

Candidate: 10067

The temporality-effect in counterfactual thinking – A replication study and exploration of validity and applicability

BA2900 - BA15 - Temporality effect and
counterfactual thinking

Bachelor's thesis in Bachelor in Psychology
Supervisor: Subramanya Prasad Chandrashekar
May 2023

Approximate word count: 8173

Candidate: 10067

**The temporality-effect in
counterfactual thinking –
A replication study and exploration of
validity and applicability**

BA2900 - BA15 - Temporality effect and
counterfactual thinking

Bachelor's thesis in Bachelor in Psychology
Supervisor: Subramanya Prasad Chandrashekar
May 2023

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



Norwegian University of
Science and Technology

Preface

I was part of a group of students at NTNU's bachelors programme at the institute of psychology, which worked on a project called "BA15 - Temporality effect and counterfactual thinking" for our bachelor thesis. This study consisted of one study that was performed as a cooperation between the six students in the group and our supervisor Subramanya Prasad Chandrashekar, and multiple independent studies which was performed by each individual student in cooperation with S. P. Chandrashekar. The studies were replication studies of previous research on the field, and we heavily focused on open science as well as testing the validity of the original studies. There was no assumption that the results of the original studies would replicate, and all data was analysed objectively. The extension that was added to the replication study of Miller and Gunasegaram's 1990 study is a result of my own ideas and thoughts discussed with my supervisor as well as my fellow students, which have been of great value in terms of discussion partners. The data collection was performed largely by S. P. Chandrashekar, but the analysis was performed independently. The choice of analysis type was derived from the original study that was replicated in the respective study.

I wish to thank my fellow students for invaluable discussions on theoretical aspects of this study, as well as the cooperation in mastering analytical programmes and statistical analysis. Especially one student stood out on the aid in terms of statistical topics. I would like to thank S. P. Chandrashekar for the guidance and patience throughout the semester, and the efforts he has made in order for our success, which he in no way was obligated to exert to such an extent. I also would like to especially thank my girlfriend for her patience, support, and cheers during the process of crafting this thesis, as well as my parents for tolerating long and theoretical phone calls late at night. There has been a lot of social support in order for this thesis to be crafted, but with the exceptions of the statements made in

this preface, and the references used explicitly in text, I declare that this thesis is my own work, based on my own ideas and thoughts.

Abstract

Counterfactual thinking and the temporality effect are concepts that have been studied to a moderate extent mostly in the field of decision-making. There has been some disagreement on the theoretical mechanisms that establish the basis for the phenomena, as well as if the temporality effect is the best way to describe the phenomena. This study investigated the temporality effect in counterfactual thinking through replicating two previously performed studies by researchers that are well-fared in the field. Three general predictions that apply to both of the studies were made, which are; 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. And lastly; 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. The results of the studies both supported all of the hypotheses. All of the original studies' results also replicated in the replication studies and their respective results. These results are then discussed in relation to other contributions on the field of counterfactual thinking and the temporality effect, as well as possible fields of application outside of academical research.

Introduction:

Counterfactual thinking may be defined as mental representations that serve as an alternate depiction of reality (Miller & Gunasegaram, 1990). These mental representations often alter past events and their consequences to the benefit of, or in other cases not to the benefit of, the individual (Byrne et al., 2000). An important aspect of counterfactual thinking is the deeply illogical nature innate to the concept. Counterfactual means contrary to the facts. Therefore, some may wonder why it seemingly is such a common aspect of human behaviour and thought patterns.

In some ways defining counterfactual thinking in a formal language may make the concept more ambiguous and difficult to grasp. In reality counterfactual thinking appears in most people's day to day lives, even though we may not all be conscious of the fact that our own thoughts are counterfactual. For instance, one may think that "Oh, if only I had not broken my ankle at 16, I would be playing professional soccer by now.". Others may think about what part of a morning routine that could have been left out, in order to not have missed the bus that drove away from the bus stop, just as one rounded the final corner 30 meters away. Given these examples, the normality of counterfactual thinking might be more exposed and relatable, while simultaneously providing a basis for research into how these patterns of thoughts may affect various aspects of our lives, such as learning, behaviour, and in the instance of this thesis, decision-making.

Theory:

Decision-making

During everyday life humans face a large number of decisions. These decisions may affect various aspects of one's daily living or general well-being. There are many factors that affect how these decisions are made, and some of them may be more beneficial to the individuals well-being than others. Some argue that decisions are made based upon three concepts. These are our perceived courses of action, meaning the various alternative behaviours one could see as useful in order to resolve the given situation that necessitates the decision in question. The second is the beliefs one has surrounding the objective states, world events and processes that may influence one's actions directly, or indirectly through other causal links. The third concept is surrounding our own values, needs and desires, in addition to our subjective definitions and perceptions of the positive or negative consequences that is perceived to be causally linked to the various possible decisive actions (Hastie, 2001). In previous psychological research there has been descriptions of multiple biases or effects that influence the decision-making process, such as the status quo bias, confirmation bias, various heuristics etc. A less elaborate way of defining decision-making as a concept could be as making a choice between different alternative actions, outcomes, options etc. This means that there is a necessity for there to exist a separate and different alternative to the given outcome that was selected in the decision-making process. These alternatives that never were realized are what are known as counter-factuals (Byrne et al., 2000).

Counterfactual thinking

Counter-factuals may be defined as the outcomes of a decision-making process or the outcome(s) of an event, that never came to be, therefore leaving them to be purely hypothetical or in simpler terms,

contrary to the facts (Miller & Gunasegaram, 1990; Byrne et al., 2000; Walsh & Byrne, 2004). Derived from these counter-factuals arises the term counterfactual thinking. Counterfactual thinking involves the human inclination to dwell on certain aspects of previous events or encounters, more specifically on various hypothetical outcomes of these events or encounters (Miller & Gunasegaram, 1990). The basis on which the counterfactual thinking rests upon is considered somewhat an unresolved matter given that researchers are not entirely sure why such a vast amount of cognitive energy has been set aside for a behaviour which at first glance does not directly affect on our current situation or automatically give us a form of pleasurable experience. As briefly mentioned in the introductory paragraph, this might for instance from an evolutionary standing-point seem as a type of behaviour or usage of cognitive resources that might be cut away at during the evolutionary process of the human. This has however not happened, leading us to believe that this form of counterfactual thinking must have some sort of beneficial effect for the human being's well-being and or survival.

Counterfactual thinking and growth mindset

A common theory as of why counterfactual thinking is so preeminent in human behaviour is that it fulfils a primary cognitive role in an individual's ability to learn from their own mistakes or failures (Sigmundsson, 2017). This theory bases its argument on the ability of the individual to be able to see their own failures in previous situations, as well as the ability to establish a mental representation of the situation as well as it's outcomes and consequences. After this process, in order to learn anything from one's own failures one still has to accept that a different course of action might have led to a better outcome, and be able to utilize this newly acquired knowledge in a similar situation at a later time. Some researchers do in these instances refer to what is known as growth mindsets and fixed mindsets (Sigmundsson, 2017). A fixed mindset could be defined as an individual whose mindset is fixed on their worldview and ability to affect it in a certain deterministic manner. One could argue that an individual with a fixed mind-set might attribute blame in instances of failure to surroundings in their

environment, such as other individuals, rules, time-restraints etc. A growth mind-set however involves the individual being open to embracing that one's own mistakes or short-comings might have affected the result in a negative manner (Sigmundsson, 2017). This could have led to counter-factual thinking being more beneficial in instances where the individual inherits a growth mind-set, where the counterfactual thinking might be revolving around factors that the individual itself could make implications on in a similar situation or event at a later point in time.

