Bachelor's thesis

NTNU
Norwegian University of Science and Technology

Candidate 10011

Temporality Effect: Replicating and Extending Miller and Gunasegaram´s (1990) Coin Toss Scenario and Walsh and Byrne´s (2004) First Experiment

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

NTNU
Norwegian University of
Science and Technology

Candidate 10011

Temporality Effect: Replicating and Extending Miller and Gunasegaram´s (1990) Coin Toss Scenario and Walsh and Byrne´s (2004) First Experiment

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

Norwegian University of Science and Technology

NTNU
Norwegian University of
Science and Technology

Preface

This bachelor thesis was carried out as part of PSY2900: In-Depth Research in Psychology under the supervision of Subramanya Prasad Chandrashekar, a postdoctoral fellow in Social Psychology at the Norwegian University of Science and Technology. The research presented in this thesis was conducted as a part of the project "Revisiting the findings on the temporality effects", and the results of this study will be used by the supervisor in his ongoing research. This is one of six bachelor theses on the temporality effect this semester.

This thesis aimed to study the temporality effect by replicating Miller and Gunasegaram's (1990) coin toss scenario and Walsh and Byrne's (2004) first experiment and extending the former. I carried out online questionnaires on a Norwegian and a US sample and analyzed the data in R using Z-tests and Chi-Square tests to achieve these objectives. I hope that the findings presented in this thesis will contribute to advancing knowledge in the field of social psychology and that the results will be helpful for the scientific literature on the psychology of counterfactual thinking and the temporality effect.

I want to give special thanks to my supervisor, who answered a plethora of questions throughout the research process and provided valuable guidance and feedback that contributed to shaping the direction of this thesis and ensuring its quality and validity. Furthermore, I would like to acknowledge all the killed "darlings" and thank one particular member of the research group who supported me throughout the writing process with their knowledgeable insights. Finally, I thank you for reading and declare that this thesis is the result of the independent work of the author,

Candidate 10011

Word count approx.: 7280

Abstract

The replication crisis in psychology has decreased the trust in social psychological studies, highlighting the need for direct replications to rebuild the credibility in psychology as a scientific field. To contribute toward resolving the replication crisis, I aimed to replicate two influential studies on the temporality effect, which is an integral part of counterfactual thinking in social psychology. It describes the tendency to judge the most recent event as more mutable and ascribe the latter player more guilt and blame for the outcome in sequences of independent chance events. I conducted pre-registered, well-powered, close replications of Miller and Gunasegaram's (1990) coin toss scenario and Walsh and Byrne's (2004) first experiment through online surveys. In addition, I extended the former study by asking participants to judge the order of events and the number of counterfactual thoughts based on players' ascribed emotions. The results support former findings on the temporality effect, although the effect sizes were slightly smaller than in the original studies. Furthermore, the studies demonstrate that it is possible to replicate social psychological studies from half a century ago. I discuss the potential underlying mechanisms of the results in light of dual processing mechanisms and the TempCounterFacts model and encourage researchers to continue exploring the temporality effect.

Keywords: temporal order effect, temporality effect, counterfactual thinking, dual processing theory, default-interventionist processing, judgment and decision making, replication

Table of contents

Temporality Effect: Replicating and Extending Miller and Gunasegaram's (1990) Coin Toss Scenario and Walsh and Byrne's (2004) First Experiment

Imagine waking up, realizing it is pouring outside, and deciding to drive to work instead of your usual bike ride. On the way there has been an accident, so you spend twice as much time as regular and miss your first meeting. You might think about an alternative scenario contradicting the factual outcome: "If I had only put on a rain jacket and biked, I could have come on time…". This cognitive process of mentally undoing events by imagining a counter possibility to a fact is referred to as *counterfactual thinking* (Miller & Gunasegaram, 1990) and has, amongst others, been studied in the fields of psychology, philosophy, and artificial intelligence (Lewis, 1973; Walsh & Byrne, 2001).

Counterfactual thinking has been observed in children as young as 2 years old who are learning to use the conditional "what if…" (Epstude & Roese, 2008). By age 6-8, children across cultures develop several psychological regularities that characterize counterfactual thinking (Epstude & Roese, 2008). Counterfactual thinking can be beneficial by enabling our ability to plan and predict the future, understand outcomes, learn from past mistakes, regulate behavior, and improve decision-making (Byrne, 2015; Byrne & Girotto, 2008; Kahneman & Miller, 1986). On the other hand, it can also be dysfunctional, leading to increased emotional experiences of regret, guilt, and blame, which are associated with reduced mental health (Eisma et al., 2021; Gilovich & Medvec, 1994; Landman, 1987).

Researchers have identified several psychological regularities defining counterfactual thinking, such as the tendency to perceive exceptional events as more mutable, meaning more changeable, compared to regular events, and react to them with stronger emotions (Kahneman & Miller, 1986). If you usually bike to work and the first day you decide to drive you get stuck in traffic, you will probably feel higher levels of regret compared to delays on your default routine. Studies have also shown that humans tend to mentally undo controllable events rather than uncontrollable events (Girotto et al., 1991; Segura & McCloy, 2003), and actions rather than failures to act when given a choice (Byrne & McEleney, 2000), which further affects the experience of regret about the outcome (Wilkinson et al., 2015).

The Temporality Effect

In this thesis, the psychological regularity of counterfactual thinking I will focus on is the tendency to judge events differently based on their representation order. In an independent sequence of events, humans tend to mentally undo the most recent events, referred to as the *temporal order effect* or *temporality effect* (Byrne & McEleney, 2000; Sherman & McConnell, 1996). Findings on the temporality effect show that it persists for both negative and positive outcomes (Byrne et al., 2000), for longer sequences of three, four, six, and nine events, and selfish and mundane events (Segura et al., 2002; Segura & McCloy, 2003). The temporality effect has been subject to multiple proposed underlying mechanisms, which will be further outlined and elaborated.

Spellman (1997) formulated the *crediting causality model,* which describes how humans make causal attributions by evaluating the relative contribution of an event to an outcome (Spellman, 1997). For instance, in a coin toss scenario in which both players need to toss either heads or tails to win, the winning probability is 50% after the first player's toss due to the two possible outcomes. However, after the second player's coin toss, the winning probability is 0% or 100%, as the players have either tossed the same coin and won or tossed different coins and lost. The consequence of the observer's probability calculations is the tendency to attribute the latter player greater causal responsibility, perceive the latter event as having a greater effect on the outcome, and therefore mutate it.

Spellman's (1997) theoretical framework is similar to Miller and Gunasegaram's (1990) second prediction about the temporality effect, stating that the latter event is mutated more frequently due to greater perceived "causal potency", understood as influence over the outcome. Although both explain the tendency to mutate the latter, they differ, as Spellman (1997) focused on the person making probability calculations when attributing the second player greater causal responsibility, while Miller and Gunasegaram (1990) described the mutability of the second event as a consequence of the first event being presupposed.

