Università degli Studi di Padova

Department of General Psychology

Master's Degree Course in Applied Cognitive Psychology

Master's Thesis

The Impact of Perceptual (Dis)fluency on Causality Heuristics in an Associative Learning Paradigm

L'Impatto della (Dis)fluenza Percettiva sulle Euristiche di Causalità in un Paradigma di Apprendimento Associativo

Supervisor:
Michele Vicovaro
University of Padua

Candidate:
Stefano Dalla Bona
Student ID: 2017380

Academic Year 2023 – 2024

«Let the sun shine on, behind me, then!
The waterfall that splits the cliffs' broad edge,
I gaze at with a growing pleasure, when
A thousand torrents plunge from ledge to ledge,
And still a thousand more pour down that stair,
Spraying the bright foam skywards from their beds.
And in lone splendour, through the tumult there,
The rainbow's arch of colour, bending brightly,
Is clearly marked, and then dissolved in air,
Around it the cool showers, falling lightly.
There the efforts of mankind they mirror.
Reflect on it, you'll understand precisely:
We live our life amongst refracted colour.»

– Johann Wolfgang von Goethe, Faust, 1832.

Abstract

When people face challenging mental tasks, they tend to become more attentive and engage in a more deliberate and careful type of reasoning, known as *system two*. This mode of thinking can reduce dependence on the intuitive and effortless kind of reasoning, known as *system one*, which is prone to cognitive biases. One such bias is the *illusion of causality*, where individuals mistakenly perceive a causal relationship between unrelated events in associative learning contexts. Díaz-Lago and Matute (2019a) found that a superficial perceptual feature, such as a difficult-to-read font, can weaken the strength of this illusion.

Our study sought to explore whether *perceptual disfluency* – making something harder to perceive – could similarly reduce the illusion's strength across different conditions. In our first experiment, we investigated whether changing the contrast between text and background in a *contingency learning task* would affect the illusion of causality. Although we successfully created conditions of fluency and disfluency in a 200-participant online experiment, the results showed no effect of contrast on the strength of the illusion. Following this null result, our second experiment, with 100 participants, focused on manipulating font type to test if we could replicate the findings of Díaz-Lago and Matute (2019a). Contrary to their results, we found that different font types had no significant impact on the illusion's strength, even though this manipulation also created varying levels of task fluency and disfluency. These findings suggest that not all forms of cognitive disfluency can influence biases in the same way. They emphasize the need to reevaluate and refine our understanding of how (dis)fluency affects cognitive processes and biases.

This thesis originates from a master's internship dedicated to the programming and execution of an experiment on the causality heuristic. This effort culminated in a peer-reviewed international journal publication (Dalla Bona & Vicovaro, 2024), which can be accessed via the following link:

https://www.researchgate.net/publication/376262740_EXPRESS_Does_perceptual_disfluency_affect_the_illusion_of_causality

Summary

1

A higher point of view on biases

1.1 Cognitive biases and heuristics

The term *cognitive illusion* (or *cognitive bias*) encompasses a broad spectrum of phenomena that collectively illustrate deviations in thinking, judgment, and memory from an objectively correct standard. Typically, biases are studied through experiments in which participants are assigned specific tasks, and deviations in their judgments from an intersubjectively shared mathematical or logical baseline are observed as systematic patterns (Haselton et al., 2015).

Interest in the field of cognitive biases can be traced back to the early 1970s (Tversky & Kahneman, 1996), with the introduction of a research program on *judgment under uncertainty* (Tversky & Kahneman, 1974). Over the years, some authors have endeavored to identify and catalog all cognitive biases (e.g., Benson, 2016), while others have asserted that an accurate and unified definition of cognitive bias is unattainable (Caverni et al., 1990).

As it has been argued (Kahneman & Frederick, 2002), biases emerge from our tendency to rely on *heuristics* – defined as sets of rules of thumb that can expedite decision-making in an efficient manner. Heuristics can be functionally interpreted in two ways: as evidence of distorted perception and flawed reasoning, resulting in biases, or as a set of processes that the human mind employs to solve problems, which are functionally effective in decision-making processes most of the time. These two perspectives can both hold true simultaneously. Quick and intuitive thinking is necessary to reduce the number of variables processed by the cognitive system, which is crucial for efficiently analyzing problems and making adaptive responses to the environment. However, this mode of thinking can also fail in correctly evaluating certain types of problems, leading to distortions (Tversky & Kahneman, 1974). In this thesis, we aim to maintain a broader viewpoint, focusing primarily on the functional causes (i.e., mechanisms and processes) and consequences of scenarios in which the human mind opts for a different set of rules than formal ones, specifically in the context of the illusion of causality.

1.2 Counter-intuitivity as a substitute for illusion

According to Roediger (1996), an analogy can be drawn between cognitive biases and optical illusions. Just as sensory processing can lead to the misperception of a physical stimulus, the processes of codification, elaboration, and retention of information can lead to numerous judgment errors. Due to the limited processing capacity of the cognitive system, in many situations of judgment under uncertainty, the human mind employs a small set of heuristics that can lead to severe and systematic errors (Tversky & Kahneman, 1974), suggesting that cognitive biases are robust, universal, and unavoidable. Pohl (2022) highlighted five analogies between cognitive biases and optical illusions:

- *Deviation from reality* – the phenomenon represents a deviation from a correct normative standard.
- *Systematic deviation from the standard* – the observed phenomenon deviates from the normative standard in a predictable manner.
- *Involuntary production of the illusion* – biases appear without deliberative will.
- *Impossibility of avoiding the illusion* – biases are hard or even impossible to avoid in their manifestation.
- *Universal appearance of the illusion* – biases appear across all people, independently of confounding or psychological variables.

As discussed by Gigerenzer (2008), this negative connotation of biases can be erroneous, leading to a one-sided view of human rationality that focuses too much on errors and results in a pessimistic view of human thinking. Furthermore, the analogy can be somewhat misleading. First, from our very first stage of cognitive processing (i.e., perception), we *systematically deviate from reality*, as classically supported by studies from *Gestalt* psychology (Atkinson & Hilgard, 2017). When reasoning, we do not usually follow standard logical and mathematical rules. Instead, what we can ask is in which context the tendency to base our judgment on heuristics results in evident errors, how salient these results are considering the deviation from a rule and the ecological validity of a task, and how the emergent phenomenon, in the specific paradigm, unveils how we process the input to produce the output. Secondly, many relevant psychic events occur in a non-conscious state, and the production of automatic responses does not require intention and controlled processes (Cornoldi et al., 2018). Indeed, the production of a biased response can be intended to be involuntary, but it could be asked why this property should be a distinctive feature for biases, as most processes are unintentional. Additionally, researchers have shown that biases can be mitigated in certain contexts (Pohl, 2022; see also Maguire et al., 2018, for an example), and the appearance of an error in reasoning can largely depend on the kind of task we propose to people (see Section 1.5).

Drawing analogies to provide a better understanding of phenomena can be important, but the *illusion* analogy may be inadequate. We propose that biases can be understood as counter-intuitive results in cognitive tasks that provide a deeper understanding of the underlying processes. They produce a paradoxical effect where the same cognitive system, intended as a whole, that generally enables adaptive responses to the environment can sometimes lead to incorrect responses within the same properties of the system. Analyzing the production of these responses is important for unveiling how we conduct certain types of reasoning.

This paradoxical definition is particularly apt when comparing cognitive biases to the study of paradoxes in other fields. For example, in logic and mathematics, paradoxes have been invaluable in extending knowledge within the field. Without delving into an extensive history of paradoxes (e.g., the liar paradox – *I am lying* – which dates back to ancient Greece), their formulation, which often involves elements of self-reference and negation, shows that these counter-intuitive propositions can serve to analyze rules of logic in greater detail and enhance our understanding of the system as a whole.

As an example, Bertrand Russell's famous paradox[1] (see Equation 1.1) showed how Gottlob Frege's attempt to reduce mathematics to logic led to a contradiction, leading to a reevaluation of the foundations of mathematics (Irvine & Deutsch, 2021).

$$A = \{a : a \notin a\} \implies A \in A \iff A \notin A \tag{1.1}$$

1.3 Classifications

Pohl (2022) proposed a valuable distinction for biases, classifying them into three categories. It is important to note that a particular bias may fall into multiple categories:

- *Illusions of thinking*: These biases involve the application of certain mathematical or logical rules (e.g., the *conjunction probability rule*) derived from a normative model, which constitutes a standard against which human performance is evaluated.
- *Illusions of judgment*: These biases occur when participants are asked to subjectively rate a specific feature of given stimuli (e.g., their pleasantness), and certain features within the context of presentation may bias participants' judgment in a particular direction.
- *Illusions of memory*: These biases occur when individuals are required to recall information that was encoded earlier, often leading to memory errors or distortions. In the *Oxford Handbook of Memory*, Roediger and McDermott (2000) offer an extensive review of various memory phenomena and distortions.

A similar distinction among cognitive biases has been proposed by Hell et al. (1993), who also included misconceptions in physics. Over the years, various classifications have been suggested, emphasizing different aspects of cognitive biases (Pohl, 2022).

Within the broader category of illusions of thinking, a useful distinction is based on the type of reasoning involved in specific tasks. Reasoning, as a form of thought, can be divided into two main types (Cornoldi et al., 2018):

[1] *Russell's paradox*: let A be the set of all sets that are not members of themselves. If A is not a member of itself, then by definition, it must be a member of itself. Conversely, if A is a member of itself, then according to its definition, it cannot be a member of itself. Thus, we arrive at a contradiction.

- *Deductive reasoning*: this type of reasoning involves drawing a conclusion from a set of premises, where the truth of the conclusion is directly related to the truth of the premises. Deductive reasoning is often employed in tasks that require logical analysis, such as solving *Aristotelian syllogisms* (Cornoldi et al., 2018). One example of a task that assesses deductive reasoning is the *Wason selection task* (WST; Wason, 1966). In this task, participants are shown four cards, each displaying a letter or a number (e.g., $E - K - 4 - 7$). They are informed that each card has a letter on one side and a number on the other. Participants are given a conditional statement – *If there is a consonant on one side, then there is an even number on the other side* – and are asked to determine which cards need to be flipped to test this rule. Participants often struggle with this task challenge due to a bias toward confirming the rule rather than testing for disconfirmation. For instance, they may select the cards K and 4 to check if they conform to the rule, rather than selecting K and 7 to test if the rule is violated. Performance on this task can improve when it is presented in a more realistic context (Cornoldi et al., 2018), highlighting how familiarity and context can affect reasoning abilities.

- *Inductive reasoning*: this type of reasoning involves drawing general conclusions or making predictions based on a set of specific observations or patterns. It often includes extrapolating general rules from limited data, estimating probabilities, categorizing events to reduce variability, and making decisions based on observed trends. A common example of a heuristic that can lead to a bias within inductive reasoning is the representativeness heuristic (Tversky & Kahneman, 1974). A bias occurs when people judge the likelihood of an event based on how much it resembles a typical or representative example, rather than on statistical probabilities or base rates. For instance, when people are asked to judge the probability of a sequence of dice rolls, such as $2 - 5 - 4 - 3 - 4 - 5 - 1 - 1 - 6 - 3$, they might perceive it as more likely than a sequence of all ones (i.e., $1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1$), because the first sequence appears more random and representative of what they expect from a fair dice roll. Despite the fact that both sequences are equally probable, the representativeness heuristic leads people to favor the one that seems more in line (i.e., representative) with their notion of randomness.

1.4 Dual-process models

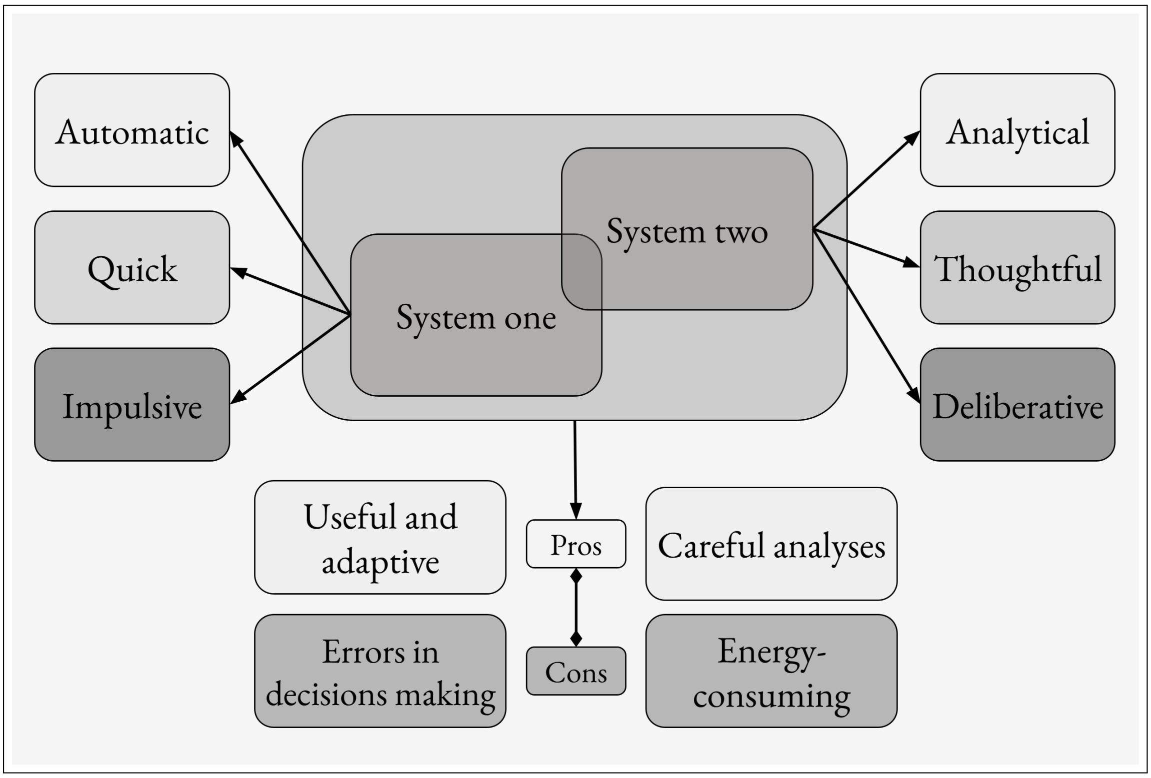

Figure 1.1: Dual-process models' most common interpretation.