The temporality effect

The temporality effect refers to the human tendency in instances where one creates mental representations of situations involving multiple independent events in a sequence, to undo the most recent event (Miller & Gunasegaram, 1990; Byrne et al., 2000; Walsh & Byrne, 2004; Spellmann, 1997). This effect has been documented by various studies, for instance by Kahneman and Miller in 1986, where they created a study in which participants were a number of letters in a sequential order, upon which the participants were then asked to replace one of the letters. They were asked to do so quickly in order to focus on the human tendency, not conscious cognition. The results of the study showed that the participants had a tendency to mutate the last letter of the sequence (Kahneman & Miller, 1986; Miller & Gunasegaram, 1990).

The temporality effect has since been argued to contribute to human decision-making, for instance in how we assign blame, perceive guilt in others, or how we attribute causality to situations, objects, individuals etc (Spellmann, 1997).

Walsh and Byrne made six tenets that they base the mechanisms of the temporality effect on. These could be considered as theoretical assumptions of how individuals view scenarios where counterfactual

thinking occurs both consciously and unconsciously (Walsh & Byrne, 2004). The first tenet is that the individual keeps in mind all possible outcomes that are true, meaning all possibilities that has happened. The second tenet is that the individual is not able to keep in mind all of the counterfactual possibilities of the situation (Walsh & Byrne, 2004). The third tenet is closely tied to the second, as it states that the set of counterfactual possibilities that the individual does keep in mind is guided by the winning conditions (Walsh & Byrne, 2004). This means that the individual prioritizes to keep in mind the counterfactual possibilities where the individuals in the scenario could have won. The fourth tenet is that people have then tendency to mutate certain facts or aspects of the facts in a scenario, in order for the situation to turn out so that the individuals win. This means that people prioritize to alter facts of a situation to counterfactuals, given that this leads to the winning conditions to be met. The fifth is that the first fact we are faced with in a scenario act as an anchor given that it matches the winning possibilities (Walsh & Byrne, 2004). In cases where this is true, the second event is then changed to match the winning possibility established by the first. If the first event does not match the winning possibilities however, the first element might be subjected to counterfactual thinking, and be changed. The sixth and final tenet describes how some parts of true possibilities are thought of explicitly, while others only implicitly, based on how we present the information. An example made to fit this thesis study could be “Cooper or Jones, but not both, flip a coin and gets a head.”. According to Walsh and Byrne the individual then only explicitly imagines certain aspects of the true events, as could be illustrated as follows;

Table 1

Illustration of theoretical mechanism described in Walsh & Byrne 2004

<u>Heads</u>	<u>Tails</u>
Cooper gets heads	Jones gets tails
Cooper gets tails	Jones get heads

The elements in table 1 that is crossed over represents the true possibilities that are described by Walsh and Byrne as implicit, while the elements highlighted are the ones that are thought about by the individual explicitly.

Walshe and Byrne argue that these tenets, and this model illustrated in table 1 is one possible theoretical mechanism that could be used as a baseline to explain why the temporality effect occurs, based upon how individuals look for the first true fact that matches the pictured win condition. They also conducted three studies in which they flipped the winning condition in order to see if this would reverse the temporality effect, and in this way further increase their theory. This means that if the win condition was presented as “Cooper or Jones, but not both, flip a coin and gets a tail.”, the model illustrating the implicit and explicit elements would appear different, and so would the temporality effect, in that it would be reversed (Walsh & Byrne, 2004).

Table 2

Illustration of theoretical mechanism described in Walsh & Byrne 2004

<u>Tails</u>	<u>Heads</u>
Cooper gets heads	Jones gets tails
Cooper gets tails	Jones get heads

When this win condition was presented to a computer program made in order to test the temporality effect, the results told Walsh and Byrne that the temporality effect had indeed flipped, strengthening this theoretical explanation of the temporality effect (Walsh & Byrne, 2004).

Temporal and Causal chains

In multiple of the research papers that have been made on decision-making and counterfactual thinking, there is a reference to multiple types of chains of events. A causal chain may be defined as a chain of

events that are sequentially set in motion by one another. An example could be an event where the alcoholism of a father causes a mother to commit suicide and the combination of the two of these leading to or contributing to juvenile history of crime or general delinquency in a son (Spellmann, 1997). This example was used in a paper by Barbara A. Spellmann which shines light upon the difference between temporal and causal chains, as well as the concept of mutability in order to establish how we credit causality in various events (Spellmann, 1997).

Temporal chains differ from causal chains in multiple ways, the most important one being that the events are not caused by one another, ergo they are not automatically causal events (Spellmann, 2017; Miller & Gunasegaram, 1990). Since the events are not caused by one another this also means that the order the events appear in may be rearranged freely without this affecting the substance of the chain (Spellmann, 1997). An example of this could be the coin toss scenario created by Miller and Gunasegaram in their 1990 study of counterfactual thinking and the temporality effect. Here the test subjects were introduced to two hypothetical individuals that were given the opportunity to win 1000 dollars (Spellmann, 1997; Miller & Gunasegaram, 1990). The premise was that they each had to flip a coin, and the upwards facing side of the coin had to be the same for both coin-tosses for the individuals to claim their price. In the study Jones has the first toss and gets a head. Cooper then goes second and tosses a tail, meaning that neither of Jones nor Cooper receive a cash price. The participants of the study were then presented with the following statement; “There are two ways that Jones and Cooper could have won \$1000. Which of these alternatives come more readily to mind: (a) Jones tossing a tail; (b) Cooper tossing a head.” The results of the study later revealed that 89% of the participants chose to mutate the second event rather than the first. This study was used by Miller and Gunasegaram to strengthen the hypothesis of the temporality effect (Miller & Gunasegaram, 1990).

Spellmann's Model – The Crediting Causality Hypothesis

There has as previously mentioned been some counterarguments and competing explanations to the temporality effect in counterfactual thinking, when it comes to explaining the patterns that show in various studies on human decision-making. One of these theories is described by Barbara A. Spellmann in an article where she introduces the crediting causality hypothesis. In the article Spellmann applies the crediting causality hypothesis to multiple studies on decision-making, one of them being Miller and Gunasegaram's 1990 study on the temporality effect in counterfactual thinking (Spellmann, 1997). In this instance she argues that the study's participants calculate the implication the action of the first individual will have on the collective chances of winning the cash prize. When the first individual flips the coin, the odds of the two individuals succeeding in getting matching sides of the coin do not go down. The percentage chance is still 50% vs 50%. When the second individual flips their coin however, the result is decided in terms of either the coins matching and the individuals win their prize, or the coins do not match, meaning the individuals earn no prize. Spellmann then argues that this reasoning means we credit causality to the second individual (event), which shows in the Miller and Gunasegaram study by the participants of the study when they are asked to select which of the characters in the scenario has the highest probability of feeling guilty for the forsaken chance at a prize (Spellmann, 1997).

Why do we replicate – The replication crisis

Michael Gibbons describes a sort of implicit contract between sciences and the general public (Anvari & Lakens, 2018). In this contract he states that science has long received the public's support and in order to continue to do so, science needs to give something in return to the public. Science then provides the public with knowledge about various aspects of our world and the human existence in an orderly and reliable manner. This is part of the social contract that is society, where no single individual may excel in all fields, but rather should attempt to master the craft where their talents shine the

brightest (Packer, 2017). Following this principle, it then becomes apparent that the researchers conducting experiments in the area of psychology need to conduct their experiments in a way so that the public can leave the responsibility of their knowledge being true and usable in the hands of the researchers. The problem arises when we no longer take the necessary means and precautions to establish an open science where the data and processes behind science is open and understandable to everyone, instead of hidden behind metaphorical veils, impenetrable by the average member of society.

In the history of social psychology, there has been multiple cases of experiments and studies that pique the public's interest both in positive and negative ways. Sadly, a common denominator in multiple of these studies are the lack of focus on open science, validity and replicability (Anvari & Lakens, 2018). Some of the studies did not yield the same results when replicated, while others were not possible to replicate at all. One explanation for this may be that the focus on validity or ethics might not have been as high a couple of decades into the past, as it is as of now. This however does not give us a reason to not be critical towards research that have a low case of validity or is hard to replicate in order to either strengthen or debunk the argument that was proposed in the original study (Anvari & Lakens, 2018). In addition to this some may argue that since psychology is a social study, it may change accordingly to other human phenomena, such as culture and politics, giving us further reason to replicate various studies as time goes by (Anvari & Lakens, 2018).