A third explanation for the temporality effect is the *availability heuristic.* According to this mechanism, judgments are influenced by how readily information comes to mind, as humans assume that it must be important because it is accessible and therefore emphasize it more (Kahneman, 2011; Tversky & Kahneman, 1973). The availability can be understood in two ways with respect to the temporality effect. The first the *recency effect*, referring to the tendency to remember recent events more clearly and therefore perceiving them as more mutable (Murdock, 1962; Turvey, 2012). The effect is a type of availability heuristic, as it depends on the ease of recall (Bar-Hillel, 2001). However, the recency effect has been contradicted by Byrne et al. (2000) technical hitch study in which two players needed to draw the identical colored card to win but the game was interrupted by a technical hitch after the first player's draw. When the first player subsequently re-drew the same card color as before the interruption, the temporality effect was stronger than when the first player re-drew a different card color (Byrne et al., 2000). The technical hitch study disproved the recency effect as participants did not consistently mutate the most recent event. It supported the second understanding of the availability heuristic, according to which the descriptions participants read affects how information is stored in their mental models and is more accessible during judgment-making processes. Although many studies have described the temporality effect by referring to the second understanding of the availability heuristic, the *TempCounterFacts model* by Walsh and Byrne (2001) was the first to outline it systematically.

The TempCounterFacts model, akin to other computational models, aims to simulate complex systems such as the judgment of mutability by observing how various variables, either in combination or isolation, lead to diverse outcomes by following a set of six principles and steps (Johnson-Laird, 2010; Walsh & Byrne, 2001). According to the TempCounterFacts framework, individuals generate two mental models based on the provided information, namely the FactsModel representing the actual "bottom-up" facts and a "top-down" Possibility Model of the winning conditions (Byrne, 2015; Walsh & Byrne, 2001, 2004). These mental models adhere to six tenets of mental possibilities, such as the *principles of truth* and *parsimony* representing only true possibilities and as little explicit information as possible, despite being able to imagine a plethora of alternatives.. If necessary, implicit information can be "fleshed out" and made explicit through an elaboration process (Schaeken et al., 1996).

According to Byrne (2005), the frugal representation of explicit and implicit information in the Possibility Model is due to the complexity of information (Byrne, 2015) and the constrained capacity of the working memory, defined as a backward search through memory (Byrne, 2005). For the purpose of the discussion, I understand and will refer to the working memory as a storage unit that maintains and manipulates goal-relevant information for a limited period (Chai et al., 2018). When the number of counterfactual alternatives exceeds the capacity of the storage unit, it prioritizes information that is explicitly presented in a scenario or information that is considered more critical to the outcome.

Consistent with the fourth and fifth tenet of the TempCounterFacts model, individuals select the first part of the FactModel and look for a match in their Possibility Model (see Supplementary Materials: Table S13 & Figure S6). Thus, the first part of the facts is presupposed and serves as an anchor, leading people to mutate the latter event in scenarios comparable to the coin toss experiment described in the first replication of Miller and

Gunasegaram (1990) (Byrne, 2005). The tendency to mutate the latter event aligns with Miller and Gunasegaram's (1990) first assumption that the immutability of the first event is more potent than the mutability of the second event. However, Walsh and Byrne (2004) argue that the immutability of the first event is not "absolute" but depends on individuals' mental representations. When individuals cannot find a match to the first part of the facts in their Possibility Model, the bottom-up fact is negated and combined with the top-down expectation to generate a CounterAlternative Model. This model consists of a counterfactual alternative to the outcome and a counterfactual conditional, leading to a reversal or elimination of the temporality effect as observed in Walsh and Byrne's (2004) first experiment (see Supplementary Materials: Table S14). In other words, the TempCounterFacts model supports the existence of the second understanding of the availability heuristic by explaining how participants generate mental representations of the presented information following certain principles. Consequently, the TempCounterFacts model contradicts Miller and Gunasegaram's (1990) prediction, the crediting causality model, and the recency effect.

A final theory relevant for the latter discussion is *dual processing theory,* which has been pursued by multiple theorists who have developed their own terminology and variations. However, most models are predominantly similar (Evans & Stanovich, 2013) and propose that the first processing mode is autonomous, fast, and intuitive, akin to a person's "gut feeling, while the second mode is slow, deliberate, resource-intensive, and referred to as "hypothetical thinking" (Cabrera et al., 2015; Kahneman, 2011; Sloman, 1996). While Kahneman (2011) conceptualized them as different systems, I have chosen to use Evans and Stanovich's (2013) understanding, defining them as modes of processing along one continuum.

Sloman (1997) proposed that the two processing modes work in a *parallel-competitive* form, meaning that they are both constantly activated and that one can attenuate the other under certain circumstances (Sloman, 1996). Evans and Stanovich (2013) criticized this view, suggesting that the first mode is constantly active with the opportunity for the second mode to intervene when stimuli are unusual or conflicting (see Supplementary Materials: Figure S5). They called this *default-interventionist processing* (Evans & Stanovich, 2013). Their view is supported by studies and theories about the limited capacity of working memory and constant effort to minimize the use of cognitive resources (Evans, 2008; Gillard et al., 2009).

Replicating the Temporality Effect

There have been multiple extensions and variations of Miller and Gunasegaram's (1990) and Walsh and Byrne's (2004) studies, but no direct replications. Replication is the practice of repeating a study's procedure with a different sample to observe whether their findings re-occur (Nosek & Errington, 2020). Reproduced findings strengthen and verify the original study, while unsuccessful attempts weaken or question it. In other words, it is essential to replicate a study using different samples, scenarios, and methods to increase confidence in the results (National Academies of Sciences et al., 2019).

It is of particular interest to replicate social psychological research considering the ongoing replication crisis that blossomed in 2011 when a study using standard psychological methods and over 1000 participants confirmed the hypothesis that humans can "feel the future" (Bem, 2011). The study triggered discussions on threats to science, such as low statistical power, a lack of quality control, p-hacking, replicability, generalization, and publication bias (Camerer et al., 2018; Munafò et al., 2017; Simmons et al., 2011). Not being able to replicate findings breaches the fundamental research purpose of creating trustworthy and replicable knowledge and can lead to mistrust in findings and ineffective interventions (Diaba-Nuhoho & Amponsah-Offeh, 2021). However, because single effect sizes are often not replicated, calculating a *continuously cumulating meta-analytical index* based on several replication studies can support inferences about the existence of a phenomenon (Braver et al., 2014; Patil et al., 2016). Furthermore, replications can be evaluated according to LeBel et al. (2019) framework, which compares the original and replicated studies (LeBel et al., 2019) (see Supplementary Materials: Figure S1 and S2).

This thesis specifically focuses on replicating Miller and Gunasegaram's (1990) and Walsh and Byrne's (2004) studies for multiple reasons. Firstly, the epistemic reason argues that replicability strengthens the understanding and insight into a phenomenon, and secondly, the pragmatic reason is that replicating develops trust in social sciences (Peels, 2019; Sikorski & Andreoletti, 2023). Additionally, as far as I know, the mentioned studies have not been replicated before although they have greatly influenced other research on the temporality effect and knowledge about psychological regularities. Finally, their simple design offers an opportunity to learn about the scientific process (Chopik et al., 2018; Jekel et al., 2020).

I have included an extension based on Miller and Gunasegaram's (1990) design, asking participants to draw inferences about players' order and counterfactual thoughts based on their attributed emotion of guilt. Including this extension is imperative as it corroborates former results, gives a more accurate measure of effect size, and increases generalizability by using a different population and scenario (Bonett, 2012).