Numerous models have been proposed to account for heuristics and biases, with dual-process models being a prominent class. In this thesis, we will focus on these models because the results from our study (see Section 3.4) can be interpreted through this specific framework. Dual-process models (Pohl, 2022) are commonly used to understand biases and generally distinguish between two types of reasoning (Evans, 2011). The first type, often referred to as *type one* reasoning, is automatic, relatively independent of working memory (WM), operates in parallel, and is usually faster. The second type, known as *type two* reasoning, is more controlled, relies on WM, processes information serially, and is slower.

Stanovich (1999) introduced the terms *system one* and *system two* to represent these two types of processes. However, Evans (2011) cautioned that these labels can be somewhat misleading, as they imply that only two cognitive systems underlie various tasks.

Functionally, system one rapidly generates intuitive responses, while system two monitors and controls these responses, potentially endorsing, correcting, or overriding them. Judgments are attributed to system one if they involve minimal modification from the initial intuitive proposal (Kahneman & Frederick, 2002). According to Stanovich (1999), both systems operate in parallel, with system one functioning continuously and system two intervening as needed.

As discussed by Kahneman (2012), system one is advantageous for quick and effortless decisions, whereas system two is beneficial for tasks requiring computational power and deliberate thought. The reliance on system one increases the likelihood of using heuristics, which can lead to biases. Although system one may be seen as more primitive compared to system two, it is not necessarily less capable (Kahneman & Frederick, 2002). In fact, complex mental operations can shift from system two to system one as individuals become more skilled in a particular task. A notable example of system one's capability is seen in elite chess players, who, through extensive practice, develop an intuitive ability to evaluate chess positions almost instantaneously.

Despite the widespread acceptance of dual-process models for interpreting heuristics and biases, influential critics argue that single-process accounts might be sufficient (Kruglanski & Gigerenzer, 2011). However, these critiques often overlook the robust evidence supporting dual processing from cognitive psychology and neuroscience (Evans, 2011).

Evans (2012) identified several common misconceptions related to dual-process models, which highlight important considerations for understanding cognitive biases. One notable misconception is the assumption that system one (or type one processes) is always responsible for biases, while system two (or type two processes) is associated with normative responses. In reality, system two, although necessary for performing well on complex cognitive tasks, is not always sufficient for ensuring correct responses. Simply engaging system two, which involves more resources and deliberate thought, does not guarantee accurate application of logical or mathematical rules. Furthermore, the notion that system one can produce incorrect responses is primarily relevant in specific, controlled experimental paradigms where participants are faced with novel and challenging tasks. In these contexts, the reliance on system one is often triggered by particular cues. Additionally, it has also been supported that cognitive biases can arise from both system one and system two processes (Pohl, 2022). To accurately capture the kind of reasoning required by an experiment, it is essential to examine how the task prompts either system one or system two. If a bias arises from system one's involvement, researchers should evaluate whether the task genuinely captures a meaningful phenomenon. While dual-process models offer valuable insights, we have to acknowledge the limitations of standard models and emphasize the importance of contextual relevance in understanding cognitive processes.

1.5 Critiques

Evidence strongly supports the existence of biases and their significant impact on everyday life, accounting for numerous phenomena. However, Pohl (2022) summarized several critiques that have been pointed out by Gigerenzer (1991; 1996; 2008; Gigerenzer et al., 2008), highlighting the need to consider these critiques to improve the quality of studies and research:

- The task might be misleading, eliciting biased behavior in participants. For example, in a famous task proposed by Wason (1960), participants are asked to extrapolate a generative rule from a series of numbers presented, such as $2-4-6$, and are then requested to produce other series according to the rule they deduce. The experimenter then confirms or rejects their responses. Participants might incorrectly infer that the rule is *even numbers in ascending order* and produce series like $8-10-12$, while the actual rule is simply *numbers in ascending order*. When asked to state the underlying rule, participants often respond incorrectly – not due to a bias, as originally intended, but because the initial number series was misleading. This example underscores the importance of researchers being aware of the potential intrinsic misleading effects of tasks and the need for careful planning. However, even if a task is highly misleading, it does not invalidate the investigation of the underlying processes, as long as the artificial pitfalls are acknowledged and managed.

- Researchers might use an inadequate presentation format or material sampling. Using inappropriate statistical formats, such as probabilities instead of frequencies, or engaging in selective sampling could distort study results. For example, researchers might focus on specific samples of materials where a positive result for their research can be found, neglecting other types of materials that are related to the same underlying processes and might support that our reasoning and memory are not inherently flawed. This selective approach can lead to biased conclusions, emphasizing the need for comprehensive and balanced material selection in research.

- The experiment might present a task that simply highlights a lack of knowledge in participants. For example, the *conjunction fallacy* (CF) illustrates individuals' inability to intuitively assess the conjunction probability rule (Fisk, 2016). This rule states that, given two distinct events, the likelihood of a single event, $P(A)$ or $P(B)$, is invariably greater than or equal to the probability of both events co-occurring, denoted as $P(A \wedge B)$. It can be argued that people simply do not know the conjunction rule, thereby artificially prompting the supposed bias. However, there are reasons to believe that the CF can be considered a real and consistent phenomenon. For instance, even in contexts where the application of a mathematical rule is more salient, people still manifest the fallacy (Maguire et al., 2018). Recently, it has been accepted by the majority of researchers as a genuine phenomenon, although some still debate its real-life implications (Fisk, 2016).

❖ The normative rule used might be wrong. To illustrate this, Gigerenzer (1991) argued that probability rules are about frequencies and do not apply to judgments of single events, typically used in cognitive biases experimental paradigms. For example, presenting the CF paradigm in terms of a frequentist interpretation of the task (i.e., examining the probability within a group of 100 people rather than presenting a single case) can result in a reduction in the incidence of the fallacy (Gigerenzer, 1991). Nonetheless, contrary to this finding, it has been shown that presenting the CF in terms of a frequentist kind of task does not always lead to a reduction of the bias (Fisk, 2016). However, this critique is important as it prompts us to consider which kind of normative standard is useful in a given context and how the results obtained can be used to draw conclusions in a more ecological environment.

❖ Some phenomena might be explained without referencing a failure in our information processing. For instance, what appears to be a cognitive bias might instead be a rational response given the context or the available information. Rather than always indicating a flaw, some biases may reflect adaptive strategies that work well in everyday environments, even if they lead to errors in experimental settings. Thus, it is essential to consider alternative explanations that do not necessarily involve faulty reasoning processes.

1.6 Reasoning upon statistics

Heuristics and biases studies have been successfully applied in various fields, underpinning everyday decision-making processes. These studies have proven valuable in economics (Cornoldi et al., 2018), applied cognitive psychology (e.g., in medical decision-making and in eyewitness testimony; Pohl, 2022), elderly psychology (De Beni & Borella, 2015), and clinical psychology (e.g., in relation to obsessive-compulsive disorder; Pohl, 2022).

Studies on heuristics and biases are also beneficial in understanding how people approach mathematical and statistical concepts in educational settings or learning environments (Garfield, 2002). They provide insights into how people interpret data presented in scientific communication through mass media, influencing assumptions and decision-making. Scientific and statistical communication bridges the scientific community and the general public (e.g., WHO, 2023). This connection was particularly evident during the COVID-19 pandemic, significantly shaping normative regulations and individual behaviors (Warren & Lofstedt, 2022).

Focusing on statistical communication, the literature highlights that conveying statistical information is fraught with inherent complexities and potential pitfalls. From a top-down perspective, statistical data can be misleading due to how it is presented (Huff, 1954) and visually represented (Pastore et al., 2017). Conversely, from a bottom-up approach, acquiring proficiency in statistical knowledge encompasses a range of skills, including data interpretation, understanding graphical representations, and calculating statistical measures (Garfield, 2002). These skills engage various psychological processes, making the avoidance of errors and misconceptions in statistical reasoning a significant challenge. Some statistical errors can be viewed as the phenomenological manifestation of specific cognitive heuristics. Individuals often infer statistical and probabilistic relationships between events and contingencies naively, frequently violating normative mathematical principles (Pohl, 2022).

1.7 Categorization of the illusion of causality

Lastly, an effort should be made to categorize the causality bias within the frameworks proposed in Section 1.3. As we will detail in Section 2.3, this bias occurs when individuals erroneously overestimate the causal link between a cue and an outcome after reviewing a series of trials, each characterized by the presence or absence of the cue and the outcome. This overestimation is believed to stem from our tendency to disproportionately emphasize evidence supporting the presence of an effect. The illusion of causality, as an *illusion of thinking*, can be understood as a biased evaluation of raw data, where individuals tend to prioritize true positives – scenarios where both the supposed cause and effect are observed – over true negatives, false positives, and false negatives. The fact that individuals extrapolate a rule based on the observation of the frequencies of different scenarios indicates that the causality heuristic aligns with *inductive* type of reasoning and falls within the broader category of *statistical heuristics*.

2

Illusion of causality

2.1 The nature of causality

As the illusion of causality pertains to individuals' perception of causality itself, it raises the fundamental question of what causality is in its essence. While causality can be roughly defined as the relation between two events, one of which is the consequence (i.e., the effect) of the other (i.e., the cause), philosophers have long explored the ontological nature of causality. Different schools of thought have emerged over centuries, as described by Broadbent (2024).

Regarding the nature of causality, *realists* (e.g., Armstrong, 2016) argued that there are real entities that exist independently of particular instances. They saw causation as something that exists beyond the particular things that are causally related, a universal relation that underlies and connects cause-effect pairs. In contrast, *nominalists* held that there are no entities other than what Lewis (1983) refers to as *distinct existences*. Nominalists argued that causation is not a particular entity and it is not something that exists beyond its particular instances. In this perspective, causation is nothing more than the sum of its specific occurrences.

Another perspective on causality is provided by Kant (1855/2007), extending ideas from David Hume (Broadbent, 2024). Kant asserted that causation is not an objective thing but a feature of our experience, arguing that causation is essential to any kind of experience.

David Hume's viewpoint holds a particularly important role within the field of psychology (Wasserman et al., 1990), as he raised the question of how we know about causal connection. For Hume (1740/2000), causal impression depends on previous experience. The impression of causality between two events is formed when they are temporally contiguous, the cause precedes the effect, and there is a constant coincidence over time between these two. However, the process by which the impression of causality is formed is not given by deliberative and inferential thinking, but rather by purely mechanistic learning.

2.2 Cognition of causality

As from the first conceptualization of the question about how we perceive causality by Hume, psychologists' interest in how causal inference works flourished, becoming a traditional research topic in psychology explored from various perspectives, including comparative cognition (Blaisdell et al., 2006), psychology of reasoning (Waldmann et al., 2006), psychology of learning (Dickinson et al., 1984), and visual perception (Michotte, 1963/2017). The psychological literature on this topic suggests that the understanding of cause-effect relationships is an ability in which humans clearly outperform any other species (Bender, 2020).

With respect to the perception area, Michotte's studies on causality hold significant relevance. In a famous study paradigm (Michotte, 1963/2017), participants observed a moving object, labeled as X, approaching and making contact with a stationary object, labeled as Y. Upon contact, the motion of X ceased, and Y began to move. When Y started moving within $1/10$ second after contact with X, and in the same direction as X, participants consistently reported that X caused Y to move. This generated a strong and reliable perception of causality. Michotte argued that causality is directly perceived, without the need for mediation by higher-level cognitive processes.

However, in this thesis, we will focus on the impression of causality (i.e., the causality bias) in contexts where associative learning occurs. Associative learning is a type of learning in which two initially unrelated objects become connected in our minds through a process known as conditioning (Cornoldi et al., 2018). For this reason, the illusion of causality can also be classified within the category of illusions of memory (see Section 1.3).

2.3 Causality bias in a learning context

The illusion of causality occurs when a subject develops the belief that there is a causal connection between two events that are actually unrelated. It refers to the perception that one event A, called the cue or potential cause, is causally linked to another event B, called the outcome or effect, when there is merely a coincidence between them. Generally, humans infer the presence of a causal link through the (single or multiple) contingencies between A and B (Matute et al., 2019), often showing great accuracy in detecting causal links that are genuinely present in the environment. This ability is critical for survival, as it underlies the capacity to make accurate predictions about future states of the world. However, sometimes contingency learning can lead to an overestimation of the degree to which a causal link is present when, in fact, the two events are independent (i.e., the probability of A is substantially independent of the probability of B), resulting in the so-called *over-estimation of zero-contingencies* (Blanco et al., 2014).

A. Structure		*Outcome* B_1	B_2
Cue	A_1	a	b
	A_2	c	d

B. Presence		*Outcome* B_1	B_2
Cue	A_1	25	5
	A_2	5	5

C. Absence		*Outcome* B_1	B_2
Cue	A_1	10	10
	A_2	10	10

Table 2.1: Fundamental contingency table and illustrative variations.

One of the most widely used paradigms that has become a standard experiment to explore causal learning in general and the illusion of causality in particular is the *contingency learning task* (CLT). In this task, participants are presented with a series of trials, each one characterized by the presence or absence of event A and event B. Indeed, the presence or absence of event A and event B gives rise to four hypothetical scenarios, where the respective frequencies can be represented on a tetrachoric table (see Table 2.1 – Panel *A. Structure*): (a) event A and event B are present (i.e., the cue and the outcome co-occur), (b) only event A is present (i.e., the cause manifests without the outcome), (c) only event B is present (i.e., the cause does not manifest, but the outcome does), and (d) event A and event B are not present (i.e., neither the cue nor the outcome is present; Vadillo & Matute, 2007). In a context of observation, these four combinations can appear with different frequencies, ranging from zero onwards.

Each trial shows the presence or absence of the potential cause A, linked to the presence or absence of an effect B so that event A, whether present or absent, precedes event B; in this way only A can signal B, giving rise to a *one-way dependency*. Typically, events A and B are chosen so that it is plausible that A can be the potential cause and B the outcome. Studies have proposed different types of events A and B that in principle can be causally related (e.g., a fertilizer as a potential cause and a flower blossom as an outcome; Matute et al., 2022).

