possibilities. The final principle involves that people mentally represent some elements of the true possibilities explicitly, while other elements are left implicit.

The crediting causality hypothesis

The crediting causality hypothesis provides an alternative explanation to the temporality effect, by arguing that the causal strength of events in a sequence varies according to what extent it changes the subjective probability of an outcome (Spellman, 1997). In context of the coin-toss scenario presented by Miller and Gunasegaram (1990), the crediting causality hypothesis explains the observed effect of mutating the last event as a result of a change in the probability of a winning outcome (i.e., both tossing heads or both tossing tails) between the first individual tossing heads and the second individual tossing tails. Before the first individual tosses the coin, there is a 50-50% chance of tossing heads or tails, and after the first player tosses heads, there is still a 50-50% chance of a winning outcome. However, when the second individual tosses tails, the probability of them winning changes to 0 to 1, meaning that if the event of the second player were different, there would be 100% chance of a winning outcome. The change in probability changes more following the second event and is therefore viewed as having greater causal strength.

The need for replication

Following the replication crisis, a growing recognition has been directed towards the importance of replicability, reproducibility, and open science (LeBel, 2018; LeBel, 2019; Open Science Collaboration, 2015; Earp & Trafimow, 2015). A study conducted by Open Science Collaboration (2015) investigating the replicability of 100 studies within the field of psychology suggested a surprisingly low replicability in the psychological science, where only 36 to 47% of the studies were successfully replicated. Considering that some of the main aims and expectations of science is to achieve a deeper understanding of the social and natural world (LeBel et al., 2018), it raises concern that a substantial part of findings fails to replicate. Reproducibility is one of the defining features within the scientific community (Open Science Collaboration, 2015), as surviving attempts of falsification repeatedly builds credibility to

findings (LeBel et al., 2018). Science depends on credibility and trust in order to be useful to practitioners, policymakers, the academic discipline, and society in general (Open Science Collaboration, 2015; Earp & Trafimow, 2015), as application of inaccurate and potentially misleading information could have consequences for decision-making and academic progress (LeBel et al., 2018; Ioannidis, 2005). Conducting well-powered replicating studies applying the principles of open science, can both develop and assess the credibility of the findings we already have (LeBel et al., 2018), and further develop frameworks by bringing attention to problems regarding how to conduct and interpret research (Ioannidis, 2005; Earp & Trafimow, 2015).

Choice of replication targets and present investigation

The implementation of common methodical guidelines and frameworks to ensure reliable research has changed a lot over time, and we see that most studies exploring the temporality effect lacks fundamental qualities to confirm reliability and validity (e.g., reporting of effect sizes, method and data transparency, and pre-registrations). We chose Miller and Gunasegaram (1990) and Segura et al. (2002) as our replication targets to provide direct, well-powered, pre-registered replications of two important contributions exploring the temporality effect in sequences of two and four independent events.

Miller and Gunasegaram (1990) were one of the first to demonstrate and establish the term temporality effect to the observed effects of mutating the last of two or more independent events. The study presented a coin-toss scenario described above, in addition to two studies in context of an exam sequence, with the aim of discovering stronger evidence elucidating the relation between mutability and temporal order (Miller & Gunasegaram, 1990). As of May 2023, the article has been cited 339 times, and the findings have impacted further scientific work on establishing the mechanisms that cause the effect of mutating the last event in temporal sequences (e.g., Spellman, 1997; Byrne et al., 2000). The researchers failed to provide a proper method and result section and used a sample of only 88 undergraduates. For these reasons, we

chose to conduct a replication of the coin-toss scenario on a Norwegian sample of 211 participants.

Segura et al. (2002) conducted a study testing if the tendency to mutate the last event in a sequence of two independent events also occurred in sequences of four events. As our primary interest in this study were the temporality effect, we chose to replicate only the first experiment investigating if the temporality effect occurs for sequences of four independent events. The experiment consisted of four scenarios, however, we only replicated three of them (i.e., scenario 2, 3 and 4), considering scenario 1 measured the temporality effect in a sequence of two independent events, which we already cover in the replication of Miller and Gunasegaram (1990). The measure items consisted of open-ended "if only" questions concerning which of the four events were altered in their mind when imagining an alternative outcome, and statistical analysis consisted of one sample *z*-tests measuring the count measures of the fourth event against the first, second and third event. The authors hypothesized that if the temporal order effect occur because the first event is presupposed, the effect should be present in sequences of four events as well as for sequences of two events. They also hypothesized that if the temporality effect occurs because the second event is particularly mutable, the second event should be mutated most often even in sequences of four independent events. Their results found support for both predictions, supporting the presupposition of the first event, where participants mutated the last event more than the first in all scenarios, and mutated the last event more than the second event in all scenarios except scenario 3, where they equally mutated the second and fourth event. Thus, the results imply that the temporality effect occur based on a presupposition or immutability of the first event rather than a particular mutability of the second event in sequences of independent events. For our replication we measured mutability with the same open-ended question as the original study, however, instead of asking which of the four events

were altered in their mind, we only asked them which of the first or fourth event that came more readily to mind when imagining an alternative series of outcomes.

Effect sizes in target articles

Neither of the two replication targets reported effect sizes. We reanalyzed the data using the information provided in the original articles to calculate the effect sizes. Effect sizes of the original studies are summarized in Table 5 (for more detailed results, see pre-registrations of the individual articles in supplementary materials, Appendix A, B, and C).

Extension

In addition to the replication studies, we also included an extension where we investigated whether the statistical probability of the counterfactual possibilities affected the tendency to mutate the last event of two independent events. Our interest with this extension also included to further examine the robustness of the temporality effect, moreover, it allowed us to gain insight to the robustness of the immutability of the first event. To examine this, we constructed two scenarios where the conditions were changed from a 50/50 to a 70/30 percent chance event. The scenarios were based on the coin-toss scenario in Miller and Gunasegaram (1990), but instead of tossing a coin, the players picked one marble each from a sack consisting of 30 red marbles and 70 blue marbles. Similar to the coin-toss scenario, both scenarios had negative outcomes, and the only difference between the two experimental conditions were which colored marble the first player picked (i.e., either red or blue marble). Our predictions were the same across both conditions, where we hypothesized that the more recent event would be more easily mutated than the first event, and that people would attribute that the second player would experience more guilt and be blamed more by the first player. Special interest will be directed to the results of the scenario in which the first person picks the red marble (i.e.,

Scenario 1), as this is where the probability of a good outcome is the most affected by the first event, and thereby challenging the robustness of the temporality effect.

Evaluating criteria for replication design and findings

The replication designs were evaluated using the criteria developed by LeBel, McCarthy, Earp, Elson, and Vanpaemel (2018), and summarize the current replication of both Miller and Gunasegaram (1990) and of Segura et al. (2002) as close replications (See supplementary materials Table S2 and S3). To interpret the results of the replications compared to the original studies, we used the framework by LeBel, Vanpaemel, Cheung, and Campbell (2019; See supplementary materials, Table S4 and S5).

Pre-registration, power analysis, and open science

The experiments were pre-registered on the Open Science Framework (OSF) before data collection was launched. All materials related to the three experiments are presented in the supplementary materials. Pre-registrations and power analysis are available in the supplementary materials (Appendix A, B and C), and on the OSF: Study 1: https://osf.io/gpcqd; Study 2: https://osf.io/yq8a9; Study 3: https://osf.io/nk6fb.

Ethical considerations

The study was approved by Norwegian Agency for Shared Services in Education and Research (Sikt; 754105; assessment provided in supplementary, Appendix D), and in line with the Norwegian Research Ethics Committees guidelines (NESH, 2022) and NTNU`s ethical guidelines (NTNU, s.a.).

Study 1 - Replication of Miller & Gunasegaram`s (1990) coin-toss scenario

Method

Participants

A total of 211 participants consisting of Norwegian adults ($Median_{age}$ = 23, M_{age} = 28.51, SD_{age} = 13.68) were collectively recruited by the research group by sharing a link to the online survey on social media platforms and placing posters with some information and a QR code to the survey around the campus of NTNU. Thus, recruitment was conducted using convenient sampling, and took place in the timeframe between February 25th and March 20th . Note that median of age is reported to give a more realistic measure of the distribution, due to a large portion of the sample (66.7%) being 24 years old or younger. The sample consisted of 138 females (65.4%), 72 male (34.1%), and 1 "other" (0.5%), and age ranged from 19 to 90 years old. Participants did not receive any form of compensation by participating.

Procedure and design

The subjects participated through a link to an online survey on the digital platform "Nettskjema", where they were presented with an information sheet and consented to take part in the study. Next, they were given a written scenario based on the coin-toss scenario in the original study by Miller and Gunasegaram (1990). Following the scenario, they were asked three questions relevant to the scenario measuring mutation, and attribution of guilt and blame, and then answered a few demographic questions. As this was a replication study, both the written scenario, measure items and the choice of analysis were operationalized and conducted in line with the original study. Due to limitations in the original study concerning choice of statistical analysis, we chose to conduct one sample z-tests to test the predictions. For further

information regarding the procedure, survey, and measure items, please refer to the supplementary materials (Appendix A).

Results

Table 1

Percentages and count measures of mutation, guilt and blame in Study 1

Dependent variables	In %	Participants
Mutation		
First	24	50
Second	76	161
Guilt		
First	7	14
Second	93	197
Blame		
First	8	17
Second	92	194
n	211	

A one sample z-test revealed a tendency to mutate the second event of two independent events (76% vs. 24%), $n = 211$, $k = 161$, $z = 7.64$, $p < .001$. For guilt and blame it indicated that people tend to attribute that the second player will experience more guilt (93% vs. 7%), $n = 211$, $k = 197$, $z = 12.6$, $p < .001$, and that the first player is more likely to blame the second player (92% vs. 8%), $n = 211$, $k = 194$, $z = 12.19$, $p < .001$. Effect sizes for all three measures is presented in Table 5.

Study 2 - Replication of Study 1 of Segura et al. (2002)

Method

Participants

664 participants ($M_{age} = 38.88$, $SD_{age} = 14.02$) were recruited online through Prolific platform and consisted of English-speaking adults. The sample consisted of 300 females

(45.2%), 346 male (52.1%), 16 non-binary (2.4%), and 2 that prefer not to say (0.3%), and age ranged from 10 to 77 years old. Given that recruiting was done through Prolific platform, we chose to ignore the fact that one participant reported themselves as 10 years of age, and assumed this was the result of a simple error in response, knowing Prolific platform only accept participants above the age of 18 years old. At the stage of pre-registration, we intended to exclude participants based on criteria such as proficiency in English, seriousness of filling out the survey, participants outside of the US, and completion of the survey too quickly (i.e., within one minute). However, we chose to not apply these exclusion criteria to achieve as generalizable results as possible.