I will discuss the findings of the two replications and extension from two frameworks: dual processing theory and the TempCounterFacts model. In alignment with dual processing theory, I will propose that humans mutate the latter event as they default to mode 1 due to working memory constraints. It is reversed, when accentuated information activates the resource intensive mode 2. Alternatively, the findings are the result of the six principles and steps of the TempCounterFacts model, according to which the temporality effect occurs as the first event is presupposed but can be reversed by making alternatives to the first event available by varying the description of the winning condition. I suggest that the observed discrepancy in the judgment of mutability and emotions is caused by the study design rather than distinct underlying mechanisms.

Disclosures

Pre-registration

All three experiments outlined in this thesis align with the open science framework and have been pre-registered and uploaded to OSF (Study 1: https://osf.io/gpcqd; Study 2: https://osf.io/dwa7b; Study 3: https://osf.io/y6g9c). This is to take responsibility for my actions and avoid becoming a Feynman fool: "The first principle is not to fool yourself – and you are the easiest person to fool" (Feynman & Sackett, 1985). The pre-registration includes a summary of the study, the hypothesis, the planned method and analysis section, a power analysis to calculate the sample size, and the inclusion and exclusion criteria of the sample.

Data, Materials, Online Resources, Reporting, and Ethical Approvals

All resources and materials used in the three experiments can be found online, along with the reasoning for the sample size, data exclusion, dependent and independent variables, and other measures. (https://osf.io/a69ne/). The Norwegian Center for Research Ethics (Sikt) has approved all studies for scientific data (see Supplementary Materials: Screenshot S1).

Replication of Miller and Gunasegaram's Coin Toss Experiment (1990)

Miller and Gunasegaram hypothesized that the most recent event in a sequence of two independent events is considered more mutable than the first, inclining participants to ascribe the latter player stronger emotions of guilt and blame. I have conducted a very close replication of their coin toss scenario (see Supplementary Materials: Figure S1) (LeBel et al., 2019).

Method

Power Considerations and Participants

Based on the power analysis (see Supplementary Materials: Screenshot S2 & S3), I aimed to recruit 200 participants by sending a digital participation link on social media and distributing QR codes on the NTNU campus Dragvoll (see Supplementary Materials for details). 211 participants were recruited (*Mage* = 28.51; *SDage = 13*.68; female: 138, male: 72, other: 1). This replication was in Norwegian and based on a Norwegian sample, while the other two were in English on a US sample.

Design and Procedure

The method employed in the replication is as outlined in the original study by Miller and Gunasegaram (1990) (see Supplementary Materials for Replication Evaluation). The study participants were informed about what the study entailed for them, who was responsible, and that participation is voluntary and anonymous before seeking their informed consent to participate. The procedure was similar for all three experiments in this theses.

Participants read the following scenario and were asked to focus on the decision-makers: "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."

Participants were then asked to answer three questions based on their understanding of the scenario. The questions tested for participant's judgment of mutability, guilt, and blame (see Supplementary Materials or pre-registration for details: https://osf.io/gpcqd).

Analysis

Although the original study did not include a description of the data analysis, it can be assumed that the authors conducted three one-sample Z-tests for each prediction. Data for all three experiments outlined in the thesis were analyzed in R, an open-source software according to the principles of open science. Missing data were excluded from the analysis.

Results and Discussion

One-sample Z-test analysis confirmed all three hypotheses and supported the original findings by Miller and Gunasegaram (1990) (see Table T1 and Supplementary Materials: Tables S3 and S4). In a sequence of two independent chance events, participants showed a great tendency to mutate the second event (76.30%) and not the first event (23.69%), $z = 7.64$, $p < .001$; Cohen's $g = 0.26$, 95% CI [0.20, 0.32]. Furthermore, 93.37% of participants attributed more guilt to the second player and 6.63% to the first player, $z = 12.60$, $p < .001$, Cohen's $g = 0.43$, 95% CI [0.40, 0.47]. Thirdly, 91.94% of participants assumed that the first player blamed the second player more than the second player blamed the first player (8.06%) for the outcome; z=12.19, $p < .001$, Cohen's $g = 0.42$, 95% CI [0.38, 0.46].

The replicated study showed slightly weaker findings than the original study for all variables, although the difference was minimal (Supplementary Materials: Tables S1-S4, Figure S3). Furthermore, the replicated effect size for all tests showed a consistent signal, except for the variable mutability, which had a weaker inconsistent effect size (see Table T1) (LeBel et al., 2019). The overlap in effect size between the original and replicated study for guilt and blame strengthens former findings, while the weaker inconsistent pattern for mutability weakens them. Despite this, the range of the confidence intervals was smaller in the replication, indicating higher accuracy. The inconsistent effect size for mutability and the overall smaller confidence interval are common challenges in replication studies, which is why the results are still of importance when considering their continuously cumulating meta-analysis (Braver et al., 2014; Patil et al., 2016).

Table T1

Comparison of the results of the original study and the replication of Miller and Gunasegaram (1990)

	Original results		Replication results		Replication summary
Dependent variable	Statistical test	Cohen's g 95% CI	Statistical test	Cohen's g 95% CI	
Mutability	$k = 78$, $n = 88$, $z = 7.25$, $p < .001$	0.39 [0.32, 0.45]	$k = 161$, $n = 211$, $z = 7.64$, $p < .001$	0.26 [0.20, 0.32]	Signal – Inconsistent (weaker)
Guilt	$k = 76$, $n = 88$, $z = 6.82$, $p < .001$	0.36 [0.29, 0.44]	$k = 197$, $n = 211$, $z = 12.60$, $p < .001$	0.43 [0.40, 0.47]	Signal - Consistent
Blame	$k = 81$, $n = 88$, $z = 7.89$, $p < .001$	0.42 [0.36, 0.48]	$k = 194$, $n = 211$, $z = 12.19$, $p < .001$	0.42 [0.38, 0.46]	Signal - Consistent

Replication of Walsh and Byrne's (2004) first experiment

Walsh and Byrne predicted that they could reverse the temporality effect, making the first event in a sequence of two more mutable by varying the descriptions of the winning condition while the facts remain . They hypothesized that, given a negative outcome based on a sequence of two independent chance events, participants would judge the second player to experience more guilt and be blamed more for the outcome than the first player. The very close replication (LeBel et al., 2019) included an exploratory analysis comparing the response pattern across the two experimental conditions, the disjunction and conjunction group.

Method

Power Considerations and Participants

The effect size was calculated to run three separate power analyses (see Supplementary Materials: Screenshot S4), leading to the decision to aim for 312 participants, 156 participants per condition (see Supplementary Materials for details). A total of 325 participants were recruited (aged 18-77; *Mage* = 39.23; *SDage* = 14.05; female: 140, male, 179, other: 6). The final sample in the analysis included 163 participants in the control and 162 participants in the experimental condition.

Design and Procedure

The design, procedures, and materials used in this replication match Walsh and Byrne (2004) original card game scenario (see Supplementary Materials for Replication Evaluation). Participants were randomly assigned one of two conditions reading different formulations of the same winning condition and asked to focus on the color of the cards and the order of draws. Participants in the experimental group read the following disjunctive description: "If one or the other but not both pick a card from a red suit, each individual wins £1,000." Participants in the control group read a conjunctive description: "If the two cards they pick are of a different color (i.e., one from a black suit and one from a red suit), each individual wins £1,000." Both conditions were given the fact: "John picked black and Michael picked black and they both lost."