One way to counteract this effect regarding our ever-changing society affecting results of previous research, as well as keeping a firm grip of the concept of open science, is to practice pre-registration of planned research (Anvari & Lakens, 2018). This means to submit a standardized form or document that establishes the planned research in terms of basis for the study, goals, methods, ethics etc. In the case of the studies that were replicated for this thesis, there were multiple cases where parts of the statistical data were incomplete, and in addition to this not pre-registered. This, for instance, allows for the

possibility that the various hypothesis presented in the original study were established after the data was analysed, in order to gain the statistically best result possible. This might make for a more interesting article that might do better in terms of publicity or fund-raising, but it may also present some downsides that in some cases can turn severe over time. One example is through breaking the previously mentioned social contract between society and science. In a hypothetical scenario in which psychological science becomes infamous for deluding the public through invalid research, be it so intentional or unintentional, the psychological sciences may not only lose hard-earned credibility, but also the very basis of their existence. This makes replicating past research not only scientifically interesting but also important to the psychological science's integrity and perceived credibility.

Methods

The research that has gone into this study is two different replication studies. One of the studies was performed in Norwegian, while the other was performed in English. These studies are replication-studies of the experiment performed originally by Miller and Gunasegaram in 1990. In addition to this, the other participants that contributed to the study, and are writing their thesis under the supervision of the same supervisor, added multiple extensions to these replication studies, in order to further expand on a previously established phenomenon, or in order to explore something different that in some manner is still relatable to the original study. In my case, I wanted to expand upon the aspect of linguistics, and how this may or may not counteract the temporality effect in counterfactual thinking. Therefore, my extension is based upon a replication of a study performed by Byrne et al. in 2000, where the coin-toss phenomenon was introduced to a group of Spanish students (Byrne et al., 2000). In my

extension the study was translated to Norwegian, and we also added another question more directly inclined to establish an understanding of individual's own awareness of linguistic biases.

Miller and Gunasegaram (1990) Replication Study

Hypotheses:

There are three hypotheses for this replication that apply to both scenario 1 and scenario 2 of the study. These are derived from the original study's hypothesis, and are meant to investigate both the effect observed in this study, as well as comparing the results of the replication to the original study's findings.

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.

Participants:

The participants in the replication study of Miller and Gunasegaram were recruited at various locations in Trondheim, Norway, as well as by promotion on social media and student specific means of online communication. This was specific to the replication of the study by Miller and Gunasegaram (1990) and the extension that was added to this specific replication. The total number of participants accumulated to $n = 211$. Of these 138 were female (65.4%), 72 were male (34.1%) and 1 did not identify with either (0.5%). All of the participants were above 18 years of age, ranging from 19 to 90. The mean age was 28 years ($M_{age} = 28.51$, $SD_{age} = 13.68$). The median age was 23 years of age, and this may be important to note as 66.7% of the participants were 24 years of age or below.

All of the participants answered the questionnaire completely anonymously through a website called Nettskjema. Nettskjema is a Norwegian web-site that specialises in data-collection through questionnaires, and is trusted to ensure that our research is in term with both GDPR and the current data-protection laws in general and the laws specifically concerning scientific research. All of the participants were required to actively consent to their submissions being used in the research as well as published in later times, and they were also informed that they may withdraw their consent.

Procedure:

The research that was conducted in preparation for this thesis was as previously stated replication studies. The first study was a replication of Miller and Gunasegaram's study on the temporal effect in counterfactual thinking which was originally conducted in 1990. The research questionnaire that was presented by Miller and Gunasegaram was replicated as closely to identical as possible in regard to the questionnaires structure and phrasing. The questionnaire was also translated to Norwegian, so that it may be conducted in the local language.

There was also added an extension to this replication, that builds of a combination of the Miller and Gunasegaram study and the study performed by Byrne et al. The concept of the extension is to further expand on the testing of the temporality effect by trying to interfere with the effect by subconsciously appealing to an individual's linguistic preference. In the case of Byrne et al., the researchers made a coin-flip scenario where in scenario 1 the first individual would flip the coin to "heads", and the second individual "tails" (Byrne et al., 2000). After this the participants of the study were tasked to complete the first part of the questionnaire regarding scenario 1. The participants were then asked to read scenario 2, in which the order of the linguistic phenomena was swapped, so that the first hypothetical individual in scenario 2 received "tails", and the second individual "heads". The goal of this was to investigate whether or not the temporality effect could be eliminated by an individual's preference towards the terms "heads" or "tails". This was then applied to the Miller and Gunasegaram study, in Norwegian, in order to investigate if the results that were gathered by Byrne et al would replicate in another language than the one utilized in the original study.

The study was pre-registered using the OSF-platform to follow the guidelines of open science as closely as possible. The pre-registrations are publicly available on the OSF-site and will also be found in the supplementary materials alongside the full questionnaire and dataset.

Measurement instruments:

The first instrument of measurement was *mutability*. This instrument was designed to investigate how likely the participants were to mutate events that underwent in a hypothetical scenario. There were always two different options that the participants were to choose from, this was different from the

original study, where the participants had to write their own answers in a blank textbox (Miller & Gunasegaram, 1990).

The second instrument of measurement was *guilt*. This instrument's main focus was to highlight how one might assume that an individual will experience an unjustified level of guilt based on a randomly decided chance event.

The third and final instrument of measurement was *blame*. The purpose of this instrument was to further elaborate on the assignment of blame, based on temporal order of chance events.

The participants were also asked for their gender and age, though this was not directly relevant to the analysis we are going to perform on the other measurement instruments.

The measurement instruments that are presented here are used in both the study performed by Miller and Gunasegaram as well as the study performed by Byrne et al and are therefore used in the replication-studies as well. They were also used in the extension that was made on the replication study of Miller and Gunasegaram and is therefore protruding in multiple of the tables that will follow in this thesis.

Design:

The design of this study is an online questionnaire that relies on self-reports. The questionnaire was directly derived from the questionnaire that was presented in the original study, with the exception of the removal of the blank textboxes being replaced with only two options to choose from on the separate

measurement instruments. For further information and details on the design of the replication study, refer to the pre-registration document in the supplementary materials.

Results Scenario 1:

In scenario 1, where the order of linguistic phenomena was unaltered from the original study by Miller and Gunasegaram, the result of the measurement instrument *mutability* suggests an inclination amongst the participants to mutate the last event in instances of two independent events (76.3% vs 23.7%), $n = 211$, $k = 161$, $z = 7.64$, $p < .001$ (Cohen's $g = 0.26$ [0.2, 0.32]). The results of the measure instrument *guilt* suggest a tendency to attribute more guilt towards the hypothetical individual that flipped their coin last (93.4% vs 6.6%), $n = 211$, $k = 197$, $z = 12.6$, $p < .001$ (Cohen's $g = 0.43$ [0.4, 0.47]). Lastly, the results of the measure instrument *blame* reveal a tendency of the participants to expect the first hypothetical individual to blame the last individual to a larger extent (91.9% vs 8.1%), $n = 211$, $k = 194$, $z = 12.19$, $p < .001$ (Cohen's $g = 0.42$ [0.38, 0.46]).