Procedure and design

Participants were randomly and evenly assigned to one of three experimental conditions: Scenario 1 (n = 220); Scenario 2 (n = 221); or Scenario 3 (n = 223). Participation was carried out through an online survey, where participants were presented with an information sheet where they consented to take part in the study. Upon consenting, they were taken to the next page, where they were asked to read a written scenario. All three scenarios were formulated with the same wording, apart from the order of colored cards that were picked (i.e., Scenario 1- Red-Black-Black-Red; Scenario 2- Red-Black-Red-Black; Scenario 3- Red-Red-Black-Black). Description of Scenario 1 is presented below:

Imagine four individuals (Jones, Michael, Frank, and Daniel) who are offered the following very attractive proposition. Each individual is given a shuffled deck of cards, and each one picks a card from their own deck. If three of the four cards they pick are of the same color (i.e., three from black suits or three from red suits), each individual win $1,000. However, if three of the cards are not the same color, none of the individuals win anything. Jones goes first and picks a ***red*** *card from his deck. Michael goes second and picks a* ***black*** *card. Frank goes third*

*and picks a **black** card. And Daniel picks a **red** card. Thus, the outcome is that neither of the individuals win anything.*

Following the written scenario, participants were presented with a visualization of the scenario for better understanding. Finally, they were asked one item measuring mutability with two response alternatives, followed by a few demographic questions (i.e., age, gender). To be included in the data collection, participants had to answer all the items noted above, including consent, measure of mutability, and demographic questions. Similar to the previous study, both the written scenario, measure items, and choice of analysis were operationalized and conducted to match the original study as close as possible. This included using the same scenario, with minor changes to match the target sample (original study were conducted on a Spanish population while ours were on an American population). The statistical analysis was the same (i.e., *z*-tests). For further information regarding the procedure and measure items, please refer to supplementary materials (Appendix B).

Results

Table 2

Percentages and count measures of mutations across three scenarios in Study 2

Dependent variables	*n*	In %	Participants
Scenario 1	220		
First		29	63
Fourth		71	157
Scenario 2	221		
First		19.5	43
Fourth		80.5	178
Scenario 3	223		
First		13.5	30
Fourth		86.5	193

In Scenario 1, a one sample z-test showed that people tend to find the last event in a sequence of four independent events more mutable than the first (71% vs. 29%), $n = 220$, $k = 157$, $z = 6.34$, $p < .001$. The same effect is found in Scenario 2 (80.5% vs. 19.5%), $n = 221$, $k =$

178, $z = 9.08$, $p < .001$, and Scenario 3 (86.5% vs. 13.5%), $n = 223$, $k = 193$, $z = 10.92$, $p < .001$. Effect sizes for all three scenarios is presented in Table 5.

Study 3 - Extension of the temporality effect: Statistical possibilities

Method

Participants

A total of 327 participants were recruited online through Prolific platform (M_{age} = 38.79, SD_{age} = 13.43). The sample consisted of 189 female (57.8%), 128 male (39.1%), and 10 non-binary (3.1%). Age ranged from 18 to 78 years old. As the previous study, we intended to exclude participants based on criteria such as proficiency in English, seriousness of filling out the survey, and participants that completed the survey too quickly (i.e., within one minute) at the stage of pre-registration. However, we chose to not apply these exclusion criteria to achieve as generalizable results as possible.

Procedure and design

Participants were randomly and evenly assigned to one of two experimental conditions: Scenario 1 (n = 163); or Scenario 2 (n = 164). As in the two previous studies, participants were presented with an information sheet and consented to take part in the study. The subjects were then presented a written scenario. The scenario was formulated in the same way across the two conditions, apart from which colored marble the first person picked. Description from scenario 1 is presented below:

Imagine two individuals (Linda and Barbara) who are offered the following very attractive proposition. Each individual is handed a sack of marbles that contains 30 red marbles, and 70 blue marbles. The players get to pick one marble each without looking. If both players pick the same-colored marble (i.e., both red marbles, or both blue marbles) they both win $1,000.

However, if the marbles are not the same color, neither of the players win anything. Linda goes first and picks a red marble. Barbara goes next and picks a blue marble. Thus, the outcome is that neither of the individuals win anything.

Following the scenario, participants were asked three measure items relevant to the scenario, including one item on mutability, one item on attribution of guilt, and one item on attribution of blame, where all the items had two response alternatives. Finally, they answered a few demographic questions. To be included in the data collection, participants had to answer all the items noted above, including consent, measure of mutability, guilt, blame, and demographic questions. The construction of the written scenario and measure items were based on similar experiments measuring the temporality effect (Miller & Gunasegaram, 1990; Byrne et al., 2000). For the analysis we chose to conduct *z*-tests to test our predictions of hypotheses H1, H2 and H3 across both scenarios, and a two-proportion Chi-square test to test for the differences in responses across Scenario 1 and Scenario 2 for the exploratory analysis. For further information regarding procedure, measure items, and hypotheses, please refer to supplementary materials (Appendix C).

Results

Table 3

Percentages and count measures of mutations, guilt and blame across two scenarios in study 3

	Scenario 1		Scenario 2	
Dependent variables	In %	Participants	In %	Participants
Mutation				
First	59	96	10	17
Second	41	67	90	147
Guilt				
First	51	83	12	20
Second	49	80	88	144
Blame				
First	50	82	10	16
Second	50	81	90	148
n	163		164	

Table 4

Summary of results and effect sizes from scenario 1 and scenario 2

Dependent variables	Scenario 1		Scenario 2	
	One sample z-test	Cohen`s *g* with 95% CI	One sample z-test	Cohen`s *g* with 95% CI
DV1: Which event is most often mutated?	n = 163, k = 67, z = 2.27, p = .023	-0.09 [-0.16, -0.01]	n = 164, k = 147, z = 10.15, p < .001	0.4 [0.35, 0.44]
DV2: Who will experience more guilt?	n = 163, k = 80, z = 0.23, p = .814	-0.01 [-0.09, -0.07]	n = 164, k = 144, z = 9.68, p < .001	0.38 [0.33, 0.43]
DV3: Who will blame the other more?	n = 163, k = 81, z = 0.08, p = .938	0 [-0.08, 0.07]	n = 164, k = 148, z = 10.31, p < .001	0.4 [0.36 , 0.45]
	Two proportions Chi-square test		Cohen`s *h* with 95% CI	
DV1: Which event is most often mutated?	$n1$ =163, $k1$ = 67, $n2$ = 164, $k2$ = 147, X^2 (1, 327) = 85.13, p < .001		-1.09 [-1.31, -0.88]	
DV2: Who will experience more guilt?	$n1$ =163, $k1$ = 80, $n2$ = 164, $k2$ =144, X^2 (1, 327) = 56.82, p < .001		-0.88 [-1.09, -0.66]	
DV3: Who will blame the other more?	$n1$ =163, $k1$ = 81, $n2$ = 164, $k2$ =148, X^2 (1, 327) = 64.05, p < .001		-0.94 [-1.16, -0.72]	

For Scenario 1, the results of a one sample z-test indicated a weak tendency to mutate the first event compared to the second event in a sequence of two independent events (41% vs. 59%), n = 163, k = 67, z = 2.27, p = .023, Cohen`s $g$ = -0.09 [-0.16 , -0.01]. Scenario 2 indicated a tendency to mutate the second event rather than the first (90% vs. 10%), $n$ = 164, $k$ = 147, $z$ = 10.15, $p$ < .001, Cohen`s g = 0.4 [0.35 , 0.44]. A two proportions Chi-square test indicated a reliable interaction ($n1$ =163, $k1$ = 67, $n2$ = 164, $k2$ = 147, X^2 (1, 327) = 85.13, p < .001, Cohen`s h = -1.09 [-1.31, -0.88]).

For attribution of who would experience more guilt, the effect was eliminated in Scenario 1 (49% vs. 51%), n = 163, k = 80, z = 0.23, p = .814, Cohen`s g = -0.01 [-0.09 , -0.07]. In Scenario 2, people showed a tendency to attribute that the second participant would

experience more guilt (88% vs. 12%), n = 164, k = 144, z = 9.68, p < .001, Cohen`s $g$ = 0.38 [0.33 , 0.43]. A two proportions Chi-square test indicated a reliable interaction (*n1* =163, *k1* = 80, *n2* = 164, *k2* =144, $X^2$ (1, 327) = 56.82, $p$ < .001, Cohen`s h = -0.88 [-1.09, -0.66]).

Attribution of who would blame the other more showed the same tendencies, where the effect was eliminated in Scenario 1 (50% vs. 50%), n = 163, k = 81, z = 0.08, p = .938, Cohen`s $g$ = 0 [-0.08 , 0.07]. While in Scenario 2, people attributed that the second participant would blame the first more (90% vs. 10%), $n$ = 164, $k$ = 148, $z$ = 10.31, $p$ < .001, Cohen`s g = 0.4 [0.36 , 0.45]. A two proportions Chi-square test between the two conditions, showed that the interaction was reliable (*n1* =163, *k1* = 81, *n2* = 164, *k2* =148, X^2 (1, 327) = 64.05, p < .001, Cohen`s h = -0.94 [-1.16, -0.72]).

Replication findings

Table 5 provides a summarized evaluation of the replication findings compared to the original studies, evaluated using the criteria set by LeBel et al. (2019). Our replication of Miller and Gunasegaram (1990) found support for the confirmatory predictions, as we found a signal for all three dependent variables (i.e., mutation, guilt, and blame), where the effect sizes were inconsistent in the same direction for mutation (smaller) and guilt (larger), and consistent for blame. The replication of Segura et al. (2002) also found support for the confirmatory predictions, where we reported a signal on the dependent variable mutation in all three scenarios which were inconsistent in the same direction (larger) for mutation in Scenario 1 and Scenario 2, and consistent in Scenario 3. Descriptions of the labels used to interpret the replication findings (i.e., signal, consistent/inconsistent, larger/smaller) can be found in supplementary materials (Table S4 and Figure S5).

Table 5

Comparison and interpretation of effect sizes between original study and replication findings

	Original results	Replication	
Comparison	Cohen`s *g* with 95% CI	Cohen`s *g* with 95% CI	Interpretation
Study 1 – Replication of Miller & Gunasegaram, 1990			
Mutation	0.39 [0.32, 0.45]	0.26 [0.20, 0.32]	Signal - inconsistent, smaller
Guilt	0.36 [0.29, 0.44]	0.43 [0.40, 0.47]	Signal - inconsistent, larger
Blame	0.42 [0.36, 0.48]	0.42 [0.38, 0.46]	Signal - consistent
Study 2 – Replication of Segura et al., 2002			
Scenario 1			
Fourth rather than first	0.12 [-0.01, 0.26]	0.21 [0.15, 0.27]	Signal - inconsistent, larger
Scenario 2			
Fourth rather than first	0.23 [0.06, 0.40]	0.30 [0.25, 0.36]	Signal - inconsistent, larger
Scenario 3			
Fourth rather than first	0.32 [0.14, 0.49]	0.36 [0.32, 0.41]	Signal - consistent

Note. Effect sizes were not reported for either of the two original studies, Cohen's *g* was calculated by reanalyzing the data using information provided in the articles; Interpretations are based on evaluation criteria by LeBel et al. (2019), see supplementary materials (Table S4 and Figure S5) for details.