After reading, participants were asked to answer three questions in a fixed order and were implored not to revise their answers. The first question was a counterfactual mutation task: "There were two ways that John and Michael could have won. Which of these alternatives comes more readily to mind?". Participants had the choice between "John picks red" and "Michael picks red". The second and third questions match the judgment of guilt and blame in Miller and Gunasegaram's study (1990), as participants were asked to judge who they predicted would experience more guilt and who would blame the other more for the outcome (see Supplementary Materials or pre-registration for details: https://osf.io/dwa7b).

Analysis

The three hypotheses were tested using one-proportion Z-tests to understand the temporality effect within each condition. The conditions were then compared in the exploratory analysis through Chi-squared tests.

Results and Discussion

In the disjunction condition, the first event was mutated more often (70.99%) than the second event (29.01%), reversing the temporality effect, $z = 5.34$, $p < .001$; Cohen's $g = -0.21$ [-0.28, -0.14], while in the conjunction condition, participants were almost equally likely to mutate the first and second event which eliminated the temporality effect (53.99% vs. 46.01%), $z = 1.02$, $p = .309$; Cohen's $g = -0.04$ [-0.12, 0.04]. However, there was no signal in the conjunction condition when compared with the original study's effect size. The difference between the conditions was reliable ($\chi 2(1, 325) = 10.01$, $p = .002$, Cohen's $h = -0.35$, 95% CI [-0.57, -0.14]) (see Supplementary Materials: Tables S8, S9, and S10).

Moreover, participants in either condition were more likely to attribute the second player guilt (Disjunction: 62.34% & Conjunction: 71.78%) than the first player (Disjunction: 37.65 & Conjunction: 28.22%). This tendency was stronger in the conjunction condition ($z = 5.56$, $p < .001$; Cohen's $g = 0.22$, 95% CI [0.15, 0.29]) compared to the disjunction condition ($z = 3.14$, $p = .002$; Cohen's $g = 0.12$ [0.05, 0.2]), although the difference was nonconclusive ($\chi 2(1, 325) = 3.27$, $p = .070$; Cohen's $h = -0.2$, 95% CI [-0.42, 0.20]).

The results also show a tendency to assume that the first player blamed the second player more often (Disjunction: 60.49% vs. Conjunction: 73.62%) than the second player blamed the first player (Disjunction: 39.51% vs. Conjunction: 26.38%) across conditions. This effect was more robust in the conjunction condition ($z = 6.03$, $p < .001$; Cohen's $g = 0.24$, 95% CI [0.17, 0.30]) than in the disjunction condition ($z = 2.67$, $p = .008$; Cohen's $g = 0.10$, 95% CI [0.03, 0.18]) ($\chi 2(1, 325) = 6.34$, $p = .012$; Cohen's $h = -0.28$, 95% CI [-0.5, -0.06]).

Overall, there were large variations in Z-score, Chi-square score, and effect size between the original (see Supplementary Materials: Table S5-S7) and replicated study (see Supplementary Materials: Table S8-S10), questioning the strength of the study's reproducibility (see Table T2). While the signal was consistent for mutability, it was inconsistent and weaker for the variables guilt and blame in both the Z-test and Chi-square test (see Supplementary Materials: Figure S4). However, the overall narrower confidence interval in the replication indicates a higher accuracy of the results.

Table T2

Summary of findings of the original study versus replication of Walsh and Byrne (2004).

		Original results		Replication results		Replication summary
	Experimental condition	Statistical test	Cohen's g (h) 95% CI	Statistical test	Cohen's g (h) 95% CI	
Mutability	Disjunction	*k* = 23, *n*= 61, *z* = 1.92, *p* = .072	-0.12 [-0.25, 0.00]	*k* = 47, *n*=162, *z* = 5.34, *p* < .001	-0.21 [-0.28, -0.14]	Signal – Consistent
	Conjunction	*k* = 17, *n* = 32, *z* = .35, *p* = .860	0.03 [-0.15, 0.21]	*k* = 75, *n* = 163, *z* = 1.02, *p* = .309	-0.04 [-0.12, 0.04]	No signal – Consistent
	Disjunction vs. Conjunction	*k1* = 23, *n1*= 61, *k2* = 17, *n2* = 32; *χ2*(1, 93) = 2.04, *p* = .154	-0.31 [-0.74, 0.12]	*k1* = 47, *n1*= 162, *k2* = 75, *n2* = 163; *χ2*(1, 325) = 10.01, *p* = .002	-0.35 [-0.57, -0.14]	Signal – Consistent
Guilt	Disjunction	*k* = 67, *n* = 77, *z* = 6.50, *p* < .001	0.37 [0.29, 0.45]	*k* =101, *n* =162, *z* = 3.14, *p* = .002	0.12 [0.05, 0.20]	Signal – Inconsistent (weaker)
	Conjunction	*k* =34, *n* = 34, *z* = 5.83, *p* < .001	0.47 [0.41, 0.53]	*k* =117, *n* = 163, *z* = 5.56, *p* < .001	0.22 [0.15, 0.29]	Signal – Inconsistent (weaker)
	Disjunction vs. Conjunction	*k1* = 67, *n1* = 77, *k2* =34, *n2* = 34; *χ2*(1, 111) = 4.85, *p* = .028	-0.74 [-1.14, -0.33]	*k1* = 101, *n1* = 162, *k2* =117, *n2* = 163; *χ2*(1, 325) = 3.27, *p* = .070	-0.20 [-0.42, 0.20]	No signal – Inconsistent (weaker)
Blame	Disjunction	*k* = 59, *n* = 69, *z* = 5.90, *p* < .001	0.36 [0.27, 0.44]	*k* = 98, *n* =162, *z* = 2.67, *p* = .008	0.10 [0.03, 0.18]	Signal – Inconsistent (weaker)
	Conjunction	*k* = 31, *n* = 34, *z* = 4.80, *p* < .001	0.41 [0.31, 0.51]	*k* = 120, *n* = 163, *z* = 6.03, *p* < .001	0.24 [0.17, 0.30]	Signal – Inconsistent (weaker)
	Disjunction vs. Conjunction	*k1* = 59, *n1* = 69, *k2* = 31, *n2* = 34; *χ2*(1, 103) = 0.66, *p* = .415	-0.18 [-0.59, 0.23]	*k1* = 98, *n1* = 162, *k2* = 120, *n2* = 163; *χ2*(1, 325) = 6.34, *p* = .012	-0.28 [-0.50, -0.06]	Signal – Inconsistent (stronger)

Extension of the Temporality Effect

The extension aimed to corroborate the findings of Miller and Gunasegaram (1990) by reversing causality. I hypothesized that the player ascribed guilt would be judged to be last in a series of two independent chance events and to have more counterfactual thoughts.

Method

Power Considerations and Participants

The power analysis suggested a total sample size of 263 (see Supplementary Materials: Screenshot S5). A total of 324 participants were recruited, 162 per variation (aged 18-77; *Mage* = 39.51; *SDage* = 13.55; female: 140, male: 179, other: 5).

Design and Procedure

The scenario and questions in the extension were adapted from Miller and Gunasegaram's study (1990), exchanging the coin toss with a simplified version of the

Ellsberg Urn experiment to increase generalizability. After randomization, participants in both variations were given the scenario: "Imagine two individuals who are offered the following very attractive proposition. Each individual is asked to pick a ball from an urn with red and yellow balls. If both players pick the same color, each individual wins £1,000. However, if they pick different colors, neither individual wins anything. The first player picks a red ball and the second picks a yellow ball and they both lose."