Table 3

Comparison and interpretation of original studies effect sizes and replication studies effect sizes

Study	Original study effect size	Replication effect size	Interpretation
Study 1 - Miller and Gunasegaram, 1990			
Scenario 1*	Cohen's g	Cohen's g	
Mutation	0.39 [0.42, 0.45]	0.26 [0.2, 0.32]	Signal – inconsistent, smaller
Guilt	0.36 [0.29, 0.44]	0.43 [0.4, 0.47]	Signal – inconsistent, larger
Blame	0.42 [0.36, 0.48]	0.42 [0.38, 0.46]	Signal - consistent

*** S1 = Heads-Tails

Based on the contents of table 3, and the results reported in text, this replication study finds that the second event is deemed more mutable than the first. This means that hypothesis 1 is confirmed. The effect size is however smaller than in the original study, which could be interpreted as the original findings not being fully replicated in this study to the same extent.

Hypothesis 2 is confirmed. There was a large number of the participants that judged the second player to experience more guilt than the first player. The effect size of this measurement instrument was higher than in the original study, meaning that the results replicated in the sense that it appeared, but in a stronger fashion.

Hypothesis 3 is also confirmed. The original study and the replication were close to identical in both the percentage of the participants that selected the various options (original study 92% vs 8% and replication 91.9% vs 8.1%). The effect sizes are close to identical as well, meaning that the original study's result replicated in this study.

Based on the comparison of the effect sizes in table 3 the original study replicated well in scenario 1.

Results Scenario 2:

In scenario 2, where the order of linguistic phenomena was altered from the original study by Miller and Gunasegaram, the result of the measurement instrument *mutability* suggests an inclination amongst the participants to mutate the last event in instances of two independent events (81.5% vs 18.5%), $n = 211$, $k = 172$, $z = 9.16$, $p < .001$ (Cohen's $g = 0.32$ [0.26, 0.37]). The results of the measure instrument

guilt suggest a tendency to attribute more guilt towards the hypothetical individual that flipped their coin last (91.9% vs 8.1%), $n = 211$, $k = 194$, $z = 12.19$, $p < .001$ (Cohen's $g = 0.42$ [0.38, 0.46]). Lastly, the results of the measure instrument *blame* reveal a tendency of the participants to expect the first hypothetical individual to blame the last individual to a larger extent (90.5% vs 9.5%), $n = 211$, $k = 191$, $z = 11.77$, $p < .001$ (Cohen's $g = 0.41$ [0.36, 0.44]).

Table 4

Illustration of effect sizes in the extension of the replication of Miller & Gunasegaram 1990

Study	Effect size	
Study 1 - Miller and Gunasegaram, 1990, Extension		
Scenario 2*	Cohen's g	
Mutation	0.32 [0.26, 0.37]	
Guilt	0.42 [0.38, 0.46]	
Blame	0.41 [0.36, 0.44]	

***S2 = Tails-Heads

In scenario 2, both hypothesis 1, 2 and 3 are confirmed, based on the results reported in text above.

Byrne et al 2000 Experiment 4 – Replication Study

Hypotheses:

For this replication study the same three hypotheses used in the replication of Miller and Gunasegaram's 1990 study will be used. This is in order to allow for comparison between the two, and having a broader data set for the hypothesis.

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.

Methods

Participants:

The participants of the replication study of Byrne et Al's 2000 experiment 4, were recruited through an online service, Prolific, which offers a service where individuals are offered compensation (minimum of 8 USD an hour) in order to participate in various research. The full number of participants was 339.

Of these 157 were female (46.3%), 176 were male (51.9%) and 6 identified as non-binary (1.8%). All of the participants were above 18 years of age, ranging from 18 to 93. The mean age was 40 years ($M_{age} = 39.92$, $SD_{age} = 13.82$). The median age was 36 years of age.

Procedure:

This study is a direct replication of a study performed by Byrne et Al in 2000, with the exception of an omitted scenario regarding playing cards, and the language used to be English instead of Spanish. The red card-black card scenario was omitted due to this thesis' focus on the coinflip scenario both in the Miller and Gunasegaram 1990 replication, and in this replication study. The study by Byrne et al. also implemented a scenario 2 in order to explore the effects of the linguistic aspect of the temporal chains in the hypothetical coin flip scenarios performed by Miller and Gunasegaram, which was the main inspiration for the extension in the replication of Miller and Gunasegaram's study. This strengthens the reasoning for omitting the card scenario from the Byrne et al. replication further, and strengthens the reasoning behind focusing on the coin flip scenario.

The original study performed by Byrne et Al was part of a research-paper consisting of five different studies (Byrne et al., 2000). They were not related in terms of their results, and were independent from one another. This allowed us to directly replicate this study, which in the original was labelled as "study 4", with a few exceptions as mentioned previously.

The study was pre-registered using the OSF-platform to follow the guidelines of open science as closely as possible. The pre-registrations are publicly available on the OSF-site and will also be found in the supplementary materials alongside the full questionnaire and dataset.

Measurement instruments:

View the section on measurement instruments on p. 18.

Design:

The design of this study is an online questionnaire that relies on self-reports. The phrasing in the questionnaire was directly derived from the questionnaire that was presented in the original study, with the exception of the removal of the blank textboxes being replaced with only two options to choose from on the separate measurement instruments and names of the fictional individuals participating in the study. As previously mentioned, a scenario surrounding different coloured cards were omitted. See p. 23 for more filling information on this. For further information and details on the design of the replication study, refer to the pre-registration document in the supplementary materials.

Results Scenario 1:

In scenario 1, where the order of the linguistic phenomenon was presented in the “natural” manner (heads-tails), the result of the measurement instrument *mutability* suggests an inclination amongst the participants to mutate the last event in instances of two independent events (8.3% vs 91.7%), $n = 168$, $k = 154$, $p < .001$ (Cohen’s $g = 0.42$ [0.37, 0.46]). The results of the measure instrument *guilt* suggest a tendency to attribute more guilt towards the hypothetical individual that flipped their coin last (8.9% vs 91.1%), $n = 168$, $k = 153$, $p < .001$ (Cohen’s $g = 0.41$ [0.37, 0.45]). Finally, the results of the measure instrument *blame* reveal a tendency of the participants to expect the first hypothetical individual to blame the last individual to a larger extent (10.7% vs 89.3%), $n = 168$, $k = 18$, $p < .001$ (Cohen’s $g = -0.39$ [-0.35, -0.44]).

Table 5*Comparison and interpretation of original studies effect sizes and replication studies effect sizes*

Study	Original study effect size	Replication effect size	Interpretation
Study 2 – Byrne et al, 2004, Experiment 4			
	Cohen's <i>g</i>	Cohen's <i>g</i>	
Scenario 1*			
Mutation	0.24 [0.10, 0.38]	0.42 [0.37, 0.46]	Signal – inconsistent - larger
Guilt	0.47 [0.40, 0.54]	0.41 [0.37, 0.45]	Signal – inconsistent – smaller
Blame	-0.45 [-0.56, -0.35]	-0.39 [-0.35, -0.44]	Signal – inconsistent -smaller
Scenario 2*			
Mutation	0.00 [-0.16, 0.16]	0.35 [0.3, 0.41]	Signal – inconsistent – larger
Guilt	0.48 [0.42, 0.53]	0.4 [0.35, 0.44]	Signal – inconsistent – smaller
Blame	-0.47 [-0.53, -0.41]	-0.39 [-0.34, -0.44]	Signal – inconsistent - smaller

*** S1 = Heads-Tails, S2 = Tails-Heads

Based on the contents of table 5, and the results reported in text, this replication study finds that the second event is deemed more mutable than the first. This means that hypothesis 1 is confirmed.

There was a large number of the participants that judged the second player to experience more guilt than the first player. Hypothesis 2 is on this basis confirmed.

Hypothesis 3 is also confirmed. The study showed that the participants had an inclination towards judging that the second person would be blamed by the first person to a further extent.

Based on the comparison of the effect sizes in table 5 the original study replicated well in scenario 1.