GENERAL DISCUSSION

We conducted two direct, close replications of Miller and Gunasegaram (1990), and Segura et al. (2002), and successfully replicated both studies. Thus, participants showed a tendency to mutate the last event rather than the first in sequences of both two and four independent events. The replication of Miller and Gunasegaram (1990) also found that people attribute that the second player will experience more guilt and be blamed more by the first player for their failure to win. Even though we successfully replicated both studies, the results do not give much insight as to which theoretical mechanisms and cognitive processes could explain the effects. The original study by Segura et al. (2002) did find results that supported the assumption of an immutability of the first event, and even though we successfully replicated the study, our replication was too limited to rule out other explanations to the effect, as we only measured the fourth event against the first, and not the second and third. The results of the

extension found that people mutated the second event more often in scenario 2, where the first person picked a blue marble, and this effect was eliminated in scenario 1, where the first person picked a red marble. The findings also indicated a reliable interaction between the two scenarios. The following discussion will focus on elucidating possible explanations to the findings in the extension, and

The body of research investigating the temporality effect has mainly focused on the immutability of the first event as an explanation to the effect (Miller & Gunasegaram, 1990; Byrne et al., 2000; Walsh & Byrne, 2004). Based on this assumption, the first event of picking a red marble in scenario 1 of the extension, should make both picking a red marble the more available winning possibility, and thereby make the second player also picking a red marble the more available counterfactual. However, we found that when the first event negatively affects the statistical probability of a winning outcome, the effect is eliminated. One explanation to the elimination of the effect could be that the mental representations of winning possibilities are influenced by how probable the winning possibilities are in the first place. The chance of both players picking a blue marble is 70% for both the first and second event, while in the winning condition of both players picking a red marble, the chances are 30% for both events. Thus, both players picking a blue marble might be more available as a winning condition. When attempting to match the first event of picking a red marble to the winning possibilities of both picking blue doesn't work, the first event is changed to match the winning condition. (Walsh & Byrne, 2004).

If the probability of a winning outcome influences the availability of winning conditions people mentally represent, there is still the question of why the effect was eliminated and not reversed. Byrne et al. (2000) suggested that the elimination, contrary to a reversed effect, might be an indication that the temporality effect continue to operate despite conflicting cognitive mechanisms that make for more available alternatives. They argue that some of the participants are guided by an alternative counterfactual to the first event, in this case the more available

winning condition of both picking blue marbles, while others exhibit the standard temporality effect, by focusing on how to match the second event to the winning condition of both picking a red marble.

Another possible explanation could be that people view the events in the sequence to be in a causal order because the first person picking the red marble negatively affects the chance of the second person's ability to follow up with the same-colored marble. It has been suggested that in some temporal sequences people may assume a causal link between events even though a causal relationship between the events is not implied in the description (Spellman, 1997). Based on that people don't seem to make the same attributions of a causal link between the events in scenario 2, as they showed the standard temporality effect, it could be that this attribution is more prominent when the outcome of the first event negatively affects the chances of a good outcome for the second event. This proposition resembles the assumptions presented in the crediting causality hypothesis, where the causal strength of events varies according to an estimation of change in subjective probability of a good outcome between the events (Spellman, 1997). However, this hypothesis does not base the assumption on people attributing a causal link between the events. If people view the sequence as causal, the causal primacy effect (Wells et al., 1987) would predict a reversed effect, meaning that the first event is considered more mutable, and as previously mentioned, our results only showed an elimination of the effect. This could, as hypothesized earlier in the discussion, imply that the temporality effect is present, and that some people exhibit the temporality effect, while others are guided by alternative counterfactuals of the first event as a result of perceiving the sequence as causal (Byrne et al., 2000).

Regardless of what assumptions discussed explain the effects observed in scenario 1 of the extension, there seems to be support of the temporality effect operating to some extent despite possible contradictory mechanisms (Byrne et al., 2000). This is supported by the

findings in our replications of Miller and Gunasegaram (1990), and Segura et al. (2002), where people showed a tendency to mutate the more recent event in sequences of two and four independent events. The body of research investigating the temporality effect is consistent with the findings of the present study, where the tendency to mutate the more recent event assumes that the effect occurs because of a presupposition of the first event in sequences of independent events. The research also accounts for how the mutability of an event might differ depending on how accessible alternatives to the event is (Byrne et al., 2000; Segura et al., 2002; Walsh & Byrne, 2004), and that this is part of a cognitive process which help navigate the critical and efficient information in order to make calculated judgements of which events to change to match a desired outcome (Walsh & Byrne, 2004). The crediting causality hypothesis attempt to propose an alternative explanation to the effect and suggest that people mutate the second event based on an estimation of change in subjective probability between the events (Spellman, 1997). However, several studies have demonstrated how the hypothesis is not sufficient in capturing significant elements of the effect. For instance, findings where the true possibilities are held constant but where the winning conditions and features of the scenario are manipulated, the predictions of the crediting causality hypothesis remain the same, however, the results show that the effect changes according to the conditions (Byrne et al., 2000; Walsh & Byrne, 2004).

Limitations and further research

The sampling methods used for the current study consisted of convenient sampling through social contacts of the research group for study 1, and Prolific platform for Study 2 and Study 3. Convenient sampling displays some obvious limitations as it may lead to a less representative sample, which could affect generalizability. However, the experiments in the current study measures cognitive processes that should be prevalent despite demographic deviations, so sampling method does not present any crucial concerns. Data quality is a concern

that has been raised in relation to the growing usage of online platforms for data collection (Peer, Rothschild, Gordon, Evernden, & Damer, 2022). Platforms for online research (i.e., Prolific platform) raises an issue regarding key aspects of data quality, such as participants attention, comprehension, honesty, and reliability (Peer et al., 2022). Our study consisted of fairly simple surveys and did not measure aspects that are particularly vulnerable to bias or dishonesty, so the quality of our data should not be particularly affected by the limitations that is associated with digital platforms for data collection.

We did not replicate the full experiment of Study 1 by Segura et al. (2002) and were therefore limited in both assessing the replicability of the total experiment, and exploring the implications the remaining measures could have to the theory as an entirety. A replication should be conducted where the fourth event is also measured against the second and third event.

A further investigation of scenarios where the statistical probabilities of the winning outcomes differ should also be considered, as this could provide valuable insight to how the mutation of events differ based on not only the accessibility of counterfactuals to the specific events, but also to the probabilities of the winning outcomes.

CONCLUSION

The current study successfully replicated two studies measuring the temporality effect in sequences of two and four independent events. When imagining alternatives to an event, people tend to mutate the more recent event in sequences of independent events. This tendency is based on the assumption that the first event is presupposed and considered immutable. When more accessible alternatives are available to the first event, the effect can be reversed or reduced. The extension found that when the statistical probabilities of possible counterfactual outcomes are manipulated, the effect is eliminated when the first event negatively affect the

chances of obtaining a more probable winning condition. Further research should investigate how differences in statistical probabilities of winning outcomes affect the mutability of events.

REFERENCES

Byrne, R. M. J. (2016). Counterfactual Thought. *Annual Review of Psychology, 67*(1), 135-157. https://doi.org/10.1146/annurev-psych-122414-033249

Byrne, R. M. J., Segura, S., Culhane, R., Tasso, A., & Berrocal, P. (2000). The temporality effect in counterfactual thinking about what might have been. *Memory & Cognition*, *28*(2), 264–281. https://doi.org/10.3758/BF03213805

Earp, B. D., & Trafimow, D. (2015). Replication, falsification, and the crisis of confidence in social psychology. *Frontiers in Psychology, 6*(621). https://doi.org/10.3389/fpsyg.2015.00621

Henne, P., Kulesza, A., Perez, K., & Houcek, A. (2021). Counterfactual thinking and recency effects in causal judgement. *Cognition, 212*(2021), p. 104708. https://doi.org/10.1016/j.cognition.2021.104708

Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLoS Medicine, 2*(8), e124. https://doi.org/10.1371/journal.pmed.0020124

Kahneman, D., & Miller, D. T. (1986). Norm Theory: Comparing Reality to Its Alternatives. *Psychological Review, 93*(2), 136-153. https://doi.org/10.1037/0033-295x.93.2.136

LeBel, E. P., McCarthy, R. J., Earp, B. D., Elson, M., & Vanpaemel, W. (2018). Advances in Methods and Practices in Psychological Science. *Psychological Science, 1*(3), 389-402. https://doi.org/10.1177/2515245918787489

LeBel, E. P., Vanpaemel, W., Cheung, I., Campbell, L. (2019). A Brief guide to Evaluate Replications. *Meta-Psychology, 3*(2019), 1-9. https://doi.org/10.15626/MP.2018.843

Miller, D. T., & Gunasegaram, S. (1990). Temporal order and the perceived mutability of events: Implications for blame assignment. *Journal of Personality and Social Psychology*, *59*, 1111–1118. https://doi.org/10.1037/0022-3514.59.6.1111

NESH. (2022, May 26th). *Guidelines for Research Ethics in the Social Sciences and the Humanities*. National Research Ethics Committees. https://www.forskningsetikk.no/en/guidelines/social-sciences-humanities-law-and-theology/guidelines-for-research-ethics-in-the-social-sciences-humanities-law-and-theology/

NTNU. (s.a.) *Collection of personal data for research projects.* Norwegian University of Science and Technology. Retrieved March 28th 2023, from https://i.ntnu.no/wiki/-/wiki/English/Collection+of+personal+data+for+research+projects

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), 943–943. https://doi.org/10.1126/science.aac4716

Peer, E., Rothschild, D., Gordon, A., Evernden, Z., & Damer, E. (2022). Data quality of platforms and panels for online behavioral research. *Behavioral Research Methods, 54*(4), 1643-1662. https://doi.org/10.3758/s13428-021-01694-3

Segura, S., & McCloy, R. (2003). Counterfactual thinking in everyday life situations: Temporal order effects and social norms. *Psicológica, 24*(1), 1-15.

Segura, S., Fernandez-Berrocal, P., & Byrne, R. M. J. (2002). Temporal and causal order effects in thinking about what might have been. *The Quarterly Journal of Experimental Psychology Section A, 55*(4), 1295–1305. https://doi.org/10.1080/02724980244000125

Spellman, B. A. (1997). Crediting causality. *Journal of Experimental Psychology: General, 126*, 323-348. https://doi.org/10.1037/0096-3445.126.4.323

Walsh, C. R., & Byrne R. M. J. (2004). Counterfactual Thinking: The temporal order effect. *Memory & Cognition, 32*(3), 369-378. https://doi.org/10.3758/BF03195831

Wells, G. L., Taylor, B. R., & Turtle, J. W. (1987). The undoing of scenarios. *Journal of Personality and Social Psychology, 53*(3), 421-430. https://doi.org/10.1037/0022-3514.53.3.421

Supplementary materials

The supplementary materials present additional information and details of the replications and extension, including pre-registrations, assessment of processing of personal data, and evaluation criteria for classification of replications and interpretations of effect sizes.