Participants assigned to the first variation of the scenario read the following additional prompt: "After the game, Lisa experiences more guilt than Ella." and in the second variation: "After the game, Ella experiences more guilt than Lisa". All participants were asked to answer two questions. The first question asked participants to decide in what order Ella and Lisa picked the balls from the urn, given that either Lisa or Ella experienced more guilt. The second question asked who of the two players would have more counterfactual thoughts (see Supplementary Materials or pre-registration for details: https://osf.io/y6g9c).

Analyses

The hypotheses were analyzed using a one-proportion Z-test.

Results and Discussion

The one-proportion Z-test indicated that participants in both variations tended to assume that the player who experienced more guilt was last (see Supplementary Material: Table S11 and S12). In the first variation, Lisa experienced more guilt and was more often placed last (84.57%) compared to Ella, who experienced less guilt (15.43%); k = 137, n=162, z = 8.80, p < .001; Cohen's g = 0.35, 95% CI [0.29, 0.40]. In the second condition, Ella was ascribed more guilt and more often assumed to be last (93,83%) and not Lisa (6.17%), k = 152, n = 162, z = 11.16, p < .001, Cohen's g = 0.44, 95% CI [0.40, 0.48].

Furthermore, participants more often assumed that the player experiencing guilt would have more counterfactual thoughts (Variation 1: 77.16% & Variation 2: 83.33%)

compared with the player feeling less guilt (Variation 1: 22.84% & Variation 2: 16.67%) (Variation 1: z = 6.91, p < .001, Cohen's g = 0.27, 95% CI [0.21, 0.34] & Variation 2: z = 8.49, p < .001, Cohen's g = 0.33, 95% CI [0.28, 0.39]).

In general, the effect sizes obtained for all measures in this study were small to moderate, which implies limited practical applicability (Ferguson, 2015). However, according to Braver et al., (2014), the findings contribute to the cumulative body of research literature on the topic, suggesting that they are of importance and describe an existing psychological regularity.

General Discussion

The thesis aimed to conduct close direct replications of two fundamental studies on the temporality effect and extend them. All three experiments supported former findings on the temporality effect and will be discussed in light of two frameworks. I have chosen first to approach the findings through dual processing theory to get a new perspective and deeper understanding of the results. I will then discuss them by referring to the TempCounterFacts model, which was formulated with the intention of exlpaining experiments on the temporality effect.

The results of the present studies demonstrate small effect sizes, raising concerns about their practical applicability (see Supplementary Materials: Figures S3 & S4). However, Schäfer and Schwarz (2019) have argued that previously reported effect sizes are likely to be inflated due to sample biases, study design, and analysis limitations, which may reduce their explanatory power (Schäfer & Schwarz, 2019). Baguley (2009) proposed a solution to address this issue by suggesting that instead of interpreting effect size using standardized measures such as Cohen's benchmark (Cohen, 1988) or global conventional benchmarks, effect sizes should be reported in a simple, unstandardized form, and include confidence intervals to estimate the accuracy of the findings (Baguley, 2009). Therefore, even though the replicated effect sizes in our replications were small, they should not be disregarded as they may have cumulative importance (Braver et al., 2014).

The Occurrence of the Temporality Effect

The present study successfully replicated Miller and Gunasegaram's (1990) coin toss scenario, providing further support for the occurrence of the temporality effect. The findings suggest that individuals tend to judge the latter event as more mutable.

Drawing upon dual processing theory, humans default to mode 1 processing due to the constrained ability of working memory to store and manipulate information (Evan & Stanovich, 2013). Therefore, the information participants were presented last was more accessible during memory retrieval and more likely to be mutated. This interpretation is consistent with the recency effect which can be described as part of mode 1 as both process information automatically and demand little effort. Although my second replication of Walsh and Byrne (2004) contradicts the recency effect as participants do not more frequently undo the last event, it does not exclude the possibility that we default to mutating the latter event when the capacity of working memory is exceeded. The interpretation of the temporality effect as default mode 1 is further supported by time-constrained studies in which the temporality effect remained (Atkinson et al., 2009). If the effect had changed under time pressure, it would indicate that it relied on mechanisms other than default mode 1 processing and the recency effect.

A second approach that may explain the underlying mechanisms of the occurrence of the temporality effect in the replication of Miller and Gunasegaram (1990) is the TempCounterFact model and its six tenets of mental possibilities (Walsh & Byrne, 2004). Building on the first three tenets about how information is mentally represented, participant's employ the fourth and fifth tenets as they use the first part of the facts when searching for a match in the Possibility Model of the winning conditions (Walsh & Byrne, 2001). The winning condition required both players to toss the same coin, influencing participant's Possibility Model to represent either two heads or two tails. The first part of their FactModel about Jonas tossing heads matched with the first part of their Possibility Model representing two heads while the second part of the facts about Kristian tossing tails did not. Consequently, participant's might have generated a CounterAlternative model with the counter alternative conditional: "Kristian tossed heads", and mutated the latter event. Thus, the TempCounterFacts model suggests that the underlying mechanism of the temporality effect is the generation of mental models in which the first event is kept relatively immutable while the second event is mutated to align with the winning condition (Walsh & Byrne, 2004).

The Reversal of the Temporality Effect

In the disjunction condition of Walsh and Byrne's (2004) first experiment, the temporality effect was reversed as participants were more likely to think of counteralternatives to the first event. The card game differed from the coin toss scenario, as the winning condition required participants to draw different cards rather than identical coins and the disjunctive description highlighted the role of red cards to win. According to the default-interventionist view, emphasized, unusual, or ambiguous information can activate mode 2 by affecting what information is maintained more accessible in individuals working memories (Evans & Stanovich, 2013). This way, the reversal of the temporality effect might be caused by mode 2 and depend on individual's ability to activate their working memory. This mechanism could be explored using a time constraint or testing participants with depleted cognitive resources. I predict that these conditions affect individuals' ability to mobilize their working memory and activate mode 2, causing them to rely on default mode 1 and exhibit the standard temporality effect.

The reversal of the temporality effect in the disjunction condition can also be interpreted through the TempCounterFacts model and its tenets. In addition to the five described under the occurrence of the temporality effect, the reversal might depend on a sixth principle, namely the explicit representation of true possibilities while other possibilities remain implicit with the potential to be fleshed out (Walsh & Byrne, 2004). In my replication, the accentuation of red cards in the disjunction condition might have affected participants' Possibility Model to represent red cards explicitly and black cards implicitly, thereby making them unable to find a match for the first part of the facts: "John picked black". Participants were consequently forced to negate the first fact and generate a counterfactual alternative: "If John had picked red", making them aware of alternatives to the first event and causing a reversal of the temporality effect.

The sixth tenet can be supplemented by the availability heuristic as explicit representations can affect what information is more accessible during memory retrieval and used when judging event's mutability. This interpretation is supported by studies that have varied the descriptions of the winning conditions (Byrne & McEleney, 2000; Walsh & Byrne, 2004) or asked participants to actively think about alternatives to the first event (Wellst et al., 1987). These studies and my findings support Walsh and Byrne's (2004) assumption that the underlying mechanism of the temporality effect is the presupposition of the first event but that the descriptions of the winning conditions can alter individual's mental models of them, affecting what information is available during memory retrieval, and ultimately the judgment of mutability.