After a certain number of trials, which is manipulated by the researcher, participants are usually asked to estimate the degree to which there is a causal connection between the events. Typically, a numeric scale from 0 to 100 is used to estimate the degree of the causal link, where 0 is interpreted as no causal connection and 100 as the maximum degree of a causal connection. While this scale makes sense from a theoretical standpoint (as it reflects the ΔP rule, which will be discussed below in Section 2.4), it can be somewhat misleading. For instance, a middle point of 50 can be interpreted by participants as an indecisive expression of causality or a moderate presence of a causal link. Thus, it is not surprising that some research finds different results based on the kind of scale used (Ng et al., 2024).

The presence or absence of the two target events A and B is manipulated by researchers in their proportions so that, from a normative standpoint (i.e., according to the ΔP rule; see Section 2.4), there is or is not some degree of a statistical link between the two events. Researchers systematically vary the frequencies of the a, b, c, d scenarios to create conditions with different levels of contingency between the cue and the outcome. For example, in a positive contingency condition, the cue and the outcome would frequently appear together (i.e., high frequency of scenario a; see Table 2.1 – Panel *B. Presence*) and rarely appear independently (i.e., low frequencies for scenarios b and c). Conversely, in a low or null contingency condition, the cue and the outcome would appear independently of each other, leading, for example, to balanced or low proportions across all four scenarios (see Table 2.1 – Panel *C. Absence*). When asked about the causal link between events, people are typically fairly accurate in assessing the presence or absence of a causal connection. However, in certain scenarios specifically manipulated by researchers, where no real causal link is present, people tend to overestimate the extent of a causal link between the events. This overestimation is the operational definition of the manifestation of an illusion of causality. Thus, the illusion is typically *induced* in participants, exposing them to a series of trials where normative indices assert the absence of a causal link, and the frequencies of the outcome, the cause, or both are increased to generate the illusion (see Section 2.4).

In some variants of this paradigm, the presence or absence of the cause can be manipulated by the participant, transitioning the procedure from passive observation to active engagement.

- ❖ In an active procedure, participants are given control over the presence of the cause. They decide when to introduce the cause, while the outcome's occurrence is still probabilistically controlled by the researcher. This active engagement shifts the paradigm towards the illusion of control (Matute et al., 2022). According to Langer (1975), the illusion of control refers to the tendency of individuals to overestimate their influence over outcomes that they have no actual control over.

- ❖ In a passive procedure, participants observe a series of trials where the presence or absence of the cause and the outcome are determined entirely by the experimenter. Participants do not influence the events and act purely as observers. The illusion of causality, as intended in our study, is primarily observed in this passive context.

2.4 Standard normative model

In the literature, different normative models of causal induction have been proposed (see Section 2.6). However, the most widely used method to measure contingency is the ΔP contingency index (Allan, 1980), a normative model for human causal learning (Matute et al., 2022). The ΔP index is calculated by subtracting the probability of observing the outcome when the cue event is not present, expressed as $P(O|\neg C)$ or $P(B_1|A_2)$ (following Table 2.1 nomenclature), from the probability of observing the outcome when the cue is present, expressed as $P(O|C)$ or $P(B_1|A_1)$ (following Table 2.1 nomenclature; Jenkins & Ward, 1965):

$$\Delta P = P(B_1|A_1) - P(B_1|A_2) \iff \frac{a}{a+b} - \frac{c}{c+d} \tag{2.1}$$

where a, b, c, and d are the observed frequencies of the four scenarios represented in Table 2.1. Three cases can be observed depending on the value of ΔP:

- ❖ If ΔP equals zero (i.e., $P(B_1|A_1) = P(B_1|A_2)$), then there is no contingency between the cue and the outcome, indicating no causal link.

- If ΔP is positive (i.e., $P(B_1|A_1) > P(B_1|A_2)$), a positive contingency and a causal link are present.

- If ΔP is negative (i.e., $P(B_1|A_1) < P(B_1|A_2)$), then the contingency is negative, suggesting an inhibitory effect of the cue event on the outcome.

Theoretically, the overestimation of the extent to which A and B are causally related can occur in any of these cases (i.e., when ΔP is negative, null, or positive). However, the causality bias has been predominantly studied in the case of null contingency (Allan, 1980).

The illusion arises mainly in the null contingency condition, specifically when:

- The frequencies of the scenarios in which the outcome is present (cells a and c in Table 2.1) are larger compared to the frequencies of the scenarios in which the outcome is absent (cells b and d in Table 2.1), despite the ΔP index being zero (Alloy & Abramson, 1979). This condition leads to the so-called *outcome-density bias* (Matute et al., 2015).

- The frequencies of the scenarios in which the cue is present (cells a and b in Table 2.1) are larger compared to the frequencies of the scenarios in which the cue is absent (cells c and d in Table 2.1), despite the ΔP index being zero (Allan & Jenkins, 1983). This condition leads to the so-called *cause-density bias* (Matute et al., 2015).

- The frequencies of both the cue and the outcome are jointly increased, leading to a higher frequency of scenario a relative to the other three scenarios, while the ΔP index remains zero (e.g., $a = 64$, $b = 16$, $c = 16$, $d = 4$; see Blanco et al., 2013). The emergence of a causality bias under this specific condition suggests that scenario a (i.e., when both the cue and the outcome are present) plays a particularly important role in the causal induction mechanism, as we previously discussed in Section 1.7.

In all these cases, the judged strength of a causal relationship tends to be consistently and systematically larger than what would be expected from the hypothetically correct normative response (Blanco, 2017).

Lastly, it is important to assert that the strength of the perceived causal link between events can be actively modulated – the researcher can create conditions with greater outcome-density or cause-density, which can contribute to generate a greater illusion of causality.

2.5 ΔP as a statistical rule

The ΔP index is a rule that people commonly (though approximately) follow when evaluating causal information in a learning context (see Section 2.3). In some specific cases in a null contingency condition, people do not follow the ΔP rule, and the discrepancy is indicated as the illusion of causality. In this section, we endeavor to show how the ΔP index is statistically related to the *chi-squared* (χ^2) statistic (Allan, 1980), supporting that it can be employed as a standard mathematical rule against which human performance can be evaluated.

Suppose we have two nominal variables A and B, and we indicate absolute double frequencies on a contingency table as n_{ij}, where $i = 1, 2, \ldots, k$ and $j = 1, 2, \ldots, h$.

The contingency table (2.2) can be represented as:

	B_1	B_2	$\cdots$	B_h	Total
A_1	n_{11}	n_{12}	$\cdots$	n_{1h}	$n_{1.}$
A_2	n_{21}	n_{22}	$\cdots$	n_{2h}	$n_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
A_k	n_{k1}	n_{k2}	$\cdots$	n_{kh}	$n_{k.}$
Total	$n_{.1}$	$n_{.2}$	$\cdots$	$n_{.h}$	N

Table 2.2: Contingency table of variables A and B.

Where:

- n_{ij} represents the frequency of observations for the combination of the i-th category of A and the j-th category of B.
- $n_{i.}$ is the total frequency for the i-th category of A across all categories of B.
- $n_{.j}$ is the total frequency for the j-th category of B across all categories of A.
- N is the grand total of all frequencies in the table.

To express the expected frequencies under the assumption of perfect independence between A and B in a contingency table, we can use the formula:

$$E_{ij} = \frac{n_{i.} \cdot n_{.j}}{N} \tag{2.2}$$

The χ^2 statistic is computed using the formula:

$$\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{h} \frac{(n_{ij} - E_{ij})^2}{E_{ij}} \tag{2.3}$$

Where n_{ij} is the observed frequency in cell (i, j) and E_{ij} is the expected frequency for that cell. A special case for applying the χ^2 statistic involves a 2×2 contingency table, often referred to as a *tetrachoric table*, where we have two binary nominal variables A and B. The table (2.3) is structured as follows:

	B_1	B_2	Total
A_1	n_{11}	n_{12}	$n_{1.}$
A_2	n_{21}	n_{22}	$n_{2.}$
Total	$n_{.1}$	$n_{.2}$	N

Table 2.3: 2x2 contingency table (tetrachoric table).

At this point, it is important to note that Table 2.1 and Table 2.3 express the same condition under different notations. The χ^2 statistic can be computed using the formula:

$$\chi^2 = \frac{N(n_{11} \cdot n_{22} - n_{12} \cdot n_{21})^2}{n_{.1} \cdot n_{.2} \cdot n_{1.} \cdot n_{2.}} \tag{2.4}$$

The Φ index, a measure of association between two variables on a tetrachoric table (that varies from 0, corresponding to no association between the variables, to 1 or -1, which respectively indicate complete association or complete inverse association), is based on frequency data represented in 2×2 tables. It can then be calculated as:

$$\Phi = \sqrt{\frac{\chi^2}{N}} \tag{2.5}$$

Both χ^2 and Φ reflect the dependence of variable A on variable B and the dependence of variable B on variable A.

However, a different measure of dependency, denoted as ΔP, represents the difference between two independent conditional probabilities and can be used to measure the one-way dependency of one variable on another.

The measure of dependency of variable B on variable A is given by:

$$\Delta P_{B \leftarrow A} = P(B_1 \mid A_1) - P(B_1 \mid A_2) = \frac{n_{11}}{n_{11} + n_{12}} - \frac{n_{21}}{n_{21} + n_{22}} \tag{2.6}$$

Where $P(B_1 \mid A_1)$ is the probability of B_1 given A_1, and $P(B_1 \mid A_2)$ is the probability of B_1 given A_2. This formula is the same expression of Equation 2.1.

Similarly, the measure of dependency of variable A on variable B is:

$$\Delta P_{A \leftarrow B} = P(A_1 \mid B_1) - P(A_1 \mid B_2) = \frac{n_{11}}{n_{11} + n_{21}} - \frac{n_{12}}{n_{12} + n_{22}} \tag{2.7}$$

Where $P(A_1 \mid B_1)$ is the probability of A_1 given B_1, and $P(A_1 \mid B_2)$ is the probability of A_1 given B_2.

Considering equations 2.4, 2.6, and 2.7, then the χ^2 is calculated as:

$$\chi^2 = N \cdot \Delta P_{B \leftarrow A} \cdot \Delta P_{A \leftarrow B} \tag{2.8}$$

That is, χ^2 reflects a two-way dependency, and $\Delta P_{B \leftarrow A}$ and $\Delta P_{A \leftarrow B}$ each reflect a one-way dependency.

2.6 Overview of alternative theoretical models

While the ΔP model is one of the most widely used approaches for explaining how naive reasoners infer causality from contingency in an associative learning context, several other theoretical models have also been proposed over the years. Perales and Shanks (2007) provided a comprehensive summary of the most significant models of covariation-based causal judgment.

In this final section we present a summary of some alternative models used to study causal learning. Theoretical models of causal induction can generally be divided into two main categories: *norm-based models* and *algorithmic models*, which we will introduce briefly in Subsections 2.6.1 and 2.6.2, respectively. In Subsection 2.6.3, we will discuss in further detail a specific algorithmic model of particular relevance.

2.6.1 Norm-based models

Norm-based models assume that people acquire causal knowledge by applying psychological processes that resemble rational strategies (Perales & Shanks, 2007). According to these models, individuals adopt a certain criterion to follow a rational analysis of causality, resulting in a correspondence, to some degree, between the output produced by normative rules and the outcome of psychological processes. The ΔP rule can be included in this set of models.

Another model that can be included in this category is the *power theory of probabilistic contrast* or *power PC* model (Cheng, 1997). This theory is grounded in the ΔP rule but extends it by incorporating the concept of interactive causes (Matute et al., 2022).

Consider a potential cause, denoted as event $A(1)$, alongside a set of other background causes, represented by event $A(0)$. These other causes comprise both observed and unobserved factors that operate in the background and may, for instance, produce the outcome even in the absence of event $A(1)$ (i.e., cell c in Table 2.1; Perales & Shanks, 2007). Assuming that the complete set of causes for a given event B can be partitioned into events $A(0)$ and $A(1)$, we can illustrate their relationship using a directed arrow graph, where B represents the common effect of these two causes ($A(1) \rightarrow B \leftarrow A(0)$; Perales & Shanks, 2007). Events $A(0)$, $A(1)$, and B can be either present or absent.

In this model, the focus is on estimating the causal power of event $A(1)$, denoted as $\alpha_{A(1)}$, defined as the probability with which an event $A(1)$ produces an event B when event $A(1)$ is present. Causal power aims to capture the probability with which the cause actually causes the effect. Indeed, causal power can also be estimated for $A(0)$, denoted as $\alpha_{A(0)}$. The causal power $\alpha_{A(k)}$ for any event $A(k)$ assumes a probability value from 0 to 1. This probability value can also be depicted as the weight assigned to one causal arrow in the graph (i.e., in this instance, one for $A(0)$ and one for $A(1)$), and $\alpha_{A(k)}$ can be understood as a random variable representing the strength of event $A(1)$ in influencing event B (Holyoak & Cheng, 2011). The causal power $\alpha_{A(1)}$ is denoted with a Greek letter, as it is a theoretical value and only indirectly estimated. It differs from the probability of event B given the presence of event $A(1)$, denoted as $P(B_1|A(1)_1)$, because the latter is directly observed (Cheng, 1997) and includes those occasions when the event A brought about event B, as well as occasions on which the event A was present but failed to bring about the event B (Luhmann & Ahn, 2005). $P(B_1|A(1)_1) = \alpha_{A(1)}$ only when no other event $A(k)$ is present or exists.

The power PC theory posits that people approach causal learning with four general prior assumptions (Holyoak & Cheng, 2011):

- Events $A(1)$ and $A(0)$ influence event B independently.
- Event $A(1)$ could produce event B but not prevent it.
- Causal powers $\alpha_{A(0)}$ and $\alpha_{A(1)}$ are independent of the frequency of occurrence of events $A(0)$ and $A(1)$.
- Event B does not occur unless it is caused.