Table of Contents

Appendix A: Pre-registration Study 1 – Replication of Miller and Gunasegaram (1990)

Project working title: **Replication of Miller & Gunasegaram`s (1990) coin toss scenario**

Authors: ***left out for blind review of the preregistration***
Affiliation: ***left out for blind review of the preregistration***

Summary

This project's aim is to test the predominant tendency people have to consider the second event in an independent two-event sequence more mutable than the first, and to test whether linguistic preferences influence this tendency in a Norwegian sample.

Hypotheses

Common predictions across Scenario 1 and Scenario 2
H1: The most recent event in a sequence of two independent events is considered more mutable than the first.

H2: Given a negative outcome based on a sequence of two independent chance events, participants will judge the second player to experience more guilt than the first player.

H3: Given a negative outcome based on a sequence of two independent chance events, the study participants will judge that the first player will blame the second player more often than the second player blaming the first player.

Contrasting Scenario 1 and Scenario 2
We do not have concrete directional predictions contrasting the responses across Scenario 1 and Scenario 2. We will report the results of the contrast in responses to Scenario 1 and Scenario 2 across each of the three predictions noted above (i.e., H1, H2, & H3).

Exploratory predictions:
The survey will be conducted in the Norwegian language. We will test if the Norwegian participants report "heads-tails" linguistic preference over "tails-heads," and if such a preference contributes to the pattern of responses across Scenario 1 and Scenario 2.

Study Materials

Study Outline:

In this survey, you will read two separate scenarios about outcomes of chance events (such as a coin toss). As you read the descriptions, please carefully try to form a detailed understanding of the situations related to the decision-makers involved. Following each scenario, you will answer three short questions about the scenario based on your understanding.

Scenario 1:

Imagine two individuals (Jonas and Kristian) who are offered the following very attractive proposition. Each individual is asked to toss a coin. If the two coins come up the same (both heads or both tails), each individual wins NOK 10,000. However, if the two coins do not come up the same, neither individual wins anything. Jonas goes first and tosses a head; Kristian goes next and tosses a tail. Thus, the outcome is that neither individual wins anything.

There were two ways that Jonas & Kristian could have won NOK 10,000. Which of these alternatives comes more readily to mind?

- Jonas tossing a tail
- Kristian tossing a head

Who would you predict will experience more guilt—Jonas or Kristian?

- Jonas
- Kristian

Will Jonas blame Kristian more or will Kristian blame Jonas more for their failure to win NOK 10,000?

- Jonas
- Kristian

Scenario 2:

Imagine two individuals (Oscar and Chris) who are offered the following very attractive proposition. Each individual is asked to toss a coin. If the two coins come up the same (both heads or both tails), each individual wins NOK 10,000. However, if the two coins do not come up the same, neither individual win anything. Oscar goes first and tosses a tail; Chris goes next and tosses a head. Thus, the outcome is that neither individual win anything.

There were two ways that Oscar & Chris could have won NOK 10,000. Which of these alternatives comes more readily to mind?

- Oscar tossing a head
- Chris tossing a tail

Who would you predict will experience more guilt—Oscar or Chris?

- Oscar
- Chris

Will Oscar blame Chris more or will Chris blame Oscar more for their failure to win NOK 10,000?

- Oscar will blame Chris more
- Chris blame Oscar more

Which of these phrasings appear the most natural to you?

- Heads or tails
- Tails or heads

Demographic questions

Thank you, you completed the main part of the survey.

A couple of quick final questions.

How old are you?

---- (text box here)

What is your gender?

- Female
- Male
- Other

Planned sample

Participants recruited will be Norwegian-speaking adults. The sample size was calculated using G*power 3.1.9.7, and was based on 90% power (and $\alpha = .05$), with the aim of detecting an effect size of Cohen`s $g = 0.2$. We aim to achieve a sample size of 200 particpants. The details from the power analysis can be found in the power analysis section below.

Suggested Analysis

The original authors did not include the results of the analysis, nor did they report which one proportion test they used, however, we found it appropriate to conduct one sample Z tests to test the predictions.

Detailed results of coin toss scenario by Miller & Gunasegaram (1990)

Table 1

Percentages and frequency count measures on predictions of experience of guilt, judgment of blame, and undoings of the first or second sequence through an "if only…" question.

Question	In %	Conversion	After rounding
Who will experience more guilt?			
First	14%	12,32	12
Second	86%	75,68	76
Who will blame the other more?			
First	92%	80,96	81
Second	8%	7,04	7
Which event is most often mutated?			
First event	11%	9,68	10
Second event	89%	78,32	78
n		88	

Note. Frequency calculations were based on percentage values reported in the original study (Miller & Gunasegaram, 1990, pp. 1111-1112).

Table 2

Results of Study 1 of Miller & Gunasegaram (1990)

Hypotheses	Dependent variables	Statistical test	Effect size with 95% CI
		One sample Z-test	**Cohen`s *g***
H1	**DV1:** Which event is most often mutated?	$k = 78$, $n = 88$, $z = 7.25$, $p < .001$	0.39 [0.32 , 0.45]
H2	**DV2:** Who will experience more guilt?	$k = 76$, $n = 88$, $z = 6.82$, $p < .001$	**0.36 [0.29 , 0.44]**
H3	**DV3**: Who will blame the other more?	$k = 81$, $n = 88$, $z = 7.89$, $p < .001$	0.42 [0.36 , 0.48]

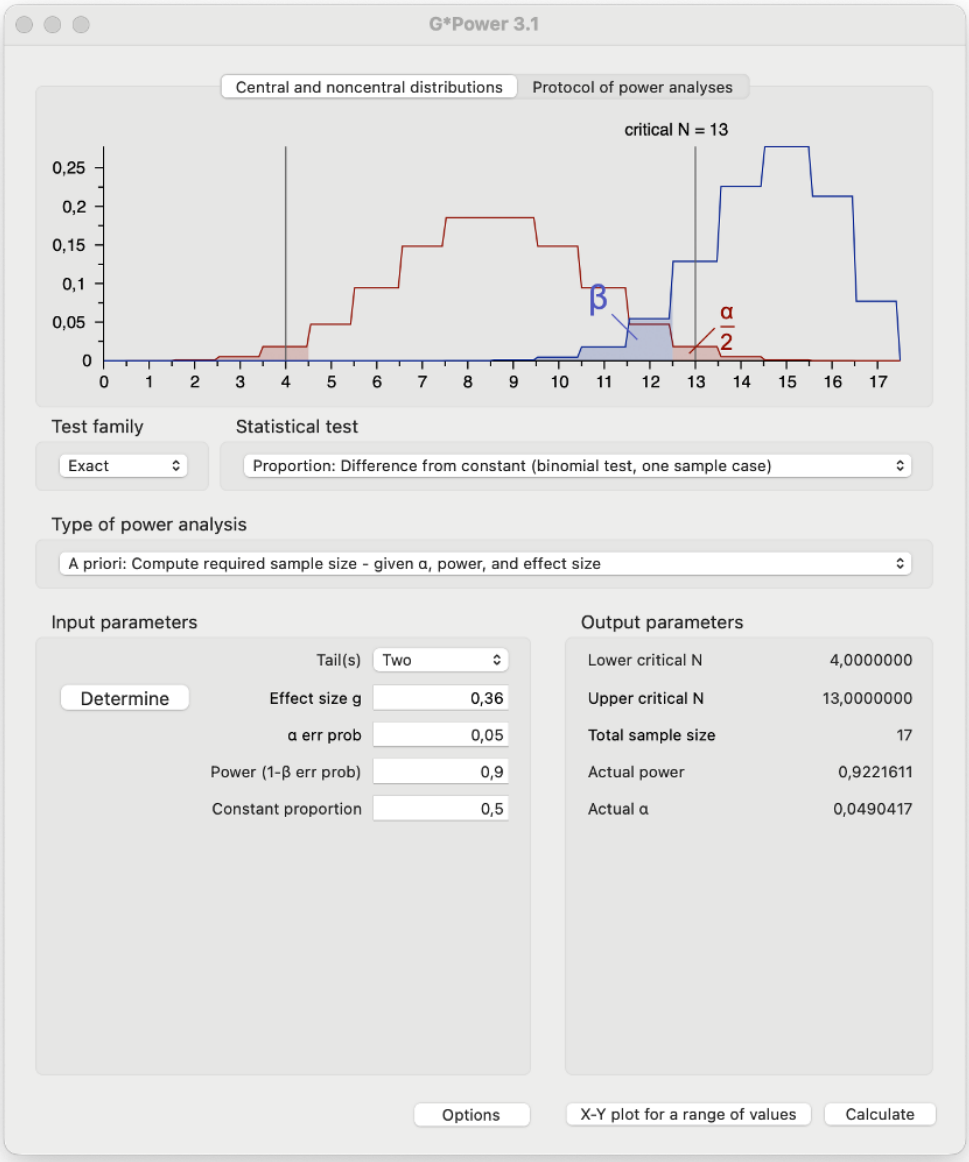
Power analysis

The rationale for reconstructing the original dataset and re-running analysis: authors of the original studies did not report the full statistical results necessary (i.e., effect size measures were missing) to run a power analysis. Hence, we had to re-conduct the analysis reported in the original study based on information available in the description of the study (see Table 1 above).