The Elimination of the Temporality Effect

Participants in the conjunction condition of the replication of Walsh and Byrne (2004) were equally likely to imagine counterfactual alternatives to the former and latter event, eliminating the temporality effect. Considering dual processing theory, the description emphasizing either card color in winning might have affected the activation of mode 2 in some participant's, while others defaulted to mode 1. According to the TempCounterFacts model, the elimination could have been caused by half of the participants representing red and the other half representing black cards in their Possibility Model. Similarly, Byrne (2000) argued that the elimination was caused by two opposing tendencies: half of the participants exhibited the temporality effect due to a presupposition of the first event, while the other half generated a counterfactual alternative to it and mutated the first event. In any case, what processing mode was employed, card color was explicitly represented, and which opposing tendency a participant relied on must have been random, as the temporality effect was eliminated across participants. Because it might have been random, the observed elimination depended on a large, representative sample to capture the variance in reasoning, as was the case in the replication.

The replicated elimination also supports Walsh and Byrne's (2004) reasoning that the temporality effect depends on the winning condition requiring players to toss the same coin or pick the same card to win. My finding further supports their suggestion as the judgment of mutability in the replication of Miller and Gunasegaram (1990) was significant and had a moderate effect size, while it was insignificant and had a small effect size in the original and replicated conjunction condition in the card game experiment.

Discrepancies in Judgment-Making

The replication of Walsh and Byrne (2004) found a discrepancy between the judgment of mutability and emotions. Participants were more likely to mentally undo the former event but ascribe the latter player more guilt and blame for the outcome. The judgment of emotions might have been caused by the underlying mechanisms comparable to the occurrence of the temporality effect, while the judgment of mutability by mechanisms causing its reversal. The possibility that the judgments occur independently of each other but through corresponding mechanisms contradicts Miller and Gunasegaram's (1990) assumption that mutability causes guilt, which, however, is weakened by my extension in which the player ascribed guilt was placed last in order independently of the mutability of events. The interpretation of similar underlying mechanisms is further challenged by Meehan and Byrne's (2005) study which found age-related differences in the development of the judgment of mutability and emotions. In their study, 8-year-olds judged comparable to adults, while 6-year-olds differed in their judgment of emotions as they ascribed both player's guilt and blame equally often (Meehan & Byrne, 2005). The discrepancy could be caused by the ongoing maturation of 6-year-old's working memory (Chai et al., 2018). Considering age-related differences in development, 6-year-olds with more developed working memories might have ascribed the latter player more guilt while still maturing 6-year-olds the former player. It is also possible that the findings were caused by chance or by the unrepresentative, small sample size including only 62 children (Meehan & Byrne, 2005). Studies have shown that working memories ability to manipulate information continues to develop until age 15 (Crone et al., 2006; Gathercole et al., 2004) and declines with age (McNab et al., 2015), which is why it would be interesting to replicate the study including a sample ranging from 6-100-year-olds to further explore the connection between judgments and working memory.

Although Meehan and Byrne's (2005) study indicates that the judgment of mutability and emotions develops at different ages, they do not exclude the possibility of them being caused by equal underlying mechanism. For instance, the discrepancy in the replication of Walsh and Byrne (2004) might have been caused by the study design rather than different mechanisms. As discussed, the reversed judgment of mutability might have been caused by the description of the winning condition affecting individual's mental representations. I predict that emphasizing emotions rather than the card color could potentially reverse this discrepancy. Including a sentence that accentuates the first player's guilt, such as: "Upon learning about the outcome of events, the former player looked shamefully to the ground", could show a tendency to mutate the latter event while ascribing the former player more guilt and blame for the outcome. Therefore, I suggest that although the judgment of mutability, guilt, and blame might develop at different ages and do not consistently co-occur, they can be caused by equivalent underlying mechanisms.

Complimenting Underlying Mechanisms

Thus far, I have discussed the findings of the replications and extension from the perspective of the dual processing theory and the TempCounterFacts model. The distinct theoretical frameworks could be regarded as complementary interpretations for the occurrence, reversal, and elimination of the temporality effect. Specifically, dual processing theory can be perceived as an explanation for how information is processed on a superficial level, while the TempCounterFacts model can elucidate the mental representation of information. Additionally, the availability heuristic might operate within the Possibility Model by making information more accessible during memory retrieval, affecting judgments.

The implication of becoming aware of psychological regularities such as the temporality effect is our ability to improve the quality of our judgments. An individual's ability to utilize both processing modes in the right circumstance can inspire us to design our schedules to make judgments when we have sufficient cognitive resources (Evan & Stanovich, 2013). Likewise, awareness of how we generate mental representations and how these affect judgments can reduce biased judgments of mutability, guilt, and blame.

Limitations and Future Directions

In replicating Miller and Gunasegaram's (1990) and Walsh and Byrne's (2004) studies, I aimed to repeat their strengths, such as the simple and intuitive study design. Additionally, I worked on finding a balance between directly replicating and improving their limitations. For instance, both original studies did not report or incorrectly reported information about the sample, methods, or analysis. To address these limitations, I re-ran the analysis and focused on transparency throughout the process by including information on all decisions and procedures for all studies (see Supplementary Materials and pre-registrations). I also made up for Walsh and Byrne's lack of randomization by including an equal number of participants in each condition.

However, I did repeat a weakness of Miller and Gunasegaram's (1990) study, which had been corrected in Walsh and Byrne's (2004) study, namely giving participants the option "neither" or "both" when judging mutability, guilt, and blame. These options are interesting, considering Walsh and Byrne's (2004) finding that in the disjunction condition, equally many participants mutated the second and both events, questioning their interpretation of the reversal of the temporality effect (see Supplementary Materials: Table S5). Qualitative feedback from participants in the first replication revealed that the lack of answer options was perceived as confusing, given that it is a game of chance. Consequently, some chose the option they suspected aligned with the researcher's intention. Although it could be argued that forcing participants to choose either answer option demonstrates a psychological regularity, future replications or extensions should offer the complete set of answer options, including "first", "second", "both", or "neither" event, for all questions about mutability and judgment of guilt and blame.

I also repeated limitations of other research on the temporality effect, such as the lack of control for individual differences (Stanovich, 1999; Stanovich & West, 2000), culture (Chen et al., 2006), age (Meehan & Byrne, 2005), method, and stimuli (Atkinson et al., 2009). All of these have shown to impact participant's judgments. For instance, the studies presented in this thesis might be affected by cultural or linguistic biases that influence how individuals think about or describe events. Consequently, future studies should consider studying non-WEIRD cultures to test whether the temporality effect is caused by human nature or culture. Additionally, prospective replications should include different age groups, as age is associated with working memory functioning which might affect judgment of mutability and emotions (Meehan & Byrne, 2005; Verhaegen et al., 2019).

My attempt to explain the underlying mechanisms of the findings through a, to my knowledge new perspective of two processing mechanisms, can be criticized for being vague and far-fetched. Nonetheless, I perceive it to be crucial to discuss the findings from different frameworks as it may highlight distinct aspects of the same psychological regularity. Although the TempCounterFacts model has stronger explanatory power, which is sensible, considering that it was formulated for this exact purpose, the mechanisms may be too complex for a psychological regularity like the temporality effect. While a too strong simplification could leave important information out, it would benefit from being revised by Occams razor (Epstein, 1984). Niininen (2013) introduced a metric describing the trade-off between accuracy and simplicity in scientific models in a non-linear manner, as simpler does not necessarily mean higher accuracy, and complexity does not mean inaccuracy (Niininen, 2013). This means that this thesis's proposed underlying mechanisms of the temporality effect could still have predictive power as it has been replicated.