In a context where there is a potentially generative event $A(1)$ (i.e., one that is assumed to *produce* event B; Perales & Shanks, 2007), the probability of observing event B, as event B can be produced independently by $A(1)$ or $A(0)$, is given by:

$$\begin{aligned} P(B_1) = P(A(1)_1) \cdot \alpha_{A(1)} + P(A(0)_1) \cdot \alpha_{A(0)} + \\ - P(A(1)_1) \cdot \alpha_{A(1)} \cdot P(A(0)_1) \cdot \alpha_{A(0)} \end{aligned} \tag{2.9}$$

That is, $P(B_1)$ is the sum of the probabilities of the constituents (i.e., events $A(0)$ and $A(1)$) minus the probability of the intersection, according to the rule of the probability of unions. The terms $P(A(0)_1)$ and $P(A(1)_1)$ are the observable probabilities of the presence of events $A(0)$ and $A(1)$, respectively. These probability terms can be used to represent the presence (i.e., when $P(A(k)_1) = 1$) or absence (i.e., when $P(A(k)_1) = 0$) of the events $A(0)$ and $A(1)$. The causal powers $\alpha_{A(0)}$ and $\alpha_{A(1)}$ correspond to the causal strengths of the background event $A(0)$ and the event $A(1)$, respectively.

The probability of event B given the presence of event $A(1)$, denoted as $P(B_1|A(1)_1)$, can be derived by conditioning Equation 2.9 on event $A(1)$ being present — implying that the term $P(A(1)_1)$ is equal to 1:

$$\begin{aligned} P(B_1 \mid A(1)_1) = \alpha_{A(1)} + P(A(0)_1 \mid A(1)_1) \cdot \alpha_{A(0)} + \\ - \alpha_{A(1)} \cdot P(A(0)_1 \mid A(1)_1) \cdot \alpha_{A(0)} \end{aligned} \tag{2.10}$$

Similarly, we can derive the probability of event B given the absence of event $A(1)$, denoted as $P(B_1|A(1)_2)$, by conditioning Equation 2.9 on event $A(1)$ being absent — implying that the term $P(A(1)_1)$ is equal to 0:

$$P(B_1|A(1)_2) = P(A(0)_1 \mid A(1)_2) \cdot \alpha_{A(0)} \tag{2.11}$$

The quantity of interest, causal power $\alpha_{A(1)}$, can then be found:

$$\begin{aligned} \alpha_{A(1)} = & \frac{P(B_1 \mid A(1)_1) - P(B_1 \mid A(1)_2)}{1 - P(A(0)_1 \mid A(1)_1) \times \alpha_{A(0)}} + \\ & - \frac{[P(A(0)_1 \mid A(1)_1) - P(A(0)_1 \mid A(1)_2)] \times \alpha_{A(0)}}{1 - P(A(0)_1 \mid A(1)_1) \times \alpha_{A(0)}} \end{aligned} \tag{2.12}$$

Equation 2.12 calculates the causal power of the event $A(1)$, yet it necessitates certain quantities that remain inaccessible or unobservable. For instance, the term $\alpha_{A(0)}$, representing the causal power of the composite alternative event $A(0)$, is itself, like all causal powers, inherently unobservable (Luhmann & Ahn, 2005). Consequently, the direct application of Equation 2.12 is impractical due to these limitations in observable data. However, when the occurrence of the candidate cause $A(1)$ is independent of the occurrence of the alternative cause $A(0)$, a condition mathematically expressed as $P(A(0)_1|A(1)_1) = P(A(0)_1|A(1)_2) = P(A(0)_1)$, Equation 2.12 simplifies as:

$$\alpha_{A(1)} = \frac{P(B_1 \mid A(1)_1) - P(B_1 \mid A(1)_2)}{1 - P(B_1 \mid A(1)_2)} \tag{2.13}$$

As we already defined in Section 2.4, ΔP is obtained by subtracting $P(B_1 \mid A(1)_2)$ from $P(B_1 \mid A(1)_1)$ (see Equation 2.1), so that Equation 2.13 can be expressed as:

$$\alpha_{A(1)} = \frac{\Delta P}{1 - P(B_1 \mid A(1)_2)} \tag{2.14}$$

Equation 2.14 indicates when and how well ΔP gives an estimate of $\alpha_{A(1)}$, and it relates causal power to probabilities that are observable, allowing the estimation of the term $\alpha_{A(1)}$. The power PC model bases its predictions on causal powers, which, in general, only partly determine ΔP (Cheng, 1997).

In conclusion, the power PC model is a parameter estimation model (Perales & Shanks, 2007) that provides normative values for optimal causal inference (Matute et al., 2022).

2.6.2 Algorithmic models

Algorithmic models propose the use of chains of algorithms to describe the psychological processes underlying causal induction. These models are also referred to as *non-normative* models, in the sense that they are not bound to any particular norm of rationality (Perales & Shanks, 2007), and, for that reason, they can allow researchers to mathematically predict the emergence of biased evaluations of causality (Matute et al., 2022).

A subset of this model family is the *rule-based* models (Perales & Shanks, 2007), for which individuals track the different frequencies or probabilities of scenarios presented during learning trials and follow specific rules to estimate the causal link between events. However, in these models, either a different rule than that of a normative model is applied, or the tracked probabilities are assigned different weights, meaning that some pieces of information are considered inherently more important than others (Matute et al., 2022).

For instance, a measure of contingency that emerged in the 1950s and can be classified within the rule-based model family was proposed by Inhelder and Piaget (1958/2013). They suggested that people compute the difference between the diagonals of the tetrachoric table (see Table 2.1) to quantify the correlation between events A and B[1]:

$$\Delta D = (a + d) - (b + c) \tag{2.16}$$

where a, b, c, and d represent the observed frequencies of the four scenarios depicted in Table 2.1. The term $(a + d)$ denotes the sum of the frequencies where events A and B either both occur or both do not occur, while $(b + c)$ represents the sum of the frequencies where the occurrence of events A and B does not align. The psychological rationale for employing this correlation method is based on the idea that individuals evaluate the evidence confirming the existence of a causal link and compare it with the evidence that disconfirms such a link.

[1] As noted by Allan (1980), another measure of contingency that resembles ΔD was proposed by Smedslund (1963) in a study examining the naive concept of correlation on a tetrachoric table. Following Allan (1980)'s notation, measures of correlation can be defined as the ratio of diagonals in Table 2.1:

$$\Delta R_1 = \frac{a + d}{b + c} \quad \text{and} \quad \Delta R_2 = \frac{a + d}{N} \tag{2.15}$$

where a, b, c, and d are the observed frequencies of the four scenarios represented in Table 2.1, and N is the total sum of frequencies.

While ΔP is a normative probabilistic measure, emphasizing the difference in the likelihood of an outcome occurring with versus without the cause, ΔD is a frequency-based measure that reflects the raw difference between confirming and disconfirming cases. In summary, ΔP is grounded in probability theory, whereas ΔD directly addresses the differences in frequencies across the diagonals of the table.

Allan (1980) pointed out the inadequacy of ΔD as a normative contingency measure, summarizing Jenkins and Ward (1965)' study, which showed that when $\Delta P = 0$, $\Delta D = 0$ only when marginal column and/or row frequencies are equal (i.e., $a + b = c + d$ and/or $a + c = b + d$). If one of these conditions is not met, then even when the relationship between events A and B is absent, ΔD could still not be equal to 0, leading to an invalid conclusion. For that reason, rather than a norm-based model, ΔD should be considered an algorithmic one.

Focusing instead on weighted rule-based models, using again the ΔP rule as a guiding example, it has been proposed to correlate causal judgment with a weighted version of ΔP (Allan, 1993) rather than an unweighted *classical* version (see Section 2.4). Specifically, cells in Table 2.1 could be weighted such that $P(B_1|A_1)$ is given more weight than $P(B_1|A_2)$ (Perales & Shanks, 2007). In this way, it should be noted that the ΔP rule is no longer considered a normative model, but rather a modified version where different weights are assigned on the tetrachoric table (e.g., $w_a > w_b > w_c > w_d$) in order to better predict and explain accurate and biased causality judgments.

Within the set of algorithmic models, we can also include *associative models*, which assume that causal links are learned by the functioning of an associative mechanism that accumulates associative strength between the events.

2.6.3 Rescorla-Wagner model

A standard associative model is the *Rescorla-Wagner model* (RWM; Rescorla & Wagner, 1972), a simple yet powerful and elegant mathematical explanation for how associations are formed and adjusted based on experience.

Causal learning can be viewed as a type of associative learning to which the RWM can be applied (Pearce & Bouton, 2001). Since the RWM has been studied in contexts involving multiple cues (i.e., multiple events A, whether presented simultaneously or not), illustrating how these cues compete for predictive power with respect to their outcome (Chapman & Robbins, 1990), we will focus on an example where two events A can be present. Consider two potential causes, events $A(1)$ and $A(2)$, an outcome event B, and a context X. Here, X represents a set of background events that are always present and can be associated with the outcome event B just like any other event A (Pearce & Bouton, 2001).

In a learning context, event B can be preceded by either $A(1)$, $A(2)$, both, or neither. According to the RWM, the change in associative strength for events $A(1)$, $A(2)$, and the context X with respect to event B is updated on each learning trial using the error-correction rule:

$$\Delta V_{A(k)} = \alpha_{A(k)}\beta(\lambda - \sum_{j\in A} V_j) \tag{2.17}$$

Where:

- $\Delta V_{A(k)}$ is the change in associative strength between event $A(k)$ (where k can be 1 or 2 in this case) or context X and event B on a given trial.
- $\alpha_{A(k)}$ is the learning rate parameter specific to each event $A(k)$ (where k can be 1 or 2 in this case) or context X, which assumes values from 0 to 1.
- β is the learning rate parameter of event B, which assumes values from 0 to 1.
- λ is the maximum associative strength that event B will support. λ is set equal to 0 when event B is absent and is set equal to 1 when event B is present.
- $\sum_{j\in A} V_j$ is the sum of the associative strengths of the events $A(k)$ (where k can be 1 or 2 in this case) and context X that are present on a given trial.

When all events A (i.e., $A(1)$ and $A(2)$) are present, the sum of their associative strengths is:

$$\sum_{j\in A} V_j = V_{A(1)} + V_{A(2)} + V_X \tag{2.18}$$

A change in associative strength is calculated for each event A present in a trial. The changes in associative strength for a generic event $A(k)$ depend on the sum of the associative strengths of all events $A(k)$ present ($A(1)$ and/or $A(2)$, and X in this case). Consequently, if one generic event $A(k)$ acquires associative strength, any other event $A(\neg k)$ is less likely to acquire associative strength (Baker et al., 1996), as at the asymptote (i.e., the point at which no more learning occurs), the sum of all associative strengths will be, following our example:

$$V_{A(1)} + V_{A(2)} + V_X = \lambda \tag{2.19}$$

This principle leads to what is known as *cue competition* (Baker et al., 1996): if one event $A(k)$ acquires associative strength, the others will not. When only one event A (i.e., $A(1)$ or $A(2)$) and one event B are involved, and the learning parameters β are assumed to be equal for both the presence and absence of an effect, the associative strength of A converges to the ΔP value at the asymptote (see the Appendix for the demonstration).

In the context of causal learning, the RWM can predict both correct and biased contingency estimations (Matute et al., 2022).

Matute et al. (2019) showed that in a setting with only one event A and context X competing for associative strength with respect to an event B in a null contingency illusory condition, the RWM algorithm predicts that the association between events A and B may initially increase above 0. This occurs because context X acquires associative strength more slowly than event A, as context X is less salient (i.e., the learning parameter α_X is smaller than the learning parameter α_A). Due to the coincidence of events A and B, and the initially weak association between context X and event B, participants in the early phase of causal learning are expected to exhibit the illusion of causality, as the association between events A and B becomes stronger. However, as more information is acquired over a certain number of trials, the associative strength between context X and event B increases, while the strength between events A and B weakens, eventually approaching 0. At this stage, participants are expected to reduce the illusion of causality, as the association between events A and B diminishes. Thus, the RWM predicts that after enough trials, there should be a convergence towards the correct contingency value, in accordance with the ΔP rule. In summary, the RWM suggests that causality biases are pre-asymptotic, meaning they are expected to occur primarily during the initial trials. However, Barberia et al. (2019) challenged this prediction by finding that participants exposed to numerous trials did not show a reduction in the illusion of causality.

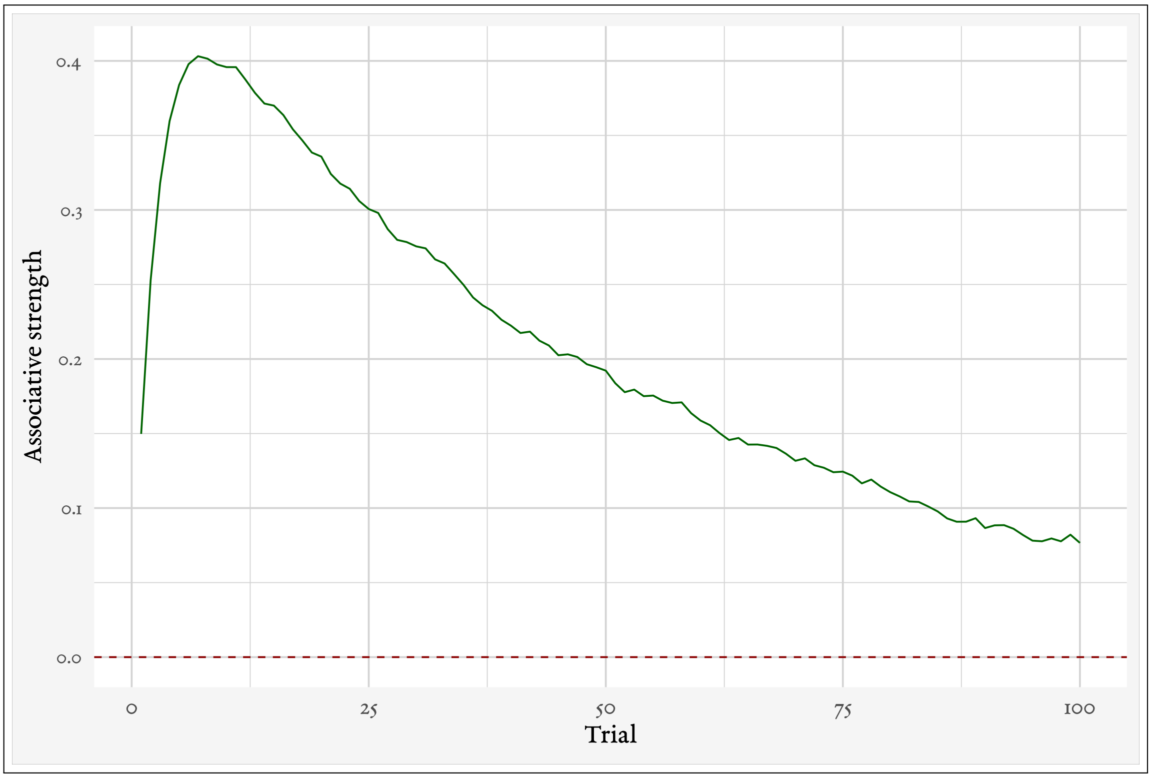

Figure 2.1: RWM pre-asymptotic causality bias simulation.