Steps for power analysis

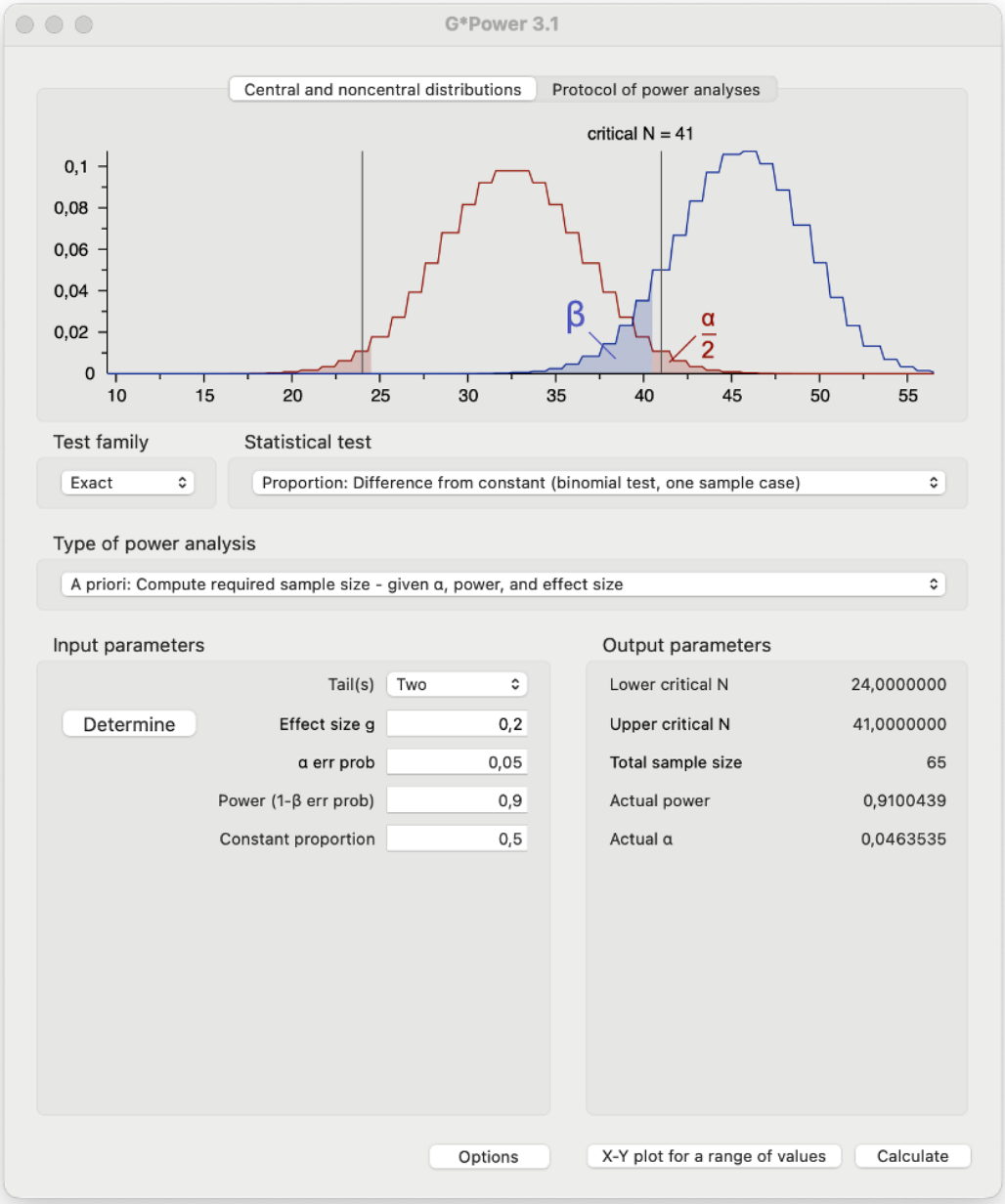
Hypotheses H1, H2 and H3

The testing of H1, H2 and H3 involved three One sample Z test. We conducted a power analysis based on the smallest effect size among these results (i.e., Cohen's g = 0.36). The results of the analysis suggest a total sample size of 17 (see screenshot below).



Smallest effect size of interest

We aim to be able to detect a small effect size of Cohen`s g = 0.2 at 0.90 power (alpha =.05). The result of the power analysis suggests a total sample size of 65 (see screenshot below).



Summary of power analysis:

The analysis suggests a replication sample size of 65 participants. However, we aim to recruit a minimum of 200 participants, because a larger sample size will be more accurate and reliable. The original study also had very large effect sizes, so to include a larger sample will be beneficial to detect smaller effect sizes.

Study materials in Norwegian

Studieoversikt:
I denne spørreundersøkelsen vil du lese to separate scenarioer om utfall i et sjansespill som myntkast. Selv om de to scenarioene kan virke like, så er de forskjellige på viktige punkt. Når du leser beskrivelsene, vennligst forsøk å forme en detaljert forståelse av situasjonene relatert til beslutningstakerne som er involvert. Etter å ha lest beskrivelsen, vil du besvare tre korte spørsmål om scenarioet basert på din forståelse av det.

Scenario 1:
Se for deg to individer (Jonas og Kristian) som blir tilbudt den følgende veldig attraktive muligheten. Hver av individene blir bedt om å kaste en mynt. Om begge myntene ender opp med det samme utfallet (begge kaster kron eller begge kaster mynt), vil hver av individene vinne 10.000kr. Dersom de to myntene ikke ender opp med det samme utfallet, vil ingen av dem vinne noe. Jonas starter og kaster kron; Kristian går neste og kaster en mynt. Altså, utfallet er at ingen av individene vinner noe.

Det var to muligheter for at Jonas og Kristian kunne vunnet de 10 000 kronene. Hvilken av disse alternativene er lettest å forestille seg?

- Jonas kastet mynt
- Kristian kastet kron

Hvem ville du forutsett at opplevde mer skyld – Jonas eller Kristian?

- Jonas
- Kristian

Vil Jonas skylde mer på Kristian, eller vil Kristian skylde mer på Jonas for at de ikke klarte å vinne 10 000kr hver?

- Jonas skylder mer på Kristian
- Kristian skylder mer på Jonas

Scenario 2:
Se for deg to individer (Oscar og Chris) som blir tilbudt den følgende veldig attraktive muligheten. Hver av individene blir bedt om å kaste en mynt. Om begge myntene ender opp med det samme utfallet (begge kaster kron eller begge kaster mynt), vil hver av individene vinne 10 000kr. Dersom de to myntene ikke ender opp med det samme utfallet, vil ingen av dem vinne noe. Oscar starter og kaster mynt; Chris er neste og kaster kron. Altså, utfallet er at ingen av individene vinner noe.

Det var to muligheter for at Oscar og Chris kunne vunnet de 10 000 kronene. Hvilken av disse alternativene er lettest å forestille seg?

- Oscar kastet kron
- Chris kastet mynt

Hvem ville du forutsett at opplevde mer skyld – Oscar eller Chris?

● Oscar
● Chris

Vil Oscar skylde mer på Chris, eller vil Chris skylde mer på Oscar for at de ikke klarte å vinne 10 000kr hver?

● Oscar skylder mer på Chris
● Chris skylder mer på Oscar

Tusen takk! Du har nå gjennomført hoveddelen av spørreundersøkelsen.

Noen få siste spørsmål.

Hvilken av disse formuleringene faller deg mest naturlig?

● "Kron eller mynt"
● "Mynt eller kron"

Hvor gammel er du?
---(text box here)

Hvilket kjønn identifiserer du deg som?

● Kvinne
● Mann
● Annet

Studie debrifing

Vi ønsker å takke deg for din deltakelse i vårt forskningsprosjekt, og håper du synes det var interessant.

De eksperimentene du har deltatt på i dag ble designet for å undersøke hvordan mennesker vurderer utfall basert på rekkefølgen som hendelsene presenteres i.

Vi ønsker at du ikke deler detaljer angående studien til noen, da de kan være potensielle deltakere, og å vite meningen med forskningsprosjektet på forhånd, kan påvirke resultatene. Tusen takk for din deltakelse. Dersom du ønsker informasjon om resultatene, eller har videre spørsmål for oss, vennligst kontakt Subramanya Prasad Chandrashekar på prasad.chandrashekar@ntnu.no når som helst.

Appendix B: Pre-registration Study 2 – Replication of Segura et al. (2002)

Project working title: **Replication of study 1 of Segura et al. (2002)**

Authors: ***left out for blind review of the preregistration***
Affiliation: ***left out for blind review of the preregistration***

Summary
This project's aim is to test if the predominant tendency people have to consider the second event in an independent two-event sequence more mutable than the first, also occurs for sequences of four independent events.

Hypotheses
H1: The last event in a sequence of four independent events is considered more mutable than the first.

Study materials

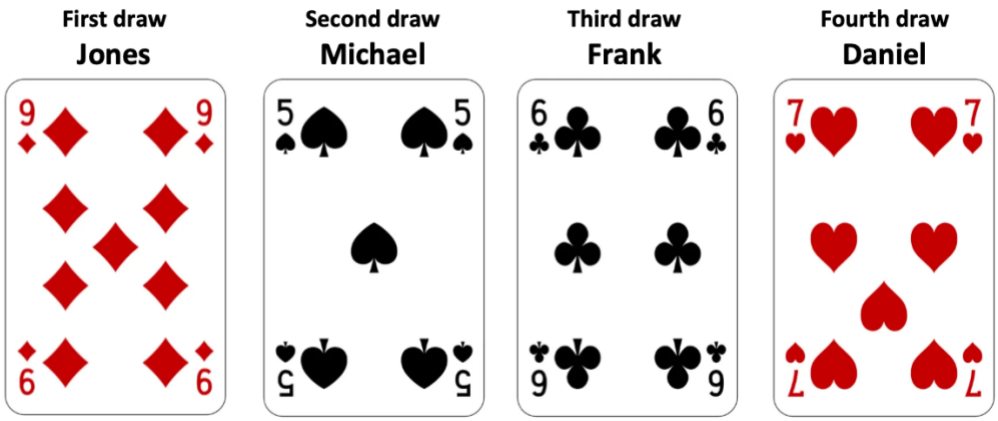
Study outline

In this survey, you will read about a situation where four people are making decisions. As you read the description, please carefully try to form a detailed impression of the situation and of the decision-makers involved. Following the scenario, you will answer one question about the scenario based on your understanding.

Scenario 1 (Red-Black-Black-Red)

Imagine four individuals (Jones, Michael, Frank, and Daniel) who are offered the following very attractive proposition. Each individual is given a shuffled deck of cards, and each one picks a card from their own deck. If three of the four cards they pick are of the same color (i.e., three from black suits or three from red suits), each individual win $1,000. However, if three of the cards are not the same color, none of the individuals win anything. Jones goes first and picks a **red** card from his deck. Michael goes second and picks a **black** card. Frank goes third and picks a **black** card. And Daniel picks a **red** card. Thus, the outcome is that neither of the individuals win anything.

The figure below is a visual summary of the outcomes you just read:



Four players could each have won £1,000 each if only one of them had picked a different card, for instance if ..."
..... *(text box)*

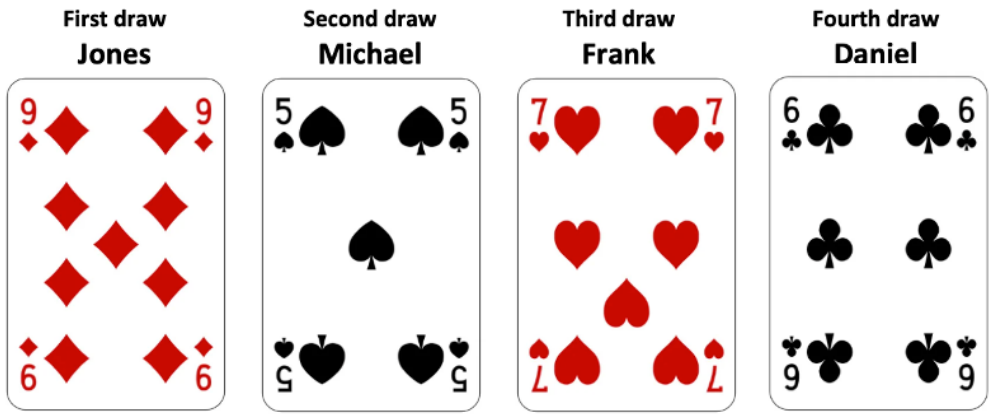
Jones, Michael, Frank, and Daniel could each have won $1,000 if only one of them had picked a different card. That is, you try to imagine an alternative series of outcomes in which the four players could have won $1,000. Which of below two alternatives comes more readily to your mind?