The replication and extension results are important, as they show that not all classic social psychology studies are unreplicable. However, it is crucial to recognize that although this thesis has taken an noteworthy step in corroborating the research literature on the temporality effect and attempted to identify underlying mechanisms, there are many unexplored possibilities for studies and discussions. To continue exploring underlying mechanisms and increase generalizability, I encourage researchers to replicate and extend using mixed-design and combine questionnaires with neuroscientific techniques such as EEG and fMRI, passive and active participation, and controlling for demographics (Cacioppo & Petty, 1982; Nisbett et al., 2001; Stanovich & West, 2000).

Conclusion

The present thesis replicated Miller and Gunasegaram's (1990) coin toss scenario, Walsh and Byrnes's (2004) first experiment and extended the former. The findings have provided further evidence for the occurrence, reversal, and elimination of the temporality effect. I have discussed the findings in light of two frameworks, dual processing theory and the TempCounterFacts model. According to the former, the latter event is mutated due to constraints of working memory making individuals rely on the automatic and fast mode 1. When information is emphasized, the resource-intensive and slow mode 2 might be activated, reversing the temporality effect. Alternatively, the findings could be caused by the six tenets and steps governing the TempCounterFacts model, stating that the first event is presupposed but can be altered by providing descriptions that change what information is explicitly represented and accessible during memory recall. The observed elimination might be caused by opposing tendencies to mutate either event or explicitly represent either color. I have suggested that similar mechanisms might cause the discrepancy between the judgment of mutability and guilt and blame despite developing at different ages. I encourage researchers to continue replicating and extending experiments on the temporality effect using minor variations. Such studies will be valuable to deepen our understanding of underlying mechanisms, providing a solid foundation for future research in this field and supporting the replication crisis.

References

American Psychological Association. (2020). *Publication manual of the American Psychological Association 2020: the official guide to APA style* (7th ed.). American Psychological Association.

Atkinson, L., Bell, D., & Feeney, A. (2009). The relationship between counterfactual thinking and emotional reactions to event outcomes: Does one account fit all? *Psychonomic Bulletin & Review*, *16*(4), 724–728. https://doi.org/10.3758/PBR.16.4.724

Baguley, T. (2009). Standardized or simple effect size: What should be reported? *British Journal of Psychology*, *100*(3), 603–617. https://doi.org/10.1348/000712608x377117

Bar-Hillel, M. (2001). Subjective Probability Judgments. In N. J. Smelser & P. B. Baltes (Eds.), *International Encyclopedia of the Social & Behavioral Sciences* (pp. 15247–15251). Pergamon. https://doi.org/10.1016/B0-08-043076-7/00637-9

Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, *100*(3), 407–425. https://doi.org/10.1037/a0021524

Bonett, D. G. (2012). Replication-Extension Studies. *Current Directions in Psychological Science*, *21*(6), 409–412. https://www.jstor.org/stable/44318618

Braver, S. L., Thoemmes, F. J., & Rosenthal, R. (2014). Continuously Cumulating Meta-Analysis and Replicability. *Perspectives on Psychological Science*, *9*(3), 333–342. https://doi.org/10.1177/1745691614529796

Byrne, R. M. J. (2015). Counterfactual Thought. *Annual Review of Psychology*, *67*(1), 135-157. https://doi.org/10.1146/annurev-psych-122414-033249

Byrne, R. M. J., & Girotto, V. (2008). Cognitive Processes in Counterfactual Thinking. In Markman, K. D., Klein, M. P., & Suhr, J. A., *Handbook of Imagination and Mental Simulation* (1st Ed., pp. 151-160). Psychology Press.

https://doi.org/10.4324/9780203809846

Byrne, R. M. J., & McEleney, A. (2000). Counterfactual thinking about actions and failures to act. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1318–1331. https://doi.org/10.1037/0278-7393.26.5.1318

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

Cabrera, D., Thomas, J. F., Wiswell, J. L., Walston, J. M., Anderson, J. R., Hess, E. P., & Bellolio, M. F. (2015). Accuracy of 'My Gut Feeling:' Comparing System 1 to System 2 Decision-Making for Acuity Prediction, Disposition and Diagnosis in an Academic Emergency Department. *Western Journal of Emergency Medicine*, *16*(5), 653–657. https://doi.org/10.5811/westjem.2015.5.25301

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*(1), 116–131. https://doi.org/10.1037/0022-3514.42.1.116

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B. A., Pfeiffer, T., Altmejd, A., Buttrick, N., Chan, T., Chen, Y., Forsell, E., Gampa, A., Heikensten, E., Hummer, L., Imai, T., Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*, *2*(9). https://doi.org/10.1038/s41562-018-0399-z

Chai, W. J., Abd Hamid, A. I., & Abdullah, J. M. (2018). Working Memory From the Psychological and Neurosciences Perspectives: A Review. *Frontiers in Psychology*, *9*. https://www.frontiersin.org/articles/10.3389/fpsyg.2018.00401

Chen, J., Chiu, C.-Y., Roese, N. J., Tam, K.-P., & Lau, I. Y.-M. (2006). Culture and Counterfactuals: On the Importance of Life Domains. *Journal of Cross-Cultural*

Psychology, *37*(1), 75–84. https://doi.org/10.1177/0022022105282296

Chopik, W. J., Bremner, R. H., Defever, A. M., & Keller, V. N. (2018). How (and whether) to teach undergraduates about the replication crisis in psychological science. *Teaching of Psychology*, *45*(2), 158–163. https://doi.org/10.1177/0098628318762900

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Academic Press.

Crone, E. A., Wendelken, C., Donohue, S. E., Van Leijenhorst, L., & Bunge, S. A. (2006). Neurocognitive development of the ability to manipulate information in working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(24), 9315–9320. https://doi.org/10.1073/pnas.0510088103

Diaba-Nuhoho, P., & Amponsah-Offeh, M. (2021). Reproducibility and research integrity: The role of scientists and institutions. *BMC Research Notes*, *14*(1), 451. https://doi.org/10.1186/s13104-021-05875-3

Eisma, M. C., Epstude, K., Schut, H. A. W., Stroebe, M. S., Simion, A., & Boelen, P. A. (2021). Upward and Downward Counterfactual Thought After Loss: A Multiwave Controlled Longitudinal Study. *Behavior Therapy*, *52*(3), 577–593. https://doi.org/10.1016/j.beth.2020.07.007

Epstein, R. (1984). The Principle of Parsimony and Some Applications in Psychology A Principle of Parsimony. *Journal of Mind and Behavior*, *5*(2) 119-130. https://www.jstor.org/stable/43853318

Epstude, K., & Roese, N. J. (2008). The Functional Theory of Counterfactual Thinking. *Personality and Social Psychology Review : An Official Journal of the Society for Personality and Social Psychology, Inc*, *12*(2), 168–192. https://doi.org/10.1177/1088868308316091

Evans, J. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, *59*(1), 255–278.

https://doi.org/10.1146/annurev.psych.59.103006.093629

Evans, J., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition. *Perspectives on Psychological Science*, *8*(3), 223–241. https://doi.org/10.1177/1745691612460685

Ferguson, C. J. (2015). An effect size primer: A guide for clinicians and researchers. In *American Psychological Association eBooks* (pp. 301–310). American Psychological Association. https://doi.org/10.1037/14805-020