Results Scenario 2:

In scenario 2, where the order of the linguistic phenomenon was presented in the “un-natural” manner (tails-heads), the result of the measurement instrument *mutability* suggests an inclination amongst the participants to mutate the last event in instances of two independent events (14.6% vs 85.4%), $n = 171$, $k = 146$, $p < .001$ (Cohen’s $g = 0.35$ [0.3, 0.41]). The results of the measure instrument *guilt* suggest a tendency to attribute more guilt towards the hypothetical individual that flipped their coin last (10.5% vs 89.5%), $n = 171$, $k = 153$, $p < .001$ (Cohen’s $g = 0.4$ [0.35, 0.44]). In scenario 2 the results of the measure instrument *blame* also reveal a tendency of the participants to expect the first hypothetical individual to blame the last individual to a larger extent (11.1% vs 88.9%), $n = 171$, $k = 19$, $p < .001$ (Cohen’s $g = -0.39$ [-0.34, -0.44]).

Based on the data collected from scenario 2, all three of the hypotheses are confirmed. Based on the comparison of the effect sizes in table 5 the original study replicated well in scenario 2.

Discussion:

Specific discussion, limitations and further research:

The original findings in both the study by Miller and Gunasegaram and the study by Byrne et Al replicated in the studies presented in this thesis. In the replication on Miller and Gunasegaram the arguably most important aspect to note is that this study was performed in Norwegian on in an entirely different society and time. This may be used to argue in order to strengthen the foundation of the temporality effect in counterfactual thinking. The broader coverage of the worlds population that is included in the studies performed on the temporality effect, the more applicable the studies are on the outside world and general public. This is positive in the sense that it deems the work of Miller and Gunasegaram as true and effective research. There are however some counterarguments that could be made in how the studies, both in terms of the original studies and the replications, were performed.

In the original study by Miller and Gunasegaram there were as previously mentioned blank text boxes where the participants would fill in their own answers (Miller & Gunasegaram, 1990). In the replication however, the participants were given to different boxes to check in order to select which of the alternatives felt the most natural to them. Some may argue that this could affect the participants to select an alternative answer that is not necessarily the same as the one they would have selected, given that they were to type their own phrasings and answers. Given that the replication study was to be fully identical, thereby also including blank text-boxes, the analysis would have taken a great deal more work in terms of coding the answers into a format that could be run through a statistical program, in our case,

R, which is free and open source. By selecting to only allow two different answers, we avoided this work-load, and therefore could distribute more time to other important aspects of the studies.

An alternative solution to this issue could have been to add a third option that stated; “Neither of the above.”. Doing so would have allowed the participants to express that they did not feel that either alternative was more appealing, and thereby not forcing all participants into the first or second checkbox. Since the participants had no option to not express a preference, some may argue that a certain percentage of the participants selected a random answer out of the two, thereby negatively affecting the results, in terms of the validity. There are several reasons as to why the studies presented may avoid this, in addition to the results being statistically significant. One of which may be of great importance is that both of the scenarios in both of the replication studies all deemed the same results, which were in favour of the temporality effect. This means that both the American participants gathered on Prolific, as well as the Norwegian participants whom answered through Nettskjema, which in total is 550 participants, all yielded relatively similar data.

Another important note in terms of the validity of our replication studies, is how the alteration of the order of linguistic constructs did not seem to affect the occurrence of the temporality effect. This is true both in the replication of the Byrne et Al study, but also for the extension added to the Miller and Gunasegaram replication study. Prior to the start of the research, the group of students that are working on their thesis’s surrounding the temporality effect discussed whether or not we thought there was a linguistic preference in terms of the order of their appearance (heads-tails vs tails-heads) in the Norwegian language, and how this would affect the replication results. The study however clearly shows that there is a preference in the Norwegian language as well, at least based on the participants of the study, where 88.6% stated that “kron eller mynt” is the phrasing that appear the most natural to them. Even though the linguistic preference is so prevalent in the studies’ population, it did not

counteract the temporality effect, as the results in instances where the linguistic order is flipped is close to the scenarios where the linguistic order was not flipped, and sometimes even more in favour of the temporality effect.

General discussion:

Another way to critique the research that has gone into the replication studies in this thesis could be by applying the theory of the temporality effect to the questionnaire itself. Some might argue that the order in which the available answers to the various questions are presented to the participant might affect their answer. This could also be tied to the model illustrated in relation to tenet number six by Walsh and Byrne, in the sense that the questionnaire that was used for our studies paved the way for a specific win condition, even though it was not supposed to. One phrasing of a scenario in the questionnaire for the Byrne et Al replication study is as follows;

“Imagine two individuals (John and Michael) 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 \$1,000. However, if the two coins do not come up the same, neither individual wins anything. John goes first and tosses a head; Michael goes next and tosses a tail. Thus, the outcome is that neither individual wins anything.” (Example 1, from Replication study of Byrne Et al 2000)

One may argue that this phrasing, specifically “John goes first and tosses a head”, might work as establishing a win condition, and allowing the participants for our study to establish “head” as the true possibility, and therefore the reference point in which Michael fails to follow, as he receives a tail. This renders us with multiple things, one of them being that this could explain why swapping the order of the linguistics phenomena (using tails-heads instead of heads-tails) did not stop the temporality effect from

occurring. If we however had done as Walsh and Byrne did with their computer programme, and attempted to reverse the temporality effect by altering the description of the scenario and its events, opposite to what is seen in the paragraph that was taken from the actual questionnaire above, and then not swapping the order of the available answers. One could argue that this may reverse the model Walsh and Byrne argues that individuals use in order to keep certain events explicitly in mind and others implicitly, and thereby reversing the temporality effect, in the same way as the study by Walsh and Byrne did in 2004, only with human participants instead of a computer programme. This might be something that could be explored further in another study or replication study at a later time.

There is almost an unlimited number of ways that one could speculate in various ideas of how the temporality effect in counterfactual thinking works and how it may be displayed in various experiments, simply by combining and adding to the studies that have been presented in this thesis, and other articles written by researchers in this specific field. This is scientifically interesting, and is very useful for the field both in terms of development of theories, but also in terms of the validity and reliability of the theories, as described earlier in regards to replication-studies.

However, one might wonder how these findings on the temporality effect in combination with the theory presented might affect the public in general, and their day to day living. As previously stated, if one were to view extensive occurrence of counterfactual thinking through the eyes of evolution, it might at first glance be difficult to understand why we dispose such large amounts of cognitive capacity to untrue realities. Counterfactual thinking is stated to be an important source of learning when it comes to reviewing one own's mistakes, as well as developing an early intention for similar situations in the future (Walsh & Byrne, 2004; Sigmundsson, 2017). The idea of a growth mindset was mentioned earlier in the thesis, and this is something that could be elaborated on in terms of how we choose to utilize this knowledge of how we attribute causality, blame and guilt through counterfactual thinking.

Someone who inhabits a fixed mindset, known as the opposite of a growth mindset, might be more likely to blame own failure on factors out of the individual's control, such as weather, time, etc, thereby rendering counterfactual thinking as relatively unproductive, as it might be hard to plan future events around scenarios that the individual has no control over. Someone with a growth mindset might, according to Sigmundsson (2017), be more likely to look for certain aspects of their own behaviour that could be altered in order to achieve a different result. This could be argued to be strikingly similar to counterfactual thinking, and also be of relevance to our specific research in regards to the temporality effect in counterfactual thinking. Both in terms that some individual might be able to shift their mindset more towards a growth mindset than a fixed mindset, on the basis of the information provided from these studies, but also in terms of our own understanding of how we may be misled by unconscious effects such as the temporality effect. This way counterfactual thinking, and our conscious understanding of it and its mechanisms might help us to lead more productive lives and avoid repeating past mistakes.