- Jones picking a black card
- Daniel picking a black card

Scenario 2 (Red-Black-Red-Black)

Imagine four individuals (Jones, Michael, Frank, and Daniel) who are offered the following very attractive proposition. Each individual is given a shuffled deck of cards, and each one picks a card from their own deck. If three of the four cards they pick are of the same color (i.e., three from black suits or three from red suits), each individual win $1,000. However, if three of the cards are not the same color, none of the individuals win anything. Jones goes first and picks a **red** card from his deck. Michael goes second and picks a **black** card. Frank goes third and picks a **red** card. And Daniel picks a **black** card. Thus, the outcome is that neither of the individuals win anything.

The figure below is a visual summary of the outcomes you just read:



Four players could each have won £1,000 each if only one of them had picked a different card, for instance if ..."
..... *(text box)*

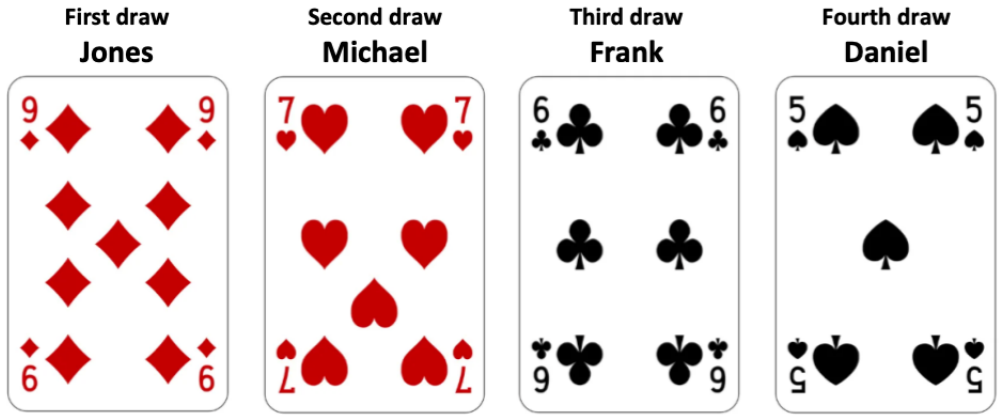
Jones, Michael, Frank, and Daniel could each have won $1,000 if only one of them had picked a different card. That is, you try to imagine an alternative series of outcomes in which the four players could have won $1,000. Which of below two alternatives comes more readily to your mind?

- Jones picking a black card
- Daniel picking a red card

Scenario 3 (Red-Red-Black-Black)

Imagine four individuals (Jones, Michael, Frank, and Daniel) who are offered the following very attractive proposition. Each individual is given a shuffled deck of cards, and each one picks a card from their own deck. If three of the four cards they pick are of the same color (i.e., three from black suits or three from red suits), each individual win $1,000. However, if three of the cards are not the same color, none of the individuals win anything. Jones goes first and picks a **red** card from his deck. Michael goes second and picks a **red** card. Frank goes third and picks a **black** card. And Daniel picks a **black** card. Thus, the outcome is that neither of the individuals win anything.

The figure below is a visual summary of the outcomes you just read:



Four players could each have won £1,000 each if only one of them had picked a different card, for instance if ..."
..... *(text box)*

Jones, Michael, Frank, and Daniel could each have won $1,000 if only one of them had picked a different card. That is, you try to imagine an alternative series of outcomes in which the four players could have won $1,000. Which of below two alternatives comes more readily to your mind?

- Jones picking a black card
- Daniel picking a red card

Participants` engagement and demographic questions

Thank you, you completed the main part of the survey.

A couple of quick final questions.

How old are you?
---- (text box here)

What is your gender?

- Female
- Male
- Non-Binary
- Other

Are you a U.S. citizen?

- Yes
- No

How serious were you in filling out this questionnaire? (1 = not at all, 5 = very much).
---- (Insert scale from 1-5 here)

Please rate your English language proficiency. (1 = very bad, 7 = very good)
---- (Insert scale from 1-7 here)

Planned sample

Data will be gathered online through the Prolific platform. The sample size was based on 90% power (and $\alpha = .05$) to detect an effect size of Cohen's $g = 0.12$. Therefore, we aim to recruit a total of 552 participants to take part in the study. See the power analysis section below for the details. The participants will be randomly and evenly assigned to each of the three conditions.

Exclusion criteria

To increase the generalizability of the results, we will focus on the total sample size.

1. Participants indicated a low proficiency in English (self-report < 4, on a 1-7 scale).
2. Participants who self-report not being serious about filling in the survey (self-report <4, on a 1-5 scale).
3. Participants who completed the survey too quickly (within one minute).

Suggested analysis

Matching the analysis choice of the original authors, we will conduct one sample z-test to test the predictions of hypothesis H1.

Detailed results of Study 1 by Segura et al. (2002)

Table 1

Percentages and frequency count measures on undoing of the first, second, third or fourth sequence through an "if only…" question.

Condition	*n*	In %	Conversion	After rounding
Two events	79			
First event overall		25	19.75	20
Second event overall		63	49.77	50
Red-Black-Black-Red	85			
First event overall		23	19.55	20
Second event overall		15	12.75	13
Third event overall		3	2.55	3
Fourth event overall		39	33.15	33
Red-Black-Red-Black	68			
First event overall		12	8.16	8
Second event overall		36	24.48	24
Third event overall		12	8.16	8
Fourth event overall		33	22.44	22
Red-Red-Black-Black	68			
First event overall		6	4.08	4
Second event overall		5	3.40	3
Third event overall		44	29.92	30
Fourth event overall		26	17.68	18

Note. Frequency count calculations were based on percentage values reported in the original study (Segura et al., 2002, Table 1, pp. 1299)

Table 2

Results from original study (Segura et al., 2002, Study 1)

Hypothesis	Dependent variables	Statistical test	Effect size with 95% CI
		One sample Z-test	**Cohen`s *g***
	Two events		
	Second rather than first	*k = 50*, *n* = 70, z = 3.59, *p* < .001	0.21 [0.11, 0.32]
	Red-Black-Black-Red		
H1	Fourth rather than first	*k = 33*, *n* = 53, z = 1.79, *p* = .098	**0.12 [-0.01, 0.26]**
	Fourth rather than second	*k = 33*, *n* = 46, z = 2.95, *p* = .004	0.22 [0.08, 0.35]
	Fourth rather than third	*k = 33*, *n* = 36, z = 5.00, *p* < .001	0.42 [0.32, 0.51]
	Red-Black-Red-Black		
H1	Fourth rather than first	*k = 22*, *n* = 30, z = 2.56, *p* = .016	0.23 [0.06, 0.40]
	Fourth rather than second	*k = 22*, *n* = 46, z = 0.29, *p* = .883	-0.02 [-0.17, 0.13]
	Fourth rather than third	*k = 22*, *n* = 30, z = 2.56, *p* = .016	0.23 [0.06, 0.40]
	Red-Red-Black-Black		
H1	Fourth rather than first	*k = 18*, *n* = 22, z = 2.98, *p* = .004	0.32 [0.14, 0.49]
	Fourth rather than second	*k = 18*, *n* = 21, z = 3.27, *p* = .0015	0.36 [0.19, 0.52]
	Fourth rather than third	*k = 18*, *n* = 48, z = 1.73, *p* = .111	-0.12 [-0.27, 0.02]

Note. The replication study will only carry 3 conditions consisting of the four-sequence scenarios; Statistical tests were based on results reported in the original study (Segura et al., 2002, pp. 1298-1300)

Additional points:

Has data collection begun for this project?

As of the date of submission of this research plan for pre-registration, the data have not yet been collected, created, or realized.

What is the estimated start and end date for the project?

The estimated end date of this project is 30.06.2023.

When, where, and what of the research output will be shared?

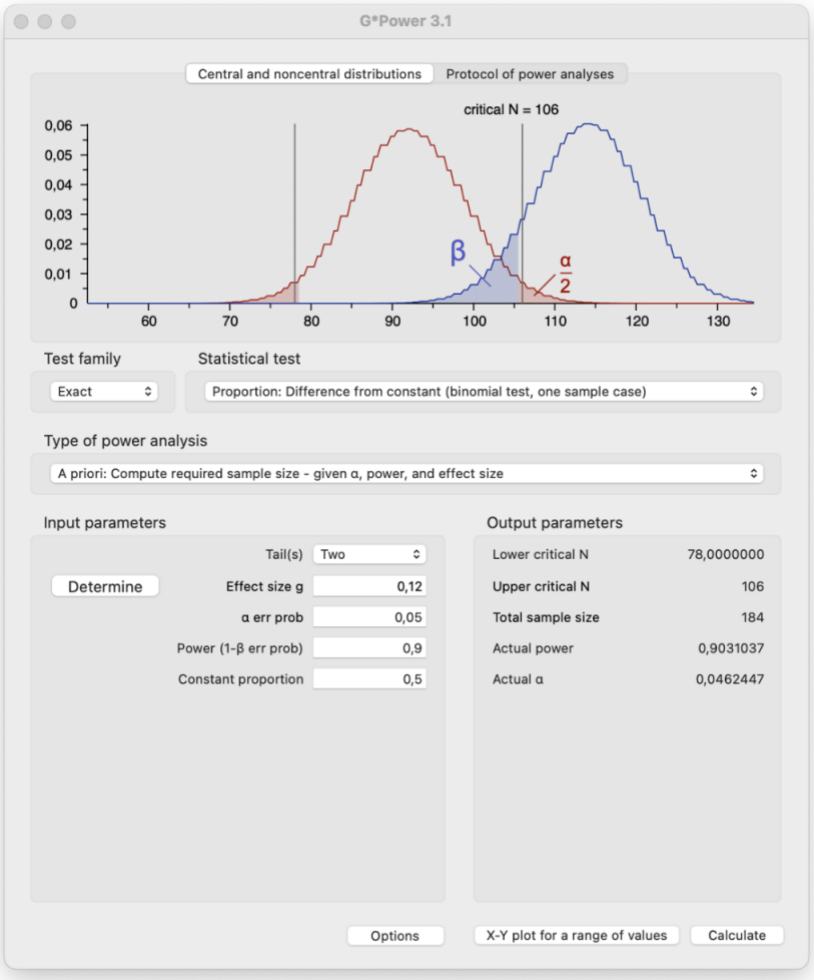
The research output will be shared on Open Science Framework (OSF), conference talks, and in peer-reviewed journals.

Power analysis

The rationale for reconstructing the original dataset and re-running the analysis: the authors of the original study did not report all the necessary results (i.e., the effect size measures were missing) to run a power analysis. Hence, we had to re-conduct the analysis reported in the original study based on information available in the descriptions in the results (see Table 1 above).