Feynman, R. P., & Sackett, P. D. (1985). *'"Surely You're Joking Mr. Feynman!"' Adventures of a Curious Character*. *American Journal of Physics*, *53*(12), 1214–1216. https://doi.org/10.1119/1.14087

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology*, *40*(2), 177–190. https://doi.org/10.1037/0012-1649.40.2.177

Gillard, E., Schaeken, W., Van Doorden, W., & Verschaffel, L. (2009). *Processing Time Evidence for a Default-Interventionist Model of Probability Judgments*, *31*. https://escholarship.org/uc/item/6sc1970f

Gilovich, T., & Medvec, V. H. (1994). The temporal pattern to the experience of regret. *Journal of Personality and Social Psychology*, *67*(3), 357–365. https://doi.org/10.1037/0022-3514.67.3.357

Girotto, V., Legrenzi, P., & Rizzo, A. (1991). Event controllability in counterfactual thinking. *Acta Psychologica*, *78*(1–3), 111–133. https://doi.org/10.1016/0001-6918(91)90007-M

Jekel, M., Fiedler, S., Allstadt Torras, R., Mischkowski, D., Dorrough, A. R., & Glöckner, A. (2020). How to Teach Open Science Principles in the Undergraduate Curriculum—The Hagen Cumulative Science Project. *Psychology Learning &*

Teaching, *19*(1), 91–106. https://doi.org/10.1177/1475725719868149

Johnson-Laird, P. N. (2010). Mental models and human reasoning. *Proceedings of the National Academy of Sciences*, *107*(43), 18243–18250. https://doi.org/10.1073/pnas.1012933107

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Kahneman, D., & Frederick, S. (2002). Representativeness Revisited: Attribute Substitution in Intuitive Judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and Biases* (1st Ed., pp. *49–81*). Cambridge University Press. https://doi.org/10.1017/CBO9780511808098.004

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

Landman, J. (1987). Regret: A theoretical and conceptual analysis. *Journal for the Theory of Social Behaviour*, *17*(2), 135–160. https://doi.org/10.1111/j.1468-5914.1987.tb00092.x

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

Lewis, D. K. (1973). Counterfactuals. *Journal of Symbolic Logic, 44*(2), 278-281. http://10.2307/2273738

Maxwell, S. R. (2016). Rational prescribing: The principles of drug selection. *Clinical Medicine*, *16*(5), 459–464. https://doi.org/10.7861/clinmedicine.16-5-459

McNab, F., Zeidman, P., Rutledge, R. B., Smittenaar, P., Brown, H. R., Adams, R. A., & Dolan, R. J. (2015). Age-related changes in working memory and the ability to ignore distraction. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(20), 6515–6518. https://doi.org/10.1073/pnas.1504162112

Meehan, J. E. & Byrne, R. M. J. (2005). *The Temporal Order Effect in Children`s*

Counterfactual Thinking. Proceedings of the 27th Annual Conference of the Cognitive Science Society, Edinburgh, Scotland. http://hdl.handle.net/2262/39431

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*(6), 1111–1118. https://doi.org/10.1037/0022-3514.59.6.1111

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., Simonsohn, U., Wagenmakers, E.-J., Ware, J. J., & Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*(1). https://doi.org/10.1038/s41562-016-0021

Murdock, B. B. (1962). The Serial Position Effect of Free Recall. *Journal of Experimental Psychology*, *64*(5), 482. https://doi.org/10.1037/h0045106

National Academies Press (US). (2019, March 3). *Replicability*. Reproducibility and Replicability in Science - NCBI Bookshelf. https://www.ncbi.nlm.nih.gov/books/NBK547524/

Niininen, P. (2013). *Simplicity and Accuracy Trade-Off in Science*. (Graduate thesis, University of Helsinki). https://helda.helsinki.fi/bitstream/handle/10138/160282/simplici.pdf?sequence=3

Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review*, *108*(2), 291–310. https://doi.org/10.1037/0033-295x.108.2.291

Nosek, B. A., & Errington, T. M. (2020). What is replication? *PLOS Biology*, *18*(3), e3000691. https://doi.org/10.1371/journal.pbio.3000691

Patil, P., Peng, R. D., & Leek, J. T. (2016). What should we expect when we replicate? A statistical view of replicability in psychological science. *Perspectives on Psychological Science : A Journal of the Association for Psychological Science*, *11*(4), 539–544. https://doi.org/10.1177/1745691616646366

Peels, R. (2019). Replicability and replication in the humanities. *Research Integrity and Peer Review*, *4*(2). https://doi.org/10.1186/s41073-018-0060-4

Schaeken, W., Johnson-Laird, P. N., & d'Ydewalle, G. (1996). Mental models and temporal reasoning. *Cognition*, *60*(3), 205–234. https://doi.org/10.1016/0010-0277(96)00708-1

Schäfer, T., & Schwarz, M. (2019). The Meaningfulness of Effect Sizes in Psychological Research: Differences Between Sub-Disciplines and the Impact of Potential Biases. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.00813

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 A: Human Experimental Psychology*, *55*(4), 1295–1305. https://doi.org/10.1080/02724980244000125

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

Sherman, S. J., & McConell, A. R. (1996). The Role of Counterfactual Thinking in Reasoning. *Applied Cognitive Psychology*, *10*(7), 113–124. https://doi.org/10.1002/(SICI)1099-0720(199611)10:7<113::AID-ACP433>3.0.CO;2-6

Sikorski, M., & Andreoletti, M. (2023). Epistemic Functions of Replicability in Experimental Sciences: Defending the Orthodox View. *Foundations of Science*. https://doi.org/10.1007/s10699-023-09901-4

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological*

Bulletin, *119*(1), 3–22. https://doi.org/10.1037/0033-2909.119.1.3

Spellman, B. A. (1997). Crediting causality. *Journal of Experimental Psychology*, *126*(4), 323–348. https://doi.org/10.1037/0096-3445.126.4.323

Stanovich, K. E. (1999). *Who Is Rational? Studies of Individual Differences in Reasoning*. Lawrence Erlbaum Associates Publishers.

Stanovich, K. E., & West, R. G. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, *23*(5), 645–665. https://doi.org/10.1017/s0140525x00003435

System 1 and System 2 Thinking - The Decision Lab. (12.03.2023). The Decision Lab. https://thedecisionlab.com/reference-guide/philosophy/system-1-and-system-2-thinking

Turvey, B. E. (2012). *Jury Psychology.* In Ramachandran, V. S. (Ed.), Encyclopedia of Human Behavior. (pp. 495-502) ScienceDirect. http://www.sciencedirect.com:5070/referencework/9780080961804/encyclopedia-of-human-behavior

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*(2), 207–232. https://doi.org/10.1016/0010-0285(73)90033-9

Walsh, C. R., & Byrne, R. M. J. (2001). A Computational Model of Counterfactual Thinking: The Temporal Order Effect. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *23*(23). https://escholarship.org/uc/item/6pn5c2gk

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

Wilkinson, M. R., Ball, L. J., & Alford, D. (2015). Counterfactual Reasoning for Regretted Situations Involving Controllable Versus Uncontrollable Events: The Modulating Role of Contingent Self-Esteem. *Advances in Cognitive Psychology*, *11*(1), 22–30. https://doi.org/10.5709/acp-0168-4

NTNU

Norwegian University of
Science and Technology