There has already been established that counterfactual thinking might be beneficial to the individual's quality of life, given the correct circumstances. If one was to take this discussion a step further, it could be in terms of how the research from these studies might shine light on how we view causality in day to day living, and how this could be utilized in various aspects of the human experience, one might look to fields like organizational psychology or applied social psychology. Barabara A. Spellmann did in the study where she established the model of crediting causality, discuss the difference between mutability and causality, as well as temporal chains of events vs causal chains of events (Spellmann, 1997). In our studies, we utilize what is described as temporal chains by the original researchers, but still, we make the participants decide matters of blame and guilt. If one were to use these studies in settings such as work places, higher education, and other stress filled environments that also involve a lot of cooperation, some might argue that this deeper understanding of what is causal and what is temporal

might lead to diffusion of certain conflicts, maybe even before they arise. It is also possible to argue that some might benefit from learning how humans sometimes “make up” causality in certain situations, where no-one or nothing is to blame.

Conclusion:

This study investigated the temporality effect in counterfactual thinking through replicating two previously performed studies by researchers that are well-fared in the field. Three generic predictions that apply to both of the studies were made, which are; 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. And lastly; 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. The results of the studies both supported all of the hypotheses. All of the original studies’ results also replicated in the replication studies and their respective results. The results were then discussed both in terms of their validity and the quality of the research that was performed, in addition to debating various possible limitations of this study and how future studies may improve or elaborate on these aspects. The results were also discussed in regards to applicability to everyday life and the general public, in order to establish relevance for the non-academic reader. This study statistically supports the temporality effect in counterfactual thinking, its occurrence. The study also highlights how counterfactual thinking as a concept may relate to important social and economical fields and our progress both as individuals and society by helping us understand each other, situations we appear in, and ourselves to a further extent.

Bibliography

Anvari, F., Lakens, D. (2018). The replicability crisis and public trust in psychological science. *Comprehensive Results in Social Psychology*, 3(3), p. 266–286.
<https://doi.org/10.1080/23743603.2019.1684822>

Byrne, R. M. J., Segura, S., Culhane, R., Tasso, A., Berrocal, P. (2000). The Temporality Effect in Counterfactual Thinking of What Might Have Been. *Memory & Cognition*, 2000(No 28(2)) p. 264 – 281.
<https://doi.org/10.3758/BF03213805>

Hastie, R. (2001). Problems for Judgment and Decision Making. *Annual Review of Psychology*, 52(1), p. 653–683.
<https://doi.org/10.1146/annurev.psych.52.1.653>

Kahneman, D., Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review*, 93(2), p. 136–153.
<https://doi.org/10.1037/0033-295X.93.2.136>

Miller, D.T. & Gunasegaram, S. (1990). Temporal Order and the Perceived Mutability of Events: Implications for Blame Assignment. *Journal of Personality and Social Psychology*, 1990(No 6), p. 1111 – 1118.
<https://doi.org/10.1037/0022-3514.59.6.1111>

Packer, M. J. (2017). *Child Development*. SAGE

Sigmundsson, H. (Ed). (2017). *Læring og ferdighetsutvikling*. Fagbokforlaget

Spellman, B. A. (1997). Crediting causality. *Journal of Experimental Psychology: General*, 126(4), p. 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), p. 369–378.

<https://doi.org/10.3758/BF03195831>

Supplementary Materials

Miller & Gunasegaram (1990) Replication Pre-Registration:

Project working title: Replication and extension 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). Even though the two scenarios may seem similar, each scenario is different in important ways. 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 participants. 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
<i>n</i>		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]