Steps for power analysis

The testing of H1 involved three One sample Z tests. We conducted a power analysis based on the smallest effect size among these results (i.e., the smallest effect size of H1 is Cohen's g = 0.12). The results of the analysis suggest a total sample size of (184 x 3) = 552 (see screenshot below).



Summary of power analysis:

The analysis suggests a replication sample size of 552 participants (i.e., 184 participants/condition across 3 conditions), which will be our aim for this study.

Appendix C: Pre-registration Study 3 – Extension

Project working title: **Extension: Statistical possibility and the temporality effect**

Authors: left out for blind review of the preregistration
Affiliation: left out for blind review of the preregistration

Summary

This project's aim is to test the robustness of the predominant tendency people have to consider the second event in an independent two-event sequence more mutable than the first.

Hypotheses

Predictions across Scenario 1 and Scenario 2

HI: The second event in a sequence of two independent events is considered more mutable than the second.

H2: Given a negative outcome based on a sequence of two independent chance events, participants will judge the second player to experience more guilt than the first player.

H3: Given a negative outcome based on a sequence of two independent chance events, the study participants will judge that the first player will blame the second player more often than the second player blaming the first player.

Contrasting Scenario 1 and Scenario 2

We do not have concrete directional predictions contrasting the responses across Scenario 1 and Scenario 2. We will report the results of the contrast in responses to Scenario 1 and Scenario 2 across each of the three predictions noted above (i.e., HI , H2, & H3).

Exploratory predictions:

We will test for the differences in the pattern of responses across Scenario 1 and Scenario 2 for each of the three DVs (i.e., mutability, guilt, blame).

Study materials

Study outline

In this survey, you will read about a situation where two people are making decisions. As you read the description, please carefully try to form a detailed impression of the situation and of the decision-makers involved. Following the scenario, you will answer three questions about the scenario based on your understanding.

Scenario 1

Imagine two individuals (Linda and Barbara) who are offered the following very attractive proposition. Each individual is handed a sack of marbles that contains 30 red marbles, and 70 blue marbles. The players get to pick one marble each without looking. If both players pick the same-colored marble (i.e., both red marbles, or both blue marbles) they both win $1,000. However, if the marbles are not the same color, neither of the players win anything. Linda goes first and picks a red marble. Barbara goes next and picks a blue marble. Thus, the outcome is that neither of the individuals win anything.

Two players could each have won $1,000 each if only one of them had picked a different colored marble, for instance if..."
……(text box)

There were two ways that Linda & Barbara could have won $1,000. That is, you try to imagine an alternative series of outcomes in which the two players could have won $1,000. Which of below two alternatives comes more readily to your mind?

- Linda had picked a blue marble
- Barbara had picked a red marble

Who would you predict will experience more guilt—Linda or Barbara?

- Linda
- Barbara

Will Linda blame Barbara more or will Barbara blame Linda more for their failure to win $1,000?

- Linda blames Barbara more
- Barbara blames Linda more

Scenario 2

Imagine two individuals (Linda and Barbara) who are offered the following very attractive proposition. Each individual is handed a sack of marbles that contains 30 red marbles, and 70 blue marbles. The players get to pick one marble each without looking. If both players pick the same-colored marble (i.e., both red marbles, or both blue marbles) they both win $1,000. However, if the marbles are not the same color, neither of the players win anything. Linda goes first and picks a blue marble. Barbara goes next and picks a red marble. Thus, the outcome is that neither of the individuals win anything.

"Two players could each have won $1,000 each if only one of them had picked a different colored marble, for instance if…"
……(text box)

There were two ways that Linda & Barbara could have won $ 1000. That is, you try to imagine an alternative series of outcomes in which the two players could have won $1 ,000. Which of below two alternatives comes more readily to your mind?

- Linda had picked a red marble
- Barbara had picked a blue marble

Who would you predict will experience more guilt—Linda or Barbara?

- Linda
- Barbara

Will Linda blame Barbara more or will Barbara blame Linda more for their failure to win $1,000?

- Linda blames Barbara more
- Barbara blames Linda more

Participants` engagement and demographic questions

Thank you, you completed the main part of the survey.

A couple of quick final questions.

How old are you?
----(text box here)

What is your gender?

- Female
- Male
- Non-binary
- Prefer not to say

How serious were you in filling out this questionnaire? (1 = not at all, 5 = very much).

----(Insert scale from 1-5 here)

Please rate your English language proficiency? (1 = very bad, 7 = very good)

----(Insert scale from 1-7 here)

Planned sample

Data will be gathered online through the Prolific platform. The sample size was based on 90% power (and a = .05) to detect an effect size of Cohen's g = 0.14. Therefore, we aim to recruit a total of 300 participants to take part in the study. See the power analysis section below for the details. The participants will be randomly and evenly assigned to each of the two conditions.

Exclusion criteria

To increase the generalizability of the results, we will focus on the total sample size.

1. Participants indicated a low proficiency in English (self-report < 3, on a 1-7 scale).
2. Participants who self-report not being serious about filling in the survey (self-report <4, on a 1-5 scale).
3. Participants who completed the survey too quickly (within one minute).

Suggested analysis

We will conduct one sample z-test to test the predictions of hypotheses HI , H2 , and H3. To test the exploratory prediction, we will conduct a two-proportions Chi-square test.

Additional points:

Has data collection begun for this project?

As of the date of submission of this research plan for pre-registration, the data have not yet been collected, created, or realized.

What is the estimated start and end date for the project?

The estimated end date of the project will be 30.06.2023.

When, where, and what of the research output will be shared?

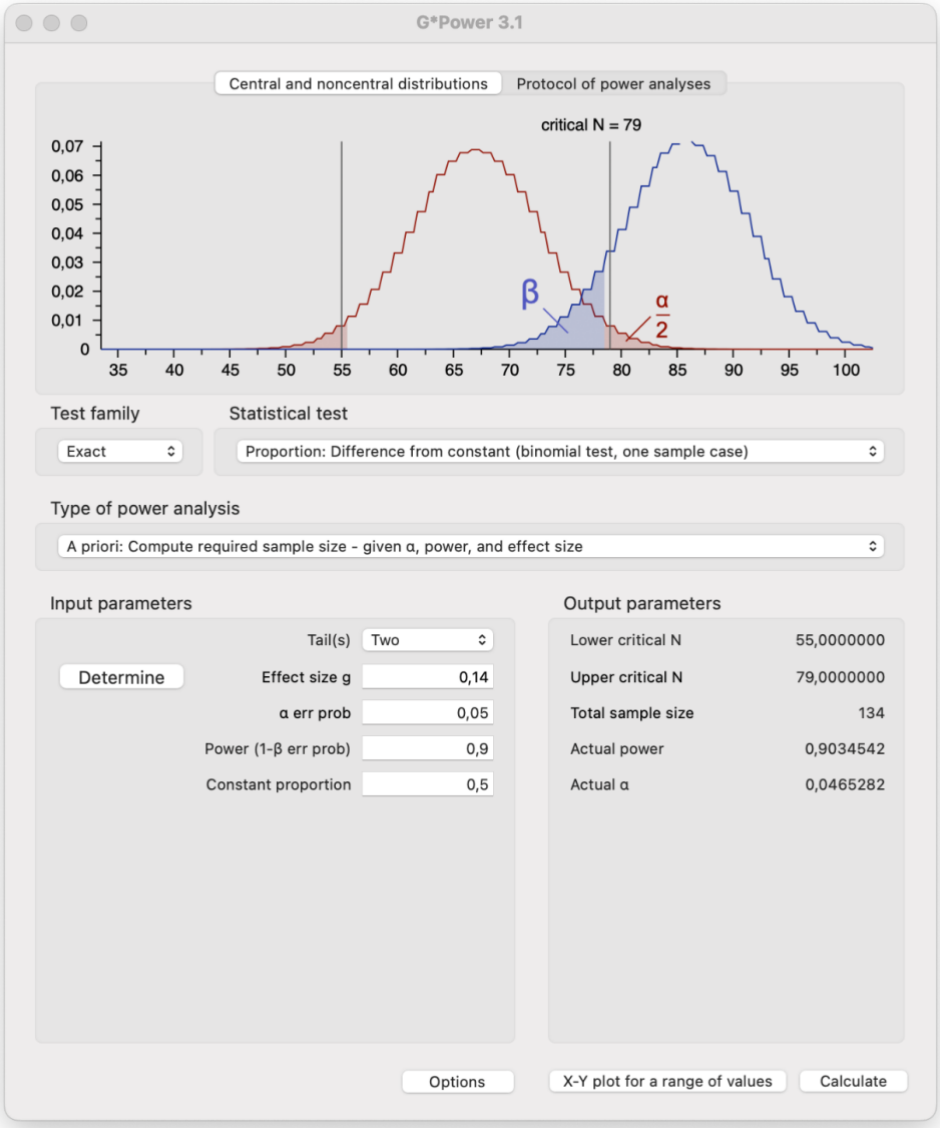
The research output will be shared on Open Science Framework (OSF), conference talks, and in peer-reviewed journals.

Power analysis

Steps for power analysis

Smallest effect size of interest

We aim to be able to detect a small effect size of Cohen`s g = 0.14 at 0.90 power (a = .05). The result of the power analysis suggests a total sample size of (134 x 2) = 268 (see screenshot below).



Summary of power analysis:

The analysis suggests a replication sample size of 268 participants. Based on this, we aim to recruit a total of 300 participants.

Appendix D – Assessment of processing of personal data

Sikt

Notification form / Temporality effect and counterfactual thinking / Assessment

Assessment of processing of personal data

Reference number
754105

Assessment type
Standard

Date
08.03.2023

Project title
Temporality effect and counterfactual thinking

Data controller (institution responsible for the project)
Norges teknisk-naturvitenskapelige universitet / Fakultet for samfunns- og utdanningsvitenskap (SU) / Institutt for psykologi

Project leader
SUBRAMANYA PRASAD CHANDRASHEKAR

Project period
15.01.2023 - 30.06.2023

Categories of personal data
General

Legal basis
Consent (General Data Protection Regulation art. 6 nr. 1 a)

The processing of personal data is lawful, so long as it is carried out as stated in the notification form. The legal basis is valid until 30.06.2023.

Notification Form

Comment
Data Protection Services has assessed the changes registered on 02.03.2023.

Another sample has been added to the notification form (sample 2). Processing is based on informed consent.

We find that the processing of personal data in this project is lawful and complies with data protection legislation, so long as it is carried out as described in the Notification Form with dialogue and attachments.

FOLLOW-UP OF THE PROJECT
We will follow up the project at the planned end date in order to determine whether the processing of personal data has been concluded.

Good luck with the project!

Appendix E – Evaluation of replication classification

Figure S1 – Taxonomy for evaluating classification of replication design by LeBel et al. (2018).