Figure 2.1 shows a graphical result from an RWM simulation conducted in R (R Core Team, 2022)[2] that replicates the RWM simulation by Matute et al. (2019), specifically in the context of an illusory condition where the frequencies of both the cue (i.e., event A) and the outcome (i.e., event B) are jointly increased (i.e., when the frequency of cell a in Table 2.1 is increased, while ΔP remains 0; see Section 2.4). The simulation has been obtained using the Equations 1, 2, 3, 4 in the Appendix. The green line shows the progression of associative strength between events A and B during trials, highlighting the pre-asymptotic bias prediction by the RWM, whereas the red dotted line represents ΔP.

[2]Learning parameters of the simulation (10^4 iterations): $\alpha_A = 0.3, \alpha_X = 0.1, \beta_A = \beta_X = 0.8$.

3

Interaction with process fluency

3.1 Modulating the illusion

In Chapter 2, we asserted that the illusion of causality can be understood as a phenomenon that consistently arises under certain circumstances. In a null contingency condition, where the ΔP index rule (see Section 2.4) would correctly advise rejecting a causal connection, participants still express a positive causal evaluation between the events when specific conditions are met (i.e., when cause or outcome frequencies are increased). By analyzing the contexts in which this bias is either heightened or reduced, we can gain valuable insights into the underlying mechanisms and cognitive processes responsible for the illusion.

First, we can inquire whether there are individual differences in the manifestation of the illusion. According to Matute et al. (2015), the illusion is not related to intelligence or personality traits. Instead, it arises from how the mind has evolved to discern causality from contingencies, with associative processes (i.e., conditioning) playing a fundamental role. However, it should not be assumed that associative learning is a simple mechanistic phenomenon. Even the most basic associative learning involves complex cognitive processes (Cornoldi et al., 2018). Furthermore, our data analyses (see Section 6.3) indicate that demographic features do not lead to variations in the magnitude of the causality bias.

Contrary to the notion of a consistently occurring phenomenon, it has been shown that certain variables can alter the illusion of causality, with mood being a notable example (Matute et al., 2015). Blanco et al. (2012), in their examination of the illusion of causality within an active procedure (i.e., the illusion of control, see Section 2.3), found that scores on depressive symptom scales can mediate the phenomenon. Consistent with the findings of Alloy and Abramson (1979), individuals with higher depressive symptoms were more accurate in their causal judgments under null contingency conditions. However, an explanation of these results suggests that this tendency may be linked to the type of procedure used (i.e., active vs. passive). Depression might reduce the tendency to initiate voluntary responses (i.e., choosing not to introduce the cause in the active procedure), whereas non-depressed participants acted with greater frequency than depressed participants to obtain the outcome (i.e., introducing the cause in the active procedure). As a result, non-depressed participants were exposed to a higher number of cause-outcome coincidences. Although this explanation pertains specifically to the active procedure, these findings underscore the significant role that the probability of the cause plays in either enhancing or reducing the illusions of causality (Matute et al., 2015).

Furthermore, the tendency to jump to conclusions (i.e., deriving conclusions based on scarce data) has been shown to mediate the effect of the illusion of causality. Participants with higher scores on this tendency made higher causal judgments in a null contingency condition (Moreno-Fernández et al., 2021). Additionally, attitudes and preferences also seem to play a role: the illusion is enhanced when tasks are presented in a framework that aligns with personal preferences and inclinations (Matute et al., 2022).

A modulation in the illusion of causality has also been observed in contexts where another potential cause is available. Vadillo et al. (2013) found that participants informed about a potential alternative explanation for the outcome showed a reduced illusion of causality compared to the group that received no suggestions about alternative explanations. Thus, informing people about the existence of alternative causes can mitigate the illusion. Nonetheless, the presence of an alternative cause can sometimes lead to erroneous conclusions. Yarritu et al. (2015) found that presenting an illusory-cause condition (i.e., $\Delta P = 0$) before an effective-cause condition (i.e., $\Delta P > 0$) could reduce the ability to detect the causal link in the latter condition.

Lastly, some research has specifically targeted bias reduction, hypothesizing and exploring whether the causality bias can be diminished, as we will discuss in the next Subsection (3.1.1).

3.1.1 Focusing on the illusion reduction

Matute et al. (2022) argued that the illusion of causality is a consistent and pervasive phenomenon with significant consequences in various domains, such as health and politics. Many social judgments and behaviors are influenced by intuitive evaluations of causal relationships between events (Crocker, 1981). Moreover, it has been suggested that the illusion is linked to pseudoscientific thinking and beliefs (Matute et al., 2011). Griffiths et al. (2019) showed that individuals prone to superstitious beliefs are also susceptible to the causality bias.

Given these concerns, researchers have focused on developing strategies to reduce the causality bias (Matute et al., 2022). Some efforts have been made to create psychoeducational interventions aimed at eliminating cognitive biases or diminishing their intensity and frequency (Lilienfeld et al., 2009), though the effectiveness of debiasing techniques remains debated (Arkes, 1981). Specific interventions targeting the reduction of causality bias have been proposed by Barberia et al. (2013; see also Barberia et al., 2018), who provided participants with explicit instructions on how to counteract the bias and think in a more scientific manner.

A reduction in the illusion has also been observed in studies that specifically manipulated the information presented in the classic CLT (see Section 2.3), as we will discuss in greater detail in the next Sections (3.2 and 3.3).

3.2 Foreign language effect

In investigating ways to indirectly reduce the causality bias, Díaz-Lago and Matute (2019b) found that a group of participants who performed the CLT in a foreign language (FL) exhibited a reduced effect of the illusion of causality compared to a group that conducted the task in their native language (NL). This result aligns with the *foreign language effect* (FLE), a phenomenon for which an increasing body of literature has shown that conducting a task in a context of a FL can affect decision-making outcomes (Circi et al., 2021).

The FLE was first described by Keysar et al. (2012), who found that participants exposed to the *asian disease dilemma* in a FL exhibited less biased responses than those who conducted the task in their NL. The asian disease problem is a task introduced by Tversky and Kahneman (1981) to study how decision-making is influenced by the way choices are presented (i.e., the way choices are *framed*). In this task, participants are asked to choose between a safe option and a risky option to deal with a hypothetical epidemic outbreak. This epidemic is expected to kill 600 people, and participants have to choose between two programs to combat the disease. Two conditions are compared: one with options presented in a *gain* frame and another with options presented in a *loss* frame. In the gain frame, participants choose between two options: $P1$, for which 200 people will be saved, and $P2$, for which there is a $1/3$ probability of saving 600 people and a $2/3$ probability of saving no one. In the loss frame, participants choose between two options: $P1$, for which 400 people will die, and $P2$, for which there is a $1/3$ probability that no one will die and a $2/3$ probability that all people will die.

When the choices are framed in terms of gains, the majority of people tend to exhibit risk-averse behavior (i.e., favoring option $P1$), whereas when the choices are framed in terms of losses, the majority of people tend to exhibit risk-seeking behavior (i.e., favoring option $P2$). The normative and expected values for options $P1$ and $P2$ are the same in the gain frame and in the loss frame. Therefore, the difference in responses between the two framings is considered a violation of the rules of rational choice (i.e., a biased response; Circi et al., 2021). Keysar et al. (2012) found that when the asian disease problem was presented in a relatively low-proficient FL, the effect of framing options in terms of gains or losses was reduced, and participants tended to choose the risk-averse option (i.e., $P1$) in both conditions to a similar extent.

At first glance, this seems like a counterintuitive result, as we would expect that the use of a FL could potentially increase cognitive difficulty, thereby promoting heuristics rather than reducing them (Keysar et al., 2012). Nonetheless, over the years, the FLE has proven consistent across different tasks in loss-aversion paradigms, decision-making, and moral dilemmas (Circi et al., 2021). For instance, Costa et al. (2014) replicated the findings from Keysar et al. (2012), extending the evidence of the phenomenon to other heuristics, and showing how decision-making, when problems are presented in a FL, is less subject to biases. Furthermore, in the area of moral dilemmas, the FLE seems to promote more *utilitarian responses*, which are judgments aimed at maximizing benefits and minimizing costs across affected individuals, in contrast with *deontological responses*, which are judgments aimed at following specific duties regardless of the consequences (e.g., Costa et al., 2014; Geipel et al., 2015; Cipolletti et al., 2016).

Without providing an extensive review of the discussion on the psychological reasons why the FLE occurs, we will briefly introduce two of the main ideas (see Subsections 3.2.1 and 3.2.2) that can also be linked in some ways to the dual-process models (see Section 1.4).

3.2.1 Emotional explanation

A first explanation of the FLE was proposed in the original study by Keysar et al. (2012). The authors argued that the FLE arises because using a FL creates psychological distance from the emotional intensity typically associated with one's NL. This psychological distance is believed to attenuate emotional reactions, thereby reducing the influence of biases that are often emotionally driven. As discussed in the meta-analysis by Del Maschio et al. (2022), the role of emotions in decision-making under either a NL or a FL can be understood in two ways: the NL may promote emotional responses that can lead to intuitive reasoning and biased decisions, or the FL may attenuate emotional responses, provoking the same effect.

Within the field of moral dilemmas, Greene et al. (2001; see also Greene et al., 2004) proposed a domain-specific dual-process model (Craigie, 2011), which assumes the existence of two cognitive subsystems that are in competition during moral reasoning tasks. The first subsystem is emotionally driven, rapid, and automatic, while the second is deliberative, slow, and effortful. There are undoubtedly certain analogies between the dual-process models discussed in Section 1.4 and Greene et al. (2001)' model. However, it is important to exercise caution when comparing these models, given that the intuitive processes underlying system one, as presented in Section 1.4, and the emotionally driven processes of the first subsystem introduced by Greene et al. (2001) are conceptually distinct (Craigie, 2011), and there are structural differences between the models (Haidt, 2001). Nonetheless, in a broader sense, we can cautiously suggest that this hypothesis implies either that using an FL reduces reliance on system one or that the use of an NL prompts the engagement of system one, as conceptualized in Greene et al. (2001)' model.

3.2.2 Cognitive explanation

Kahneman and Frederick (2002), referencing the dual-process models discussed in Section 1.4, suggests that contextual factors that elevate mental stress or cognitive load can significantly influence which system becomes more dominant in decision-making. Specifically, increased mental stress or cognitive load (through, for example, the disruption of process fluency on a task; see next Section 3.3), can enhance the reliance on system two processes, thus leading to more deliberate and analytical thinking (Del Maschio et al., 2022). Conversely, these psychological conditions may reduce the influence of system one, which operates through automatic and heuristic-based judgments. In the context of the FLE, this theoretical framework suggests that using a FL could shift decision-making away from system one thinking towards greater reliance on system two processes (Costa et al., 2014). When individuals engage in decision-making tasks in a FL, the cognitive demands of processing a non-native language may increase cognitive load and necessitate more effortful, analytical processing, characteristic of system two, consequently leading to more reasoned and less biased decisions. However, as noted by Del Maschio et al. (2022) and Circi et al. (2021), this explanation is not without limitations. Some studies have shown that a FL context does not necessarily reduce cognitive biases when participants are presented with emotionally neutral tasks (e.g., Geipel et al., 2015; Vives et al., 2018). These findings suggest that the relationship between FL usage and reduced cognitive biases may be more complex and context-dependent than initially assumed.

3.3 Process fluency

Another study, conducted by the same researchers who found that the illusion of causality can be reduced when the task is conducted in a FL (Díaz-Lago & Matute, 2019b), showed that superficial aspects of the information presented, such as the font in which text is displayed, can influence the illusion of causality (Díaz-Lago & Matute, 2019a). In their experiment, participants engaged in a CLT where scenario frequencies, as depicted in Table 2.1 (see Section 2.3), were manipulated to induce an outcome-density bias (see Section 2.4). Participants were randomly assigned to one of two conditions: one group completed the task in an easy-to-read font, while the other completed it in a hard-to-read font. The results showed a significant reduction in the causality bias among participants in the hard-to-read font condition compared to those in the easy-to-read condition. Notably, this difference was not only statistically significant but also of *medium effect size* (see Section 4.2), indicating a substantial impact from this seemingly minor change in the presentation of information.

To explain the influence of font type on the illusion of causality observed by Díaz-Lago and Matute (2019a), it is helpful to build upon the framework discussed in Subsection 3.2.2 regarding the cognitive explanation of the FLE. In particular, the broader concept of *processing fluency* (PF; see next Subsection 3.3.1) can shed light on how changes in cognitive effort may alter decision-making and provoke a reduction in the illusion of causality.