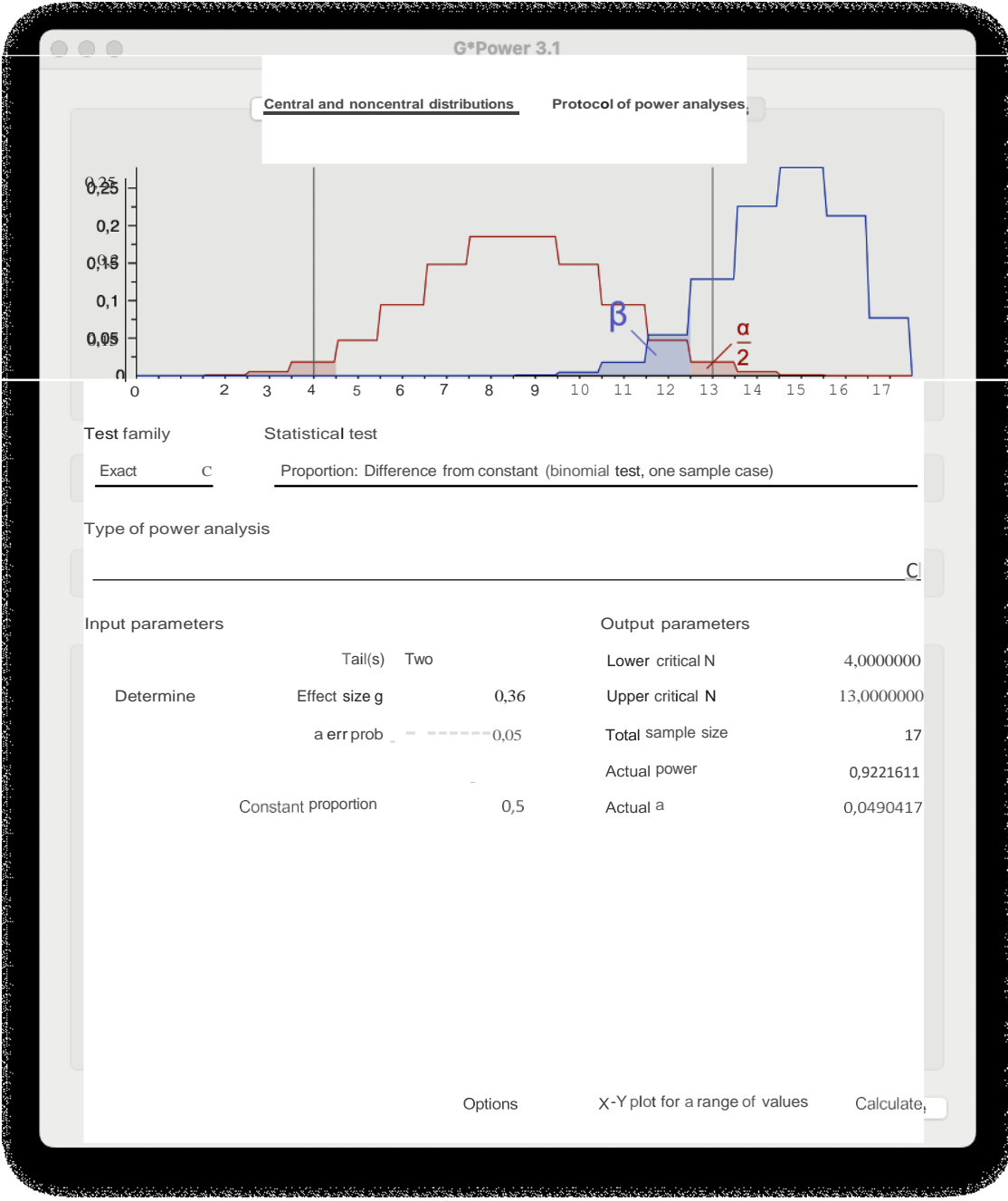
Power Analysis

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

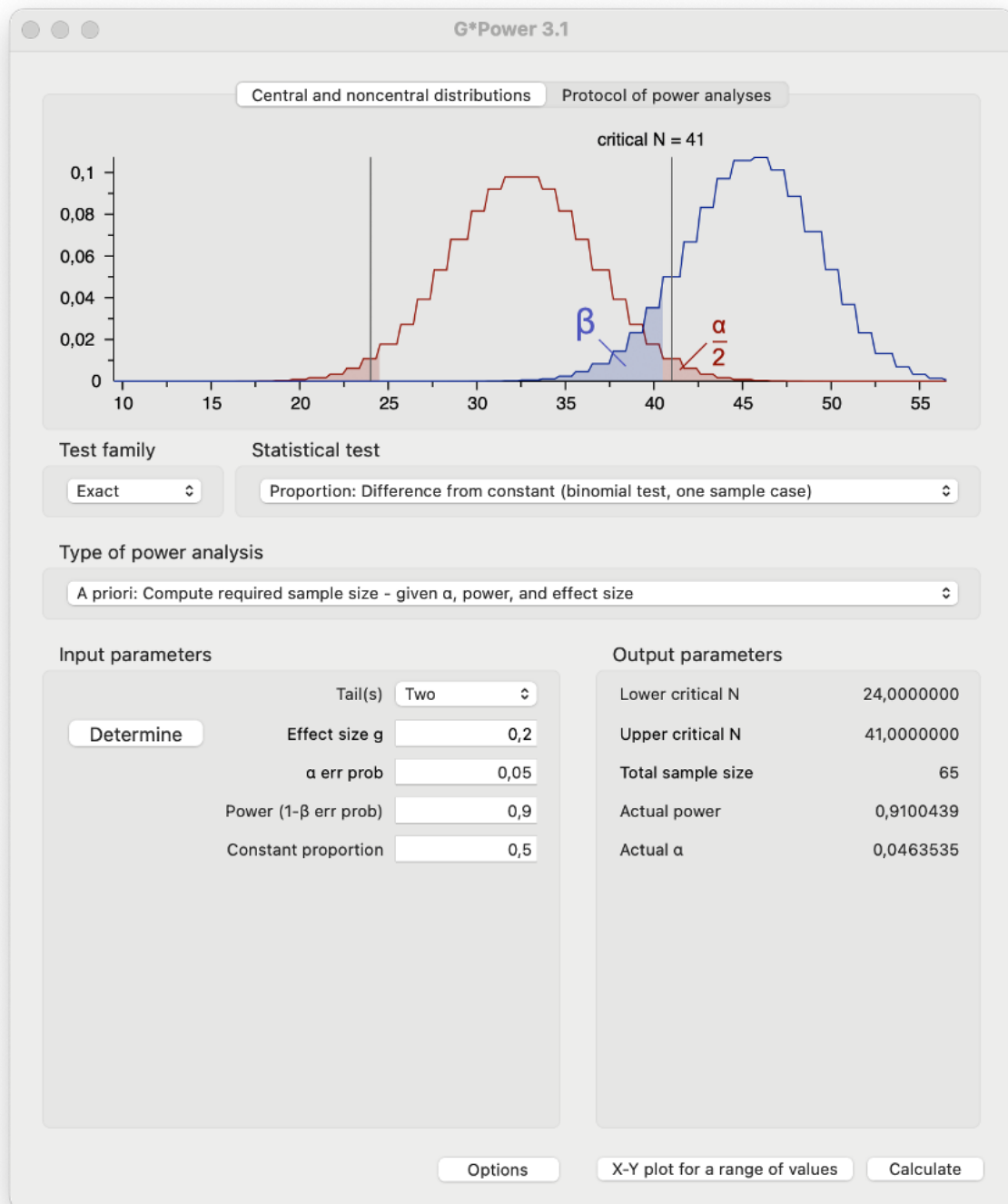
Hypotheses H1, H2 and H3

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.

Appendix A

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 skyld mer på Kristian, eller vil Kristian skyld 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

Byrne et al. (2000) Experiment 4 Replication Pre-Registration:

Project working title: Replication of Experiment 4 of Byrne et al. (2000)

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

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

Summary

study's main aim is to test if people's linguistic preferences may counteract the temporality effect in counterfactual thinking. The original study included four experimental conditions, two for coin toss scenarios and two for card scenarios. In the current replication effort, we only focus on the two conditions of the coin scenario.

Hypotheses

Common predictions across Scenario 1 and 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 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).

Study Materials

Study Outline:

In this survey, you will be presented with several scenarios where two people are making decisions. You will then be asked to complete an unfinished sentence, to determine how the situation was interpreted by you. Carefully try to form a detailed image of the situation and of the decision-makers involved. Following each scenario, you will answer two questions about the scenario revolving your impression or opinion on the situation or the decision-makers.

Scenario 1 (Heads-Tails)

Imagine two individuals (John and Michael) 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 \$1,000. However, if the two coins do not come up

the same, neither individual wins anything. John goes first and tosses a head; Michael goes next and tosses a tail. Thus, the outcome is that neither individual wins anything.

There were two ways that John and Michael could have won \$1,000. Which of these alternatives comes more readily to mind?

- John tossing tails
- Michael tossing heads

Who would you predict will experience more guilt—Jones or Michael?

- John
- Michael

Will Jones blame Michael more or will Michael blame Jones more for their failure to win \$1,000?

- John blames Michael more
- Michael blames John more

Scenario 2 (Tails-Heads)

Imagine two individuals (John and Michael) 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 \$1,000. However, if the two coins do not come up the same, neither individual wins anything. John goes first and tosses a tail; Michael goes next and tosses a head. Thus, the outcome is that neither individual wins anything.

There were two ways that John and Michael could have won \$1,000. Which of these alternatives comes more readily to mind?

- John tossing heads
- Michael tossing tails

Who would you predict will experience more guilt—John or Michael?

- John
- Michael

Will John blame Michael more or will Michael blame John more for their failure to win \$1,000?

- John blames Michael more
- Michael blames John more

Demographic questions

Thank you, you completed the main part of the survey. Please answer the following questions about yourself.

How old are you?

---- (text box here)

What is your gender?

- Female
- Male
- Non-binary
- Other

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 $g = -0.143$. 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.

Suggested Analysis

Matching the analysis choice of the original authors, we will conduct One sample binomial test to test the predictions of H1, H2 and H3.

Detailed results of Experiment 4 of Byrne et al. (2000)

Table 1

Condition	<i>n</i>	In %	Conversion	After rounding
Heads-Tails	49			
<i>Mutability</i>				
First event overall	22		10.78	11
Second event overall	64		31.36	31
<i>Guilt</i>				
First	0		0	0
Second	62		30.38	30
<i>Blame</i>				
First	38		18.62	19
Second	2		0.98	1
Tails-Heads	29			
<i>Mutability</i>				
First event overall	41		20.09	20
Second event overall	41		20.09	20
<i>Guilt</i>				
First	3		1.47	1
Second	79		38.71	39
<i>Blame</i>				
First	69		33.81	34
Second	3		1.47	1
Black-Red	50			
<i>Mutability</i>				
First event overall	31		15.19	15
Second event overall	55		26.95	27
<i>Guilt</i>				
First	2		0.98	1
Second	49		24.01	24
<i>Blame</i>				
First	43		21.07	21
Second	0		0	0
Red-Black	29			
<i>Mutability</i>				
First event overall	21		10.29	10
Second event overall	59		28.91	29
<i>Guilt</i>				
First	0		0	0
Second	76		37.24	37
<i>Blame</i>				
First	83		40.67	41
Second	0		0	0

Note. Frequency count calculations were based on percentage values reported in the original study (Byrne Et Al, 2000; Table 4, p. 272)

Table 2*Results of Experiment 4 of Byrne et al. (2000)*

Hypothesis	Dependent variables	Statistical test	Effect size with 95% CI
		One sample Binomial	Cohen`s g
	Heads-Tails		
H1	Mutability	$n = 42, k = 31, p = .002$	0.24 [0.10, 0.38]
H2	Guilt	$n = 30, k = 30, p < .001^*$	0.47 [0.40, 0.54]
H3	Blame	$n = 20, k = 19, p < .001$	-0.45 [-0.56, -0.35]
	Tails-Heads		
H1	Mutability	$n = 40, k = 20, p = 1.00$	0.00 [-0.16, 0.16]
H2	Guilt	$n = 40, k = 39, p < .001$	0.48 [0.42, 0.53]
H3	Blame	$n = 35, k = 1, p < .001$	-0.47 [-0.53, -0.41]

Note. Statistical tests were based on results reported in the original study (Byrne et al., 2000, pp. 271-272)

* To avoid error during effect size estimate and CI calculation, when the k and n are equal, we calculated the effect size estimates based on $k-1$ as the closest approximation of k .