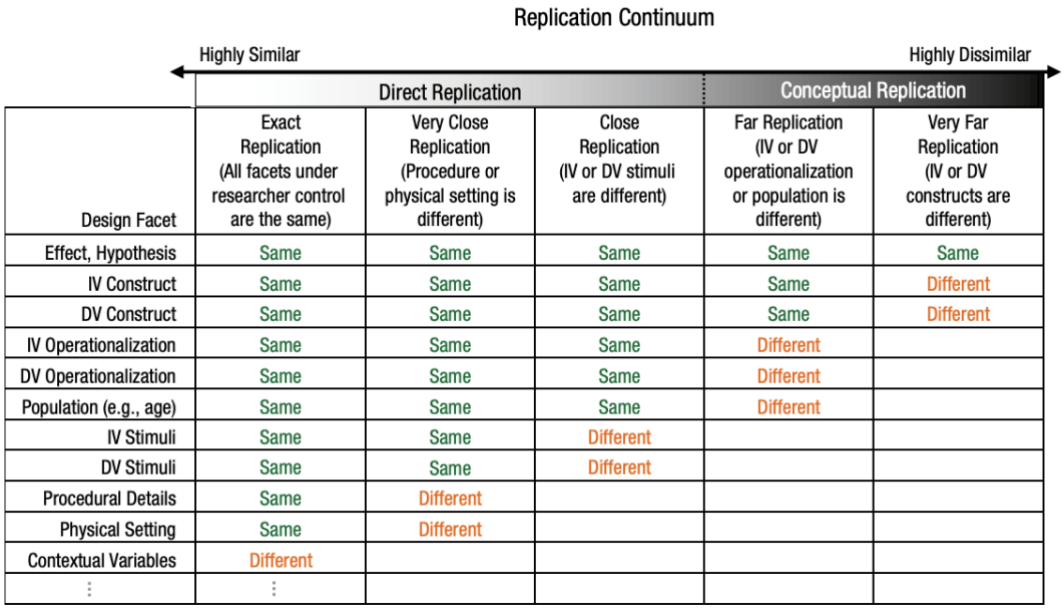


Table S2 - Classification of the replication of Miller & Gunasegaram (1990), based on LeBel et al. (2018)

Design facet	Replication
Effect, hypothesis	Same
IV Operationalization	Similar
DV Operationalization	Similar
Population (e.g., age)	Different
IV Stimuli	Same
DV Stimuli	Same
Procedural details	Similar
Physical setting	Different
Contextual Variables	Different
Replication classification:	Close replication

Table S3 - Classification of the replication of Study 1 of Segura et al. (2002), based on LeBel et al. (2018)

Design facet	Replication
Effect, hypothesis	Same
IV Operationalization	Similar
DV Operationalization	Similar
Population (e.g., age)	Different
IV Stimuli	Similar
DV Stimuli	Similar
Procedural details	Same
Physical setting	Different
Contextual Variables	Different
Replication classification:	Close replication

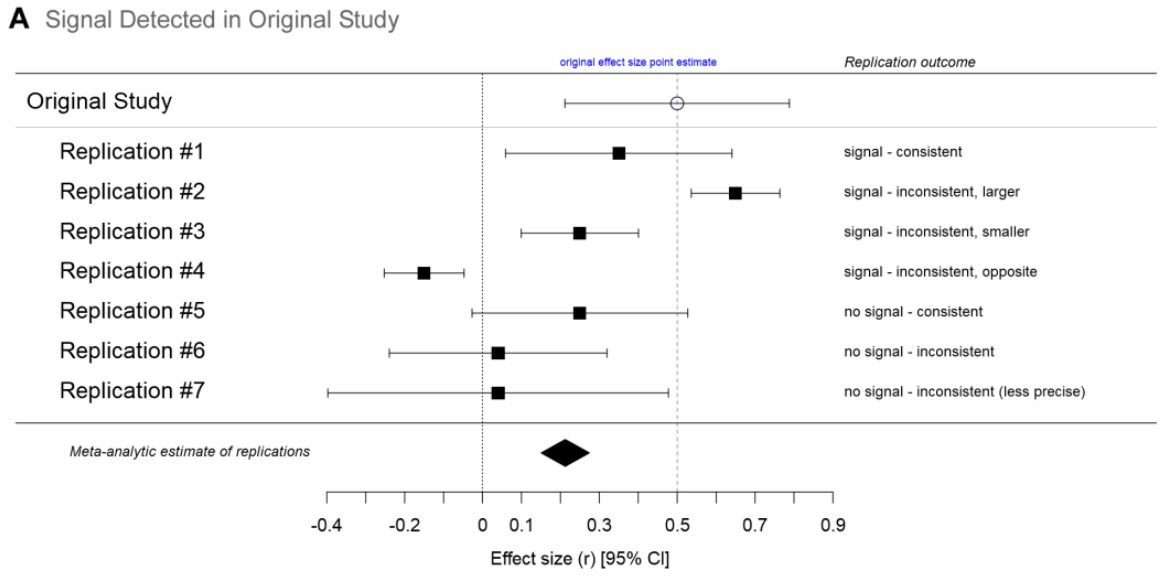
Rationale for replication classification: We evaluated both replications to be direct, close replications based on the criteria set by LeBel et al. (2018). The deviations from the original studies concerning population does not affect our evaluation, as the effect is not limited to a specific target sample. Design facets evaluated as "similar" represent minor deviations from original study.

Appendix F – Taxonomy for evaluating comparison of effect sizes in replications

Table S4. Description of criteria for evaluating comparison of effect size between original study and replication by LeBel et al. (2019). Includes the labels relevant to the current study.

Label	Description
Signal/No signal	If a signal is detected or not (i.e., confidence interval of the replication effect size (ES) excludes zero)
Consistent	Confidence interval of replication ES includes the respective ES of the original study
Inconsistent	Confidence interval of replication ES does not include the respective ES of the original study
Smaller/larger	Labels smaller and larger is used to describe if the confidence interval of the replication ES is larger or smaller compared to the ES of the original study, and is only used if the effect size in replication study is in the same direction as the original study

Figure S5. Taxonomy for evaluating comparison of effect size between original study and replication by LeBel et al. (2019).



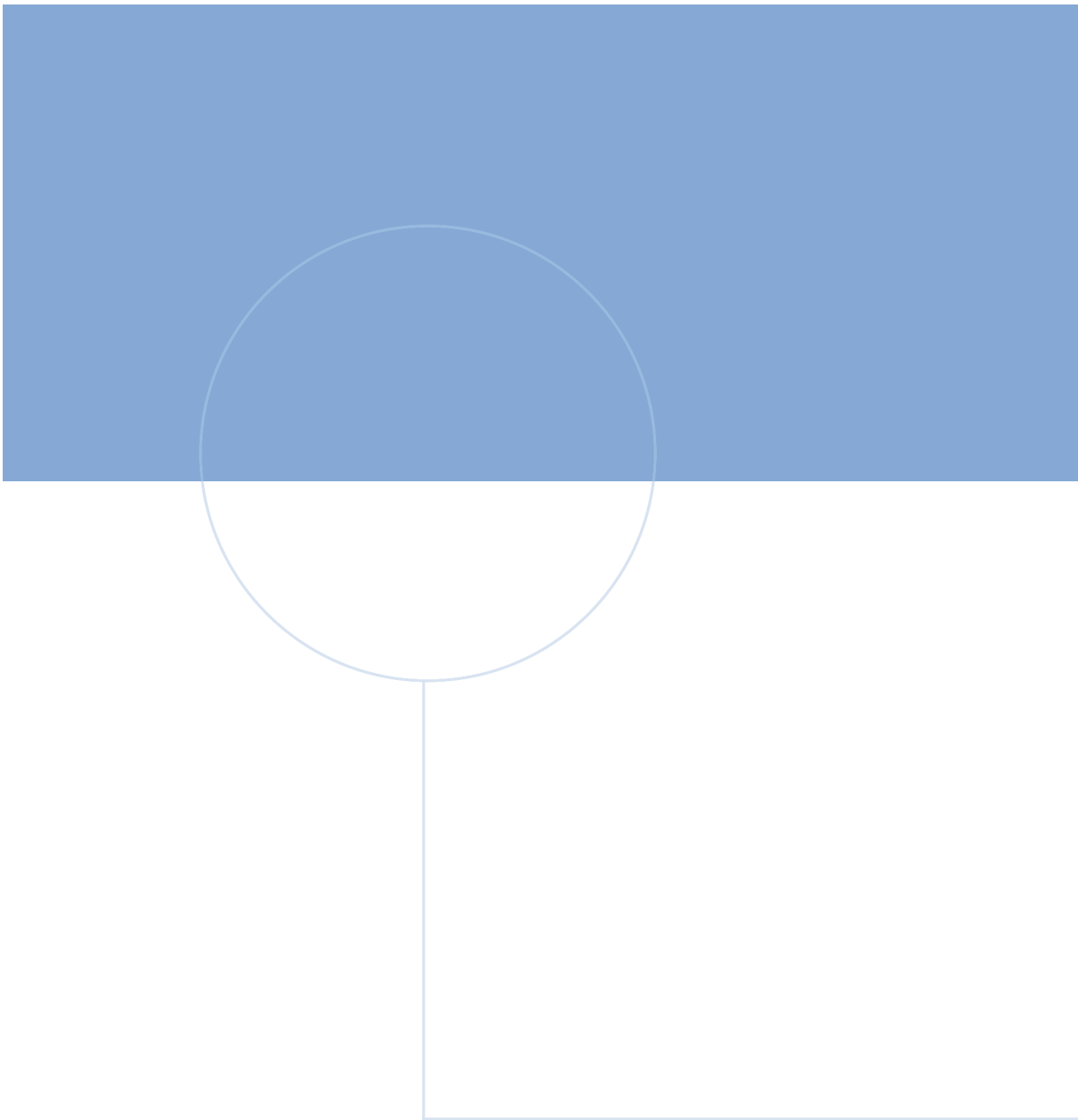
References

LeBel, E. P., McCarthy, R. J., Earp, B. D., Elson, M., & Vanpaemel, W. (2018). Advances in Methods and Practices in Psychological Science. *Psychological Science, 1*(3), 389-402. https://doi.org/10.1177/2515245918787489

LeBel, E. P., Vanpaemel, W., Cheung, I., Campbell, L. (2019). A Brief guide to Evaluate Replications. *Meta-Psychology, 3*(2019), 1-9. https://doi.org/10.15626/MP.2018.843

Miller, D. T., & Gunasegaram, S. (1990). Temporal order and the perceived mutability of events: Implications for blame assignment. *Journal of Personality and Social Psychology*, *59*, 1111–1118. https://doi.org/10.1037/0022-3514.59.6.1111

Segura, S., Fernandez-Berrocal, P., & Byrne, R. M. J. (2002). Temporal and causal order effects in thinking about what might have been. *The Quarterly Journal of Experimental Psychology Section A*, *55*(4), 1295–1305. https://doi.org/10.1080/02724980244000125



NTNU

Norwegian University of
Science and Technology