3.3.1 A tribe of fluency

PF refers to the subjective ease with which information is processed (Oppenheimer, 2008), serving as a metacognitive cue that significantly influences judgments and decision-making (Alter & Oppenheimer, 2009). Thought processes are accompanied by the metacognitive experience of the ease or difficulty with which mental representations can be retrieved by the mind, or the fluency or disfluency with which new information can be processed (Schwarz, 2004). These experiences are in themselves informative to some extent, as the knowledge of processing ease or difficulty can lead to useful inferences about the external environment, showing how human judgments reflect not only the content of thoughts but also the metacognitive experience of processing those thoughts (Alter & Oppenheimer, 2009). People form *naive theories* about the causes of their fluency experiences, which, in turn, guide how fluency influences domain-specific judgments (Schwarz, 2004). These naive theories are shaped by past experiences and the current context (Oppenheimer, 2008). For instance, if a written text is syntactically complex and difficult to read – therefore disfluent – it may lead readers to classify the text as overly complex and unpleasant. Conversely, applying the same level of syntactical complexity to poetry might make it seem more intricate and interesting.

As argued by Oppenheimer (2008), the effects of processing (dis)fluency can be generated by a wide array of cognitive processes, making it a difficult construct to capture. Any variable capable of altering processing fluency could potentially lead to similar effects from a cognitive standpoint (Schwarz, 2004). Moreover, every psychological experimental task can be described on a continuum that ranges from effortless to effortful, resulting in a corresponding metacognitive experience that spans from fluent to disfluent (Alter & Oppenheimer, 2009). Thus, fluency experiences can arise as a byproduct of diverses cognitive processes, as illustrated in Figure 3.1, which has been adapted from the review by Alter and Oppenheimer (2009).

PF has been shown to influence judgments across a wide array of domains. For instance, Schwarz et al. (1991), in a study on retrieval fluency, asked one group of participants to recall 6 examples of assertive behavior (an easy task) and another group to recall 12 examples (a more difficult task). Participants who had to generate many examples found the process more difficult than those who had fewer examples to retrieve. Subsequently, participants rated their own assertiveness. Results showed that their evaluations were based on how easily examples of assertive behavior came to mind rather than on the number of examples they had generated.

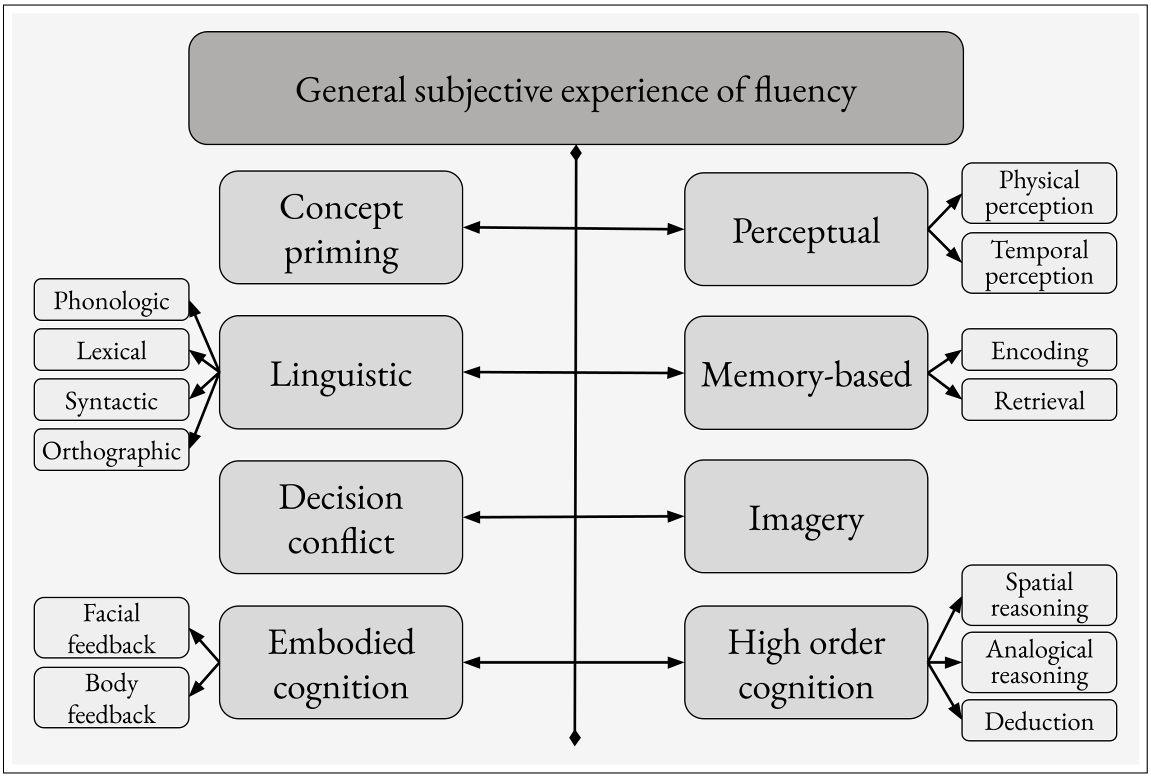

Figure 3.1: Various instantiations of fluency.

PF has also been studied in relation to other phenomena. For instance, Reber and Schwarz (1999) showed that fluent statements are judged as more likely to be true than disfluent ones, as people tend to associate fluency with truth and disfluency with falsehood (Schwarz, 2004). Moreover, fluency is linked to judgments of liking (Alter & Oppenheimer, 2009). Reber et al. (1998) found that participants rated fluent stimuli presented against highly contrastive backgrounds as more aesthetically pleasing than identical stimuli against less contrastive backgrounds. Additionally, research has shown that greater confidence in responses to questions is based on the ease (i.e., fluency) with which those responses come to mind (Kelley & Lindsay, 1993).

Studies on PF have also extended beyond basic research, including applications in fields such as marketing. For example, a study by Novemsky et al. (2007) indicates that the presentation conditions of a product, which promote either fluency or disfluency, can play a key role in determining whether a consumer makes a purchase. Specifically, when consumer products are made disfluent, consumers are more likely to defer choice or opt for a default option compared to when product names are fluently processed.

3.3.2 PF in relation to biases

PF is an influential factor in cognitive processing that can also be interpreted within the framework of dual-process models (Oppenheimer, 2008). For instance, it has been suggested that experienced difficulty (i.e., disfluency) on a task can prompt individuals to adopt a more systematic approach to information processing. Alter et al. (2007) conducted an experiment in which participants took the *cognitive reflection test* (CRT; Frederick, 2005), a measure of the extent to which individuals rely on system one processing. The test includes items where an intuitive mode of reasoning (i.e., system one) leads to incorrect answers, but participants can override these initial responses by engaging in more deliberate, analytical reasoning (i.e., system two). The study found that participants who took the CRT in a hard-to-read font provided more correct answers than those who took it in an easy-to-read font, suggesting that disfluency led to a shift towards more systematic processing strategies.

The relationship between PF and the modulation of cognitive biases is well-supported across various domains. For example, increased disfluency has been shown to reduce susceptibility to cognitive errors like the *Moses illusion* (Song & Schwarz, 2008) and to weaken the framing effect (Korn et al., 2018). These findings highlight how lower fluency can lead to more effortful and less biased reasoning by promoting a shift from intuitive to analytical processing.

Given these insights, we can hypothesize that high fluency, characterized by an effortless cognitive experience, reinforces intuitive judgments aligned with system one processes. In contrast, disfluency, marked by a challenging cognitive experience and increased cognitive load, prompts more analytical and deliberate thinking associated with system two processes (Kahneman & Frederick, 2002). In the study by Díaz-Lago and Matute (2019a), the use of a hard-to-read font likely induced perceptual disfluency (see Figure 3.1), leading participants to rely less on automatic, heuristic-based reasoning (system one) and more on deliberate, systematic processing (system two). This shift in cognitive processing could explain the observed reduction in the illusion of causality in the hard-to-read font condition. Considering that disfluency can be triggered by various alterations that increase cognitive load and task difficulty, it is plausible that the reduction in the illusion of causality observed in both a hard-to-read font context (Díaz-Lago & Matute, 2019a) and a FL context (Díaz-Lago & Matute, 2019b) may stem from a common underlying mechanism involving heightened task difficulty.

3.4 Study hypotheses

While the connection between PF and the illusion of causality is compelling, it is important to remark the multidimensional nature of the PF construct. As we exposed in Subsection 3.3.1, PF encompasses various manipulations that affect the subjective ease of information processing. However, these manipulations may target different stages and types of cognitive processing. For example, manipulating language fluency, as in the FLE, primarily affects syntactic and semantic processing, which occurs at a later stage of text comprehension, whereas manipulating font type influences only the perceptual processing of written text, an earlier stage of comprehension. Consequently, while both forms of fluency manipulation might shift processing from system one to system two, the underlying cognitive processes involved are distinct and warrant further empirical investigation. The study by Díaz-Lago and Matute (2019a) suggests that perceptual disfluency, manipulated through font type, can reduce the magnitude of the illusion of causality. Building on this finding, our first experiment aimed to test whether perceptual disfluency, manipulated through the contrast of the written stimuli with the background, similarly affects the illusion of causality. Thus, we formulated two hypotheses:

- *H1:* The illusion of causality will be reduced in a low-contrast condition compared to a high-contrast condition in the CLT. Specifically, the mean judged strength of the cause-effect relationship in the low-contrast condition is expected to be less than in the high-contrast condition.

- *Ho:* The mean judged strength of the cause-effect relationship in the low-contrast condition will be equal to the mean judged strength in the high-contrast condition.

In the next Chapter (4), we will present the design analysis, participant recruitment, and experimental procedure for our first experiment. In Chapter 5, we will detail the main results related to our hypotheses. Given the null result observed, we decided to conduct a second experiment to further investigate the phenomenon. We will discuss this experiment and its results in Chapter 6. Finally, in Chapter 7, we will offer our analysis and interpretation of the findings.

4

First experiment: structure

4.1 Conditions

Within the context of a classic CLT paradigm (see Section 2.3), we manipulated the physical-perceptual characteristics of the stimuli by adjusting the contrast of the written texts against the background, using different colors on a white background. This manipulation aimed to alter the perceptual fluency with which participants could perceive the stimuli and perform the task. As shown by Reber and Schwarz (1999), contrast manipulation is a reliable source of variation in perceptual fluency and has been successfully employed in other PF studies (e.g., in a research on judgments of agent competence; Thompson & Ince, 2013).

We developed an online version of a standard CLT, characterized by two conditions: *high contrast* (HC), where the written stimuli were displayed in dark blue on a white background, and *low contrast* (LC), where the written stimuli were displayed in light yellow on a white background. The colors were chosen based on the PF literature (e.g., Reber & Schwarz, 1999; Thompson & Ince, 2013). Yellow text on a white background has been shown to induce significant disfluency, whereas blue text on a white background has generally been shown to be easy-to-read.

Additionally, we consulted the *human-computer interaction* literature in order to determine the precise *hexadecimal* (HEX) color codes for the online experiment (e.g., Hall & Hanna, 2004; Zuffi et al., 2007). For the LC condition, we initially selected HEX #FFFF00 (i.e., yellow), which appeared sufficiently disfluent. However, due to variability in contrast across different devices, we empirically observed that the contrast ratio needed adjustment. Consequently, we further decreased the contrast ratio by using HEX #FFFF73 (i.e., a lighter yellow).

Another initial idea was to randomize the color of the text on a given randomized colored background, maintaining a fixed contrast ratio between the background and the stimuli. This approach aimed to evaluate the effect of contrast independently of the specific colors used. However, we decided against this method to avoid introducing a confounding variable related to polarity (i.e., dark text on a light background vs. light text on a dark background). This decision ensured that our manipulation strictly focused on the contrast and its effects on perceptual fluency without introducing additional variables that could affect the outcomes.

To test whether the contrast manipulation affected the magnitude of the causality bias, we combined the two contrast conditions with two different contingency conditions in a 2 (*contrast*: HC vs. LC) x 2 (*contingency*: *true contingency* vs. *null contingency*) between-subjects factorial design. The true contingency condition was characterized by a positive ΔP, indicating an actual causal link between the cue and the outcome event. The null contingency condition, characterized by a null ΔP, indicated the absence of a real causal link between the cue and the outcome event. An outcome-density bias was induced (see Section 2.4), as previous studies (e.g., Díaz-Lago & Matute, 2019a, 2019b) have shown this condition to give rise to a robust illusion of causality. If perceptual disfluency, induced by the LC condition, prompts system two reasoning and reduces the magnitude of the illusion of causality, then in the null contingency condition, lower causality ratings are expected in the LC condition compared to the HC condition. As shown by previous studies (Díaz-Lago & Matute, 2019a, 2019b), no effect of perceptual disfluency is expected in the true contingency condition, which served as our control group.

To ensure transparency and replicability, we pre-registered the hypotheses (see Section 3.4), along with the design features, on the Open Science Framework (OSF) website at the following link: https://osf.io/74d6g. The codes for the experiment, the raw data, and the script used for the main analyses are available on OSF at the following link: https://osf.io/c26qa.

4.2 Participants

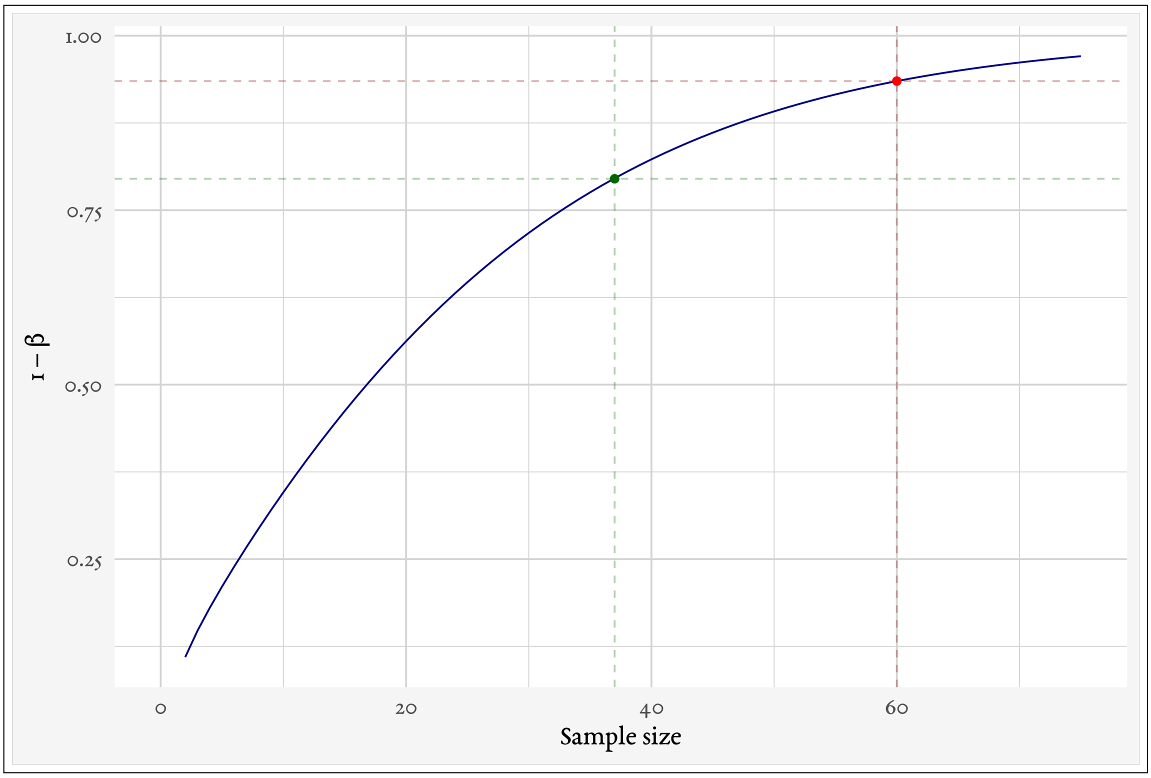

Figure 4.1: Simulated power function for the critical comparison.