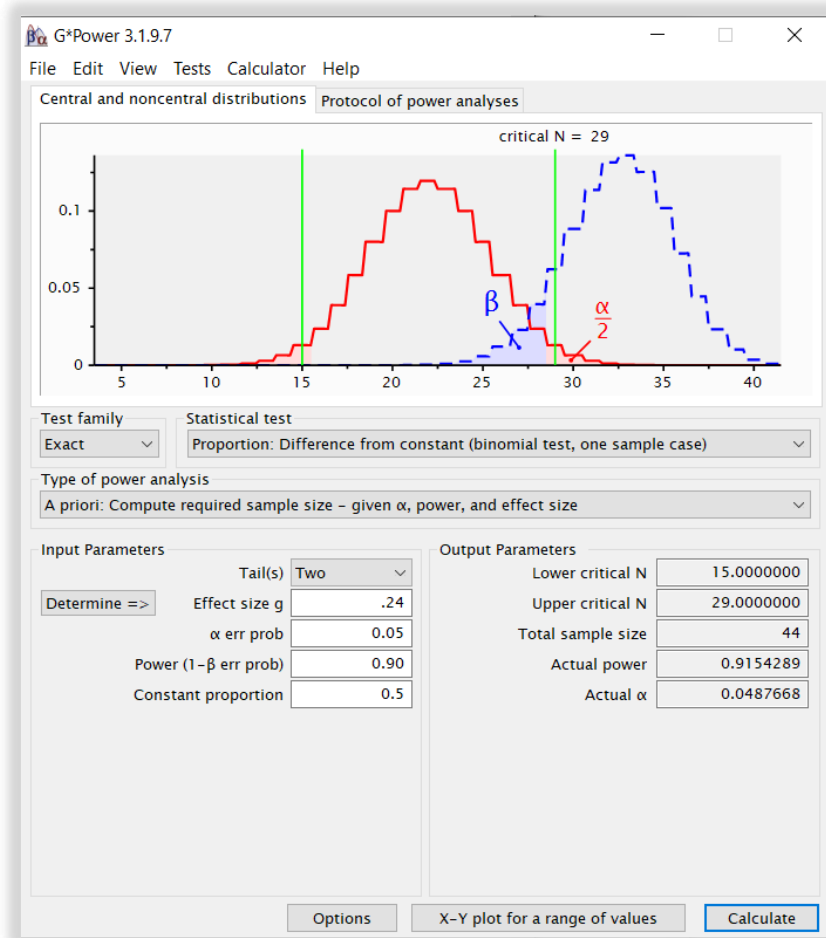
Power analysis:

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 descriptive table reported as part of the results (see Table 1 above).

Steps for power analysis

Hypothesis H1, H2 and H3:

The testing of H1, H2, and H3 involved One sample Binomial tests. The smallest effect reported across the two experimental conditions involving coin scenario (i.e., **Heads-Tails** and **Tails-Heads**) was Cohen`s $g = 0.00$. Given that planning a sample size on this effect is practically impossible, we conducted a power analysis based on the next smallest effect size among the results (i.e., Cohen's $g = 0.24$). The results of the analysis suggest a total sample size $(44 \times 2) = 88$ (see screenshot below).



Summary of power analysis:

The analysis suggests a replication sample size of 88 participants. However, we aim to achieve a sample size of 300 participants.

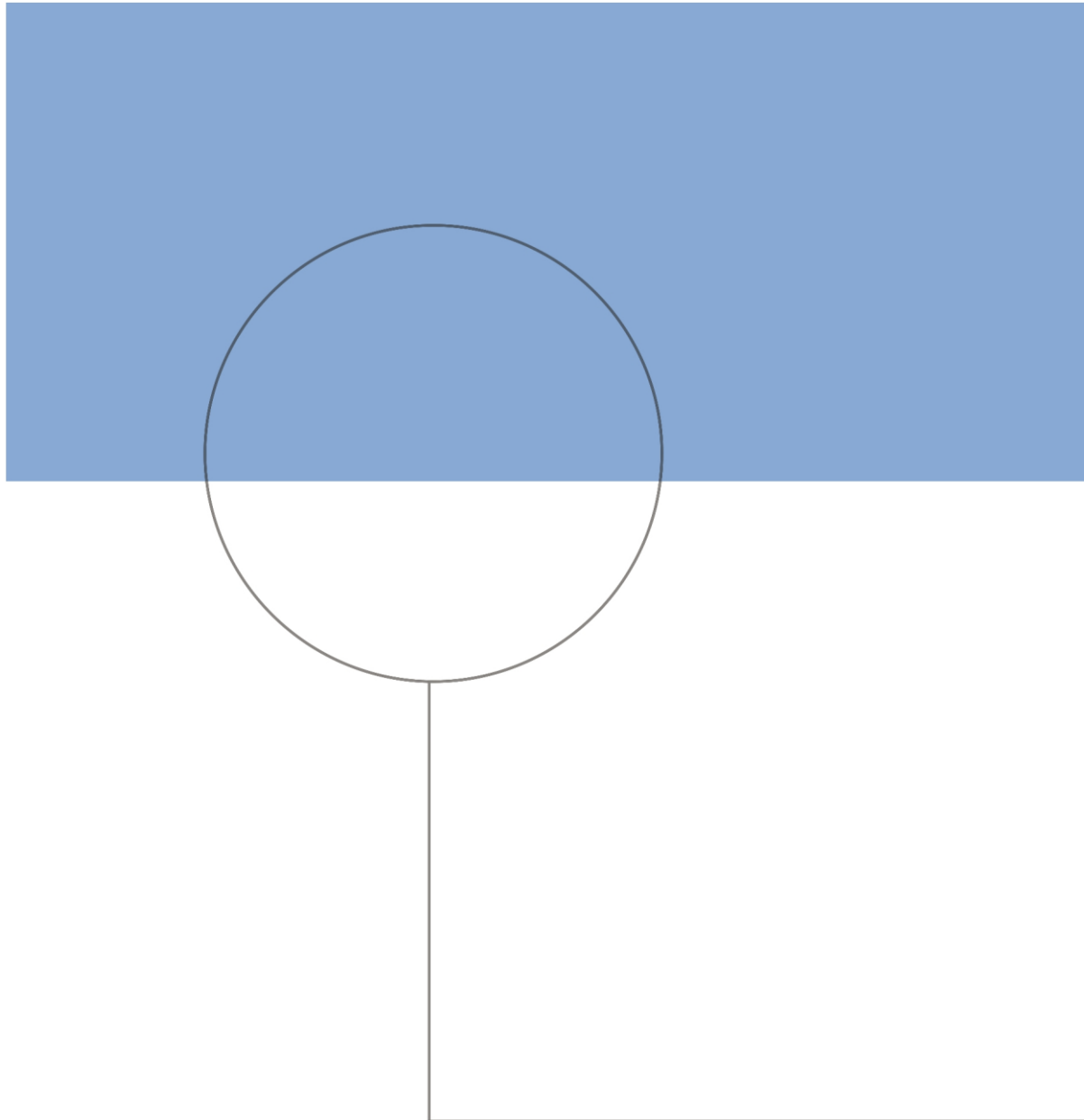
Additional points:

- Has data collection begun for this project?
No data has been collected
- What is the estimated start and end date for the project?
30.06.2023
- When, where, and what of the research output will be shared?
OSF, conference talks, and peer-reviewed journals.

Reference list:

- Byrne, R. M., Segura, S., Culhane, R., Tasso, A., & Berrocal, P. (2000).
The temporality effect in counterfactual thinking about what might have

been. *Memory & Cognition*, 28, 264-281.



&INTNU

Norwegian University of
Science and Technology