To determine the appropriate sample size, a *power analysis* must be conducted (see Figure 4.1). Through power analysis, we can ascertain the required sample size N based on fixed values for the probability of a *type I* error (α), the probability of a *type II* error (β) and its complement (power; $1 - \beta$), and the effect size, which reflects the magnitude of the difference between central tendency measures and their variability. We performed a power analysis using the *pwr* package (Champely, 2020) in *R* (R Core Team, 2022). For the hypothesized effect size, we referred to the results of a previous study by Díaz-Lago and Matute (2019a), which reported a medium effect size ($d = 0.58$) for the difference in the rated strength of cause-effect relationships between two conditions (i.e., easy-to-read font vs. hard-to-read font) in the null contingency condition. We set the power to 0.8 to detect this medium effect size of 0.58 with a standard α error probability of 0.05.

The result of the analysis was $N = 37$, with N referring to the number of participants in each group (see the green dot in Figure 4.1). Given that there were four groups (i.e., 2 contingency × 2 contrast), a total of 148 participants was necessary. With this sample size, as confirmed by a prospective design analysis using the *PRDA* package (Callegher et al., 2021) in R, our *type* M error was 1.16 and our *type* S error was 0 [1].

To further increase the statistical power of the critical comparison between the two contrast conditions (HC vs. LC) in the null contingency condition, we decided to collect 60 participants for each of these two groups, raising the statistical power above 0.9 (see the red dot in Figure 4.1). This result was confirmed by a retrospective design analysis with the *PRDA* package in R, which showed an M error of 1.04 and an S error of 0. For the two true contingency groups (HC vs. LC), we recruited 40 participants each, targeting a total of 200 participants. Thus, participants were randomly assigned to the four conditions in the following proportion: 3 (null contingency, LC): 3 (null contingency, HC): 2 (true contingency, LC): 2 (true contingency, HC).

Participants were recruited through various advertisements on social networks and university flyers. Those who agreed to participate had a chance to win 25 euros through a lottery, with 6 prizes distributed randomly among those who completed the experiment.

A total of 209 participants took part in the experiment. However, data from 9 participants were excluded based on the following *a priori* exclusion criteria:

- ❖ Completing the experiment twice (5 participants).
- ❖ Reading the fictitious story in less than 10 seconds (4 participants).
- ❖ Completing the experiment in less than 180 seconds (0 participants).
- ❖ Completing the trial section in less than 160 seconds (0 participants).
- ❖ Responding to each of the two final questions in less than 2 seconds (0 participants).

The final sample consisted of 142 females and 58 males, with an average age of 25.92 years ($SD = 9.59$). A Pearson's chi-square test of independence showed no significant difference in the distribution of sexes across the four groups, $\chi^2(3) = 3.99, p = 0.26$. Given the positive skewness in the age distribution (see Figure 4.2), a Kruskal-Wallis test was conducted to assess age differences across groups. The test indicated no systematic differences in age across the four groups, $\chi^2(3) = 3.20, p = 0.36$.

[1] Type S error refers to the probability of obtaining a statistically significant result in the opposite direction to the plausible one, while type M error represents the factor by which a statistically significant effect is, on average, exaggerated.

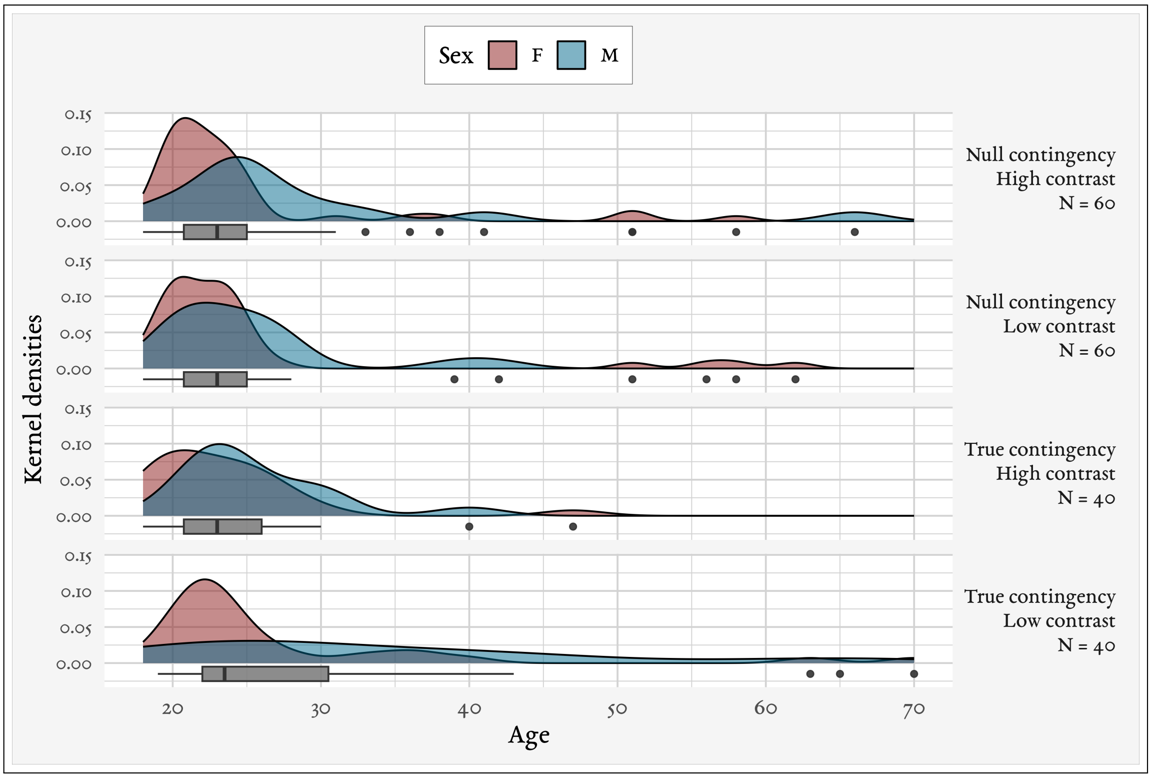

Figure 4.2: Participants' age and biological sex distribution.

4.3 Procedure

Before starting the experiment, participants read the online informed consent form approved by the *Ethics Committee for Psychological Research* at the *University of Padova* [2] and then gave their consent to participate through a response key.

Participants were automatically assigned to one of the four conditions using the *VESPR online studies* portal (Morys-Carter, 2022). As participants started flowing through the data collection procedure, this portal continuously self-balanced the assignment to conditions according to the proportions indicated in Section 4.2. After assigning participants to conditions, the program directed them to start the experiment on the *Pavlovia* online platform (https://pavlovia.org), which hosted the study. The experiment was programmed from the ground up using *PsychoPy* (Peirce et al., 2019), and the code was compiled in *PsychoJS*. To ensure consistency and control over the experimental conditions, participants were required to use a computer to launch the experiment. The screen background was set to white. Participants were asked three times to position themselves in a sufficiently illuminated room without direct light

[2] Protocol number 5010, November 3rd, 2022.

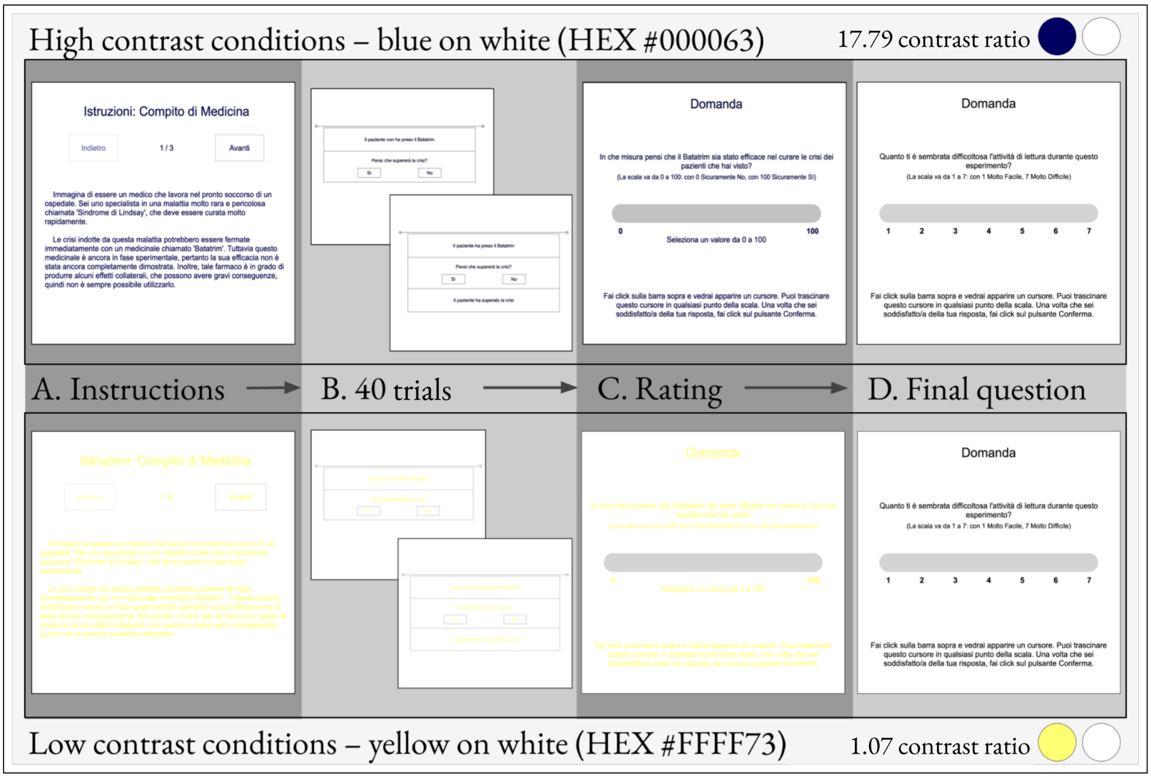

Figure 4.3: CLT structure and contrast differences.

on the screen. These reminders were provided when they received the recruitment link, in the informed consent form, and on the first instructions screen.

The experiment consisted of a standard CLT (see Figure 4.3), using an adaptation of the allergy task (Wasserman et al., 1990). The task was presented in Italian language. In the first phase of the experiment (see Figure 4.3 – Panel *A. Instructions*), a fictitious story was presented to the participants. Participants impersonated emergency room personnel and they were instructed to determine if there was a causal relationship between the presence of a medicine *Batatrim* (i.e., the potential cause or cue) and the healing of the disease *Lindsay Syndrome* (i.e., the outcome).

Then, in the second phase (see Figure 4.3 – Panel *B.* 40 *trials*), participants were exposed to a succession of 40 patient records (i.e., 40 trials, ITI= 1 sec), in a random order. Each record described one of four different scenarios (see Table 4.1), given by the possible presence or absence of the cue (i.e., *the patient had* or *had not taken the medicine to recover from the disease*) and the possible presence or absence of the outcome (i.e., *the patient had* or *had not recovered*). Through the manipulation of the frequency of the four scenarios, we created two different contingency conditions, namely a null contingency condition, in which $\Delta P = 0$, and a true contingency condition, in which $\Delta P = 0.60$. The exact frequencies of the four scenarios

	Null contingency				*True contingency*		
	B_1	B_2	$P(X\|Y)$		B_1	B_2	$P(X\|Y)$
A_1	15	5	$P(B_1\|A_1) = 0.75$	A_1	15	5	$P(B_1\|A_1) = 0.75$
A_2	15	5	$P(B_1\|A_2) = 0.75$	A_2	3	17	$P(B_1\|A_2) = 0.15$
	$\Delta P = 0$				$\Delta P = 0.60$		

Table 4.1: Frequency of each scenario and corresponding conditional probabilities.

in each condition are reported in Table 4.1. It should be noted that, in the null contingency condition, the probability of the presence of the outcome (i.e., $P = .75$) was much higher than the probability of the absence of the outcome (i.e., $P = .25$). According to the results from previous studies, this should lead to a outcome-density bias (Matute et al., 2015). Each patient record was composed of three horizontal panels. The upper panel remained visible for the whole duration of the trial, and informed the participant about the presence/absence of the cue (i.e., *The patient has taken the Batatrim*). The middle panel remained visible for the whole duration of the trial as well, and it presented a predictive question, to maintain the attention on the task. The participant was asked about whether the participant will heal after taking the medicine, by clicking with the mouse on one of the two buttons. No time limits were provided for the response. After the response was recorded, a third panel appeared below the middle one, which informed the participant about whether the patient had recovered or not. It is important to notice that the response provided by the participant through the mouse click had no influence on the information provided in the third panel. The three panels disappeared from the screen after 2 seconds, and then a new patient record was presented. As in Díaz-Lago and Matute (2019a)' study, we avoided including any pictures of the drug and the patient, to force the participants not to rely on shortcuts.

In the third phase of the procedure (see Figure 4.3 – Panel *C. Rating*), participants were asked to judge the strength of the causal relationship between the two events (i.e., *To what extent do you think that Batatrim was effective in healing the crises of the patients you have seen?*), using a visual analog scale from 0 (*Definitely Not*) to 100 (*Definitely Yes*). Once participants clicked on the scale, a cursor appeared, and participants could drag the cursor along the entire range between 0 and 100 to pick the exact judged discrete number. A numeric feedback was presented under the visual scale.

In the last phase (see Figure 4.3 – Panel *D. Final question*), participants were asked to judge the disfluency of the task through a single question (*How difficult have you found the reading activity during this experiment?*), using a 7-point Likert scale (1= *Very Easy*; 7= *Very Difficult*) that was similar to that used for the causality rating. We used a single item because, in this specific domain, the application of a single question has been shown to be robust from a psychometric standpoint and more understandable for participants with respect to multi-item scales (Graf et al., 2018).

As for the manipulation of perceptual (dis)fluency in the HC condition, the text for the instructions, patients records, and causality rating was presented in blue on a white background (HEX #000063; 17.79 contrast ratio; see the upper half of Figure 4.3), whereas in the LC condition the text was presented in yellow on a white background (HEX #FFFF73; 1.07 contrast ratio; see the lower half of Figure 4.3). In the reading difficulty rating phase, a black text on white background (HEX #000000; 21 contrast ratio; see the fourth panel of Figure 4.3) was used both in the HC and in the LC conditions. In each phase of the experiment an *arial* font was used, scaled to 0.03 height (i.e, the maximum height of any letter did not exceed 3 percent of the height of the screen).

5

First experiment: results

5.1 General data analysis procedure

After all 200 participants completed the experiment, we aggregated the individual data files generated by *PsychoPy* for each participant into a single dataframe. We then used both *R* (R Core Team, 2022) and *JASP* (JASP Team, 2023) to conduct descriptive and inferential data analyses, apply exclusion criteria (see Section 4.2), study the sample, evaluate the effectiveness of the experimental manipulation, and extract the results. Graphics were produced using the *ggplot2* package in *R* (Wickham, 2016). Data analyses were primarily conducted within the *null hypothesis significance testing* (NHST) framework, as already implied by the presence of the power analysis (see Section 4.2). For the critical tests in both experiment one and experiment two (see Sections 5.3 and 6.6), in addition to the NHST approach, we employed *Bayesian* analyses to provide more direct evidence for either H_1 or H_0 (see the hypotheses formulated in Section 3.4). Bayesian data analyses were also utilized in models comparison (see Section 7.2), as the Bayesian approach offers distinct advantages in models selection.

In this chapter, we will present the results of the first experiment, beginning with the analysis of the perceptual fluency data (see Section 5.2), followed by the results from the analysis of the causality rating task (see Section 5.3).

5.2 Manipulation check (perceptual fluency)

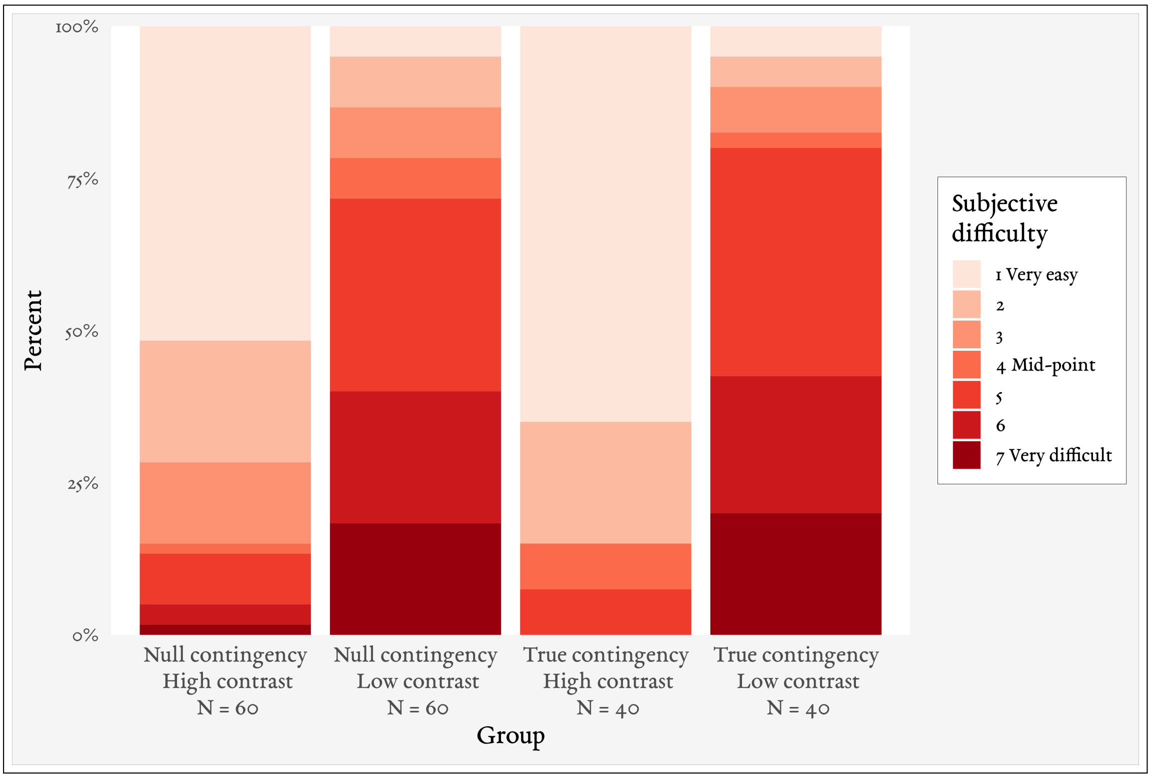

Figure 5.1: Subjective evaluation of reading difficulty expressed as percentages.

First, we expected the LC condition to induce a disfluency effect on participants. Perceptual fluency was measured using both subjective and objective indices.

For the subjective measure, we considered scores from the single-item question on subjective reading difficulty (see Section 4.3), which are represented in Figure 5.1 as percentages and Figure 5.2 as discrete distributions.

A two-way between-subjects *analysis of variance* (ANOVA) with factors contingency (null vs. true) and contrast (HC vs. LC) showed a significant main effect of contrast, $F(1, 196) = 184.99, p < .001, \eta_p^2 = .48$. Consistent with expectations, reading difficulty was rated higher in the LC condition ($M = 4.98, SD = 1.68$) than in the HC condition ($M = 1.96, SD = 1.46$). The main effect of contingency and the two-way interaction were not statistically significant [$F(1, 196) = 0.179, p = .67, \eta_p^2 = .0004$; $F(1, 196) = 1.704, p = .19, \eta_p^2 = .004$], suggesting that contingency had no effect on perceived reading difficulty ($M = 3.51, SD = 2.14$ in the null contingency condition and $M = 3.41, SD = 2.23$ in the true contingency condition).

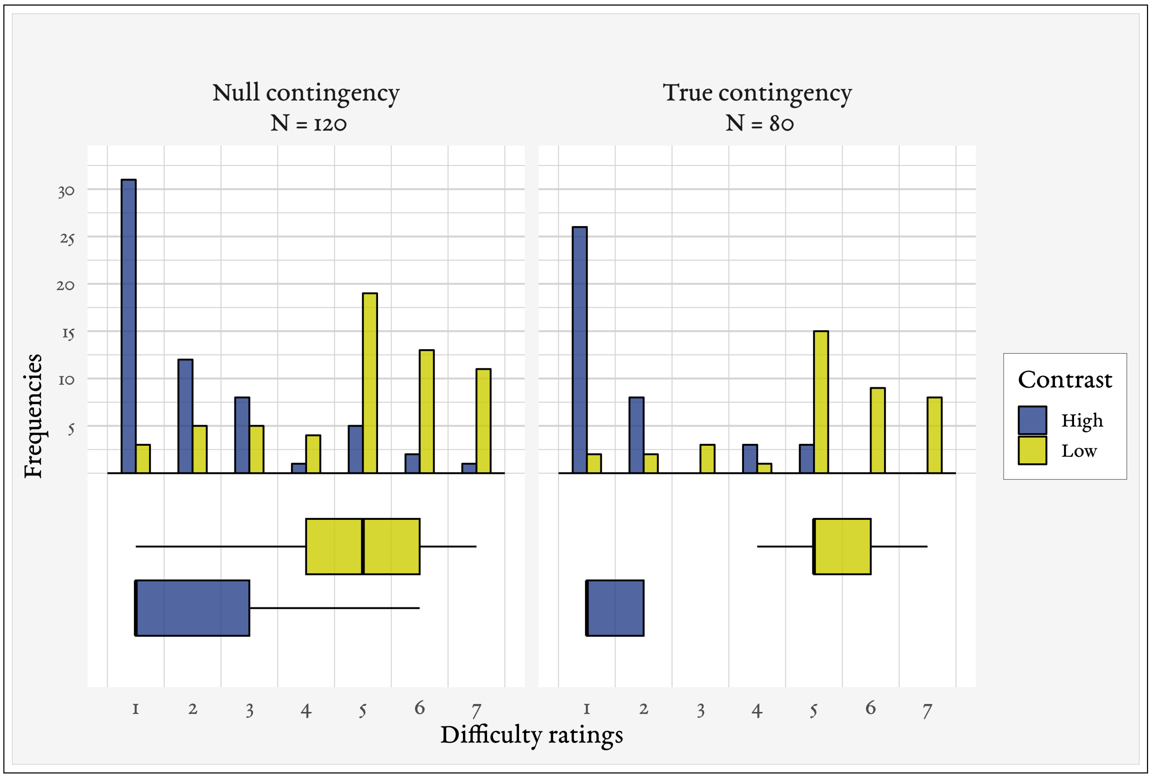

Figure 5.2: Subjective evaluation of reading difficulty expressed as discrete distributions.

For the objective measure, we considered the total time to complete the entire experimental procedure (see Figure 5.3). Experiment time was analyzed with the same independent variables as the rated reading difficulty. The main effect of contrast was statistically significant $F(1, 196) = 4.61, p = .033, \eta_p^2 = .02$, due to longer experiment time in the LC condition ($M = 373.25$ sec, $SD = 105.68$ sec) than in the HC condition ($M = 341.03$ sec, $SD = 106.26$ sec). The main effect of contingency and the two-way interaction were not statistically significant [$F(1, 196) = 1.27, p = .26, \eta_p^2 = .006$; $F(1, 196) = 0.34, p = .56, \eta_p^2 = .002$], suggesting that contingency had no effect on experiment time ($M = 350.25$ sec, $SD = 96.26$ sec in the null contingency condition and $M = 367.48$ sec, $SD = 121.09$ sec in the true contingency condition).

Furthermore, although the relationship between the two measures of (dis)fluency (i.e., total experiment time and subjective reading difficulty) is likely spurious, as they both depend on the experimental manipulation of contrast (i.e., LC vs. HC), it is noteworthy that the objective index and the subjective index were moderately correlated ($r_p = 0.20$), as shown in Figure 5.4, indicating the existence of an association between the two (dis)fluency measures.

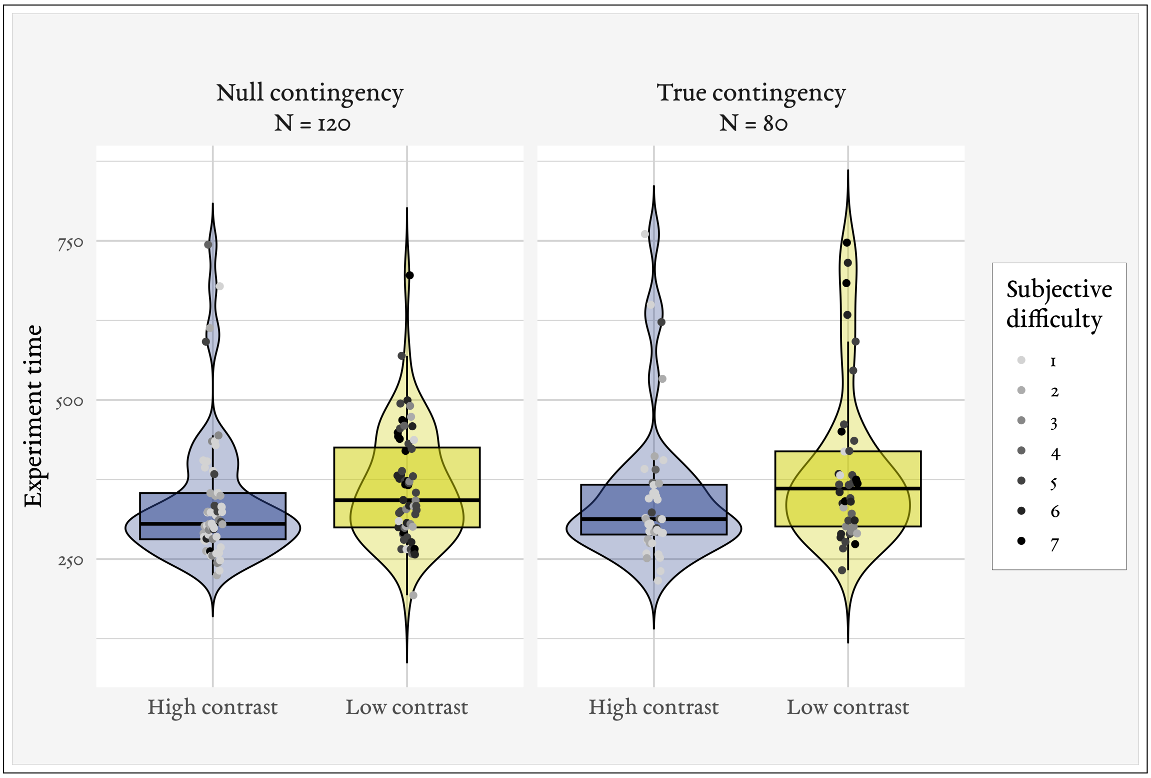

Figure 5.3: Experiment total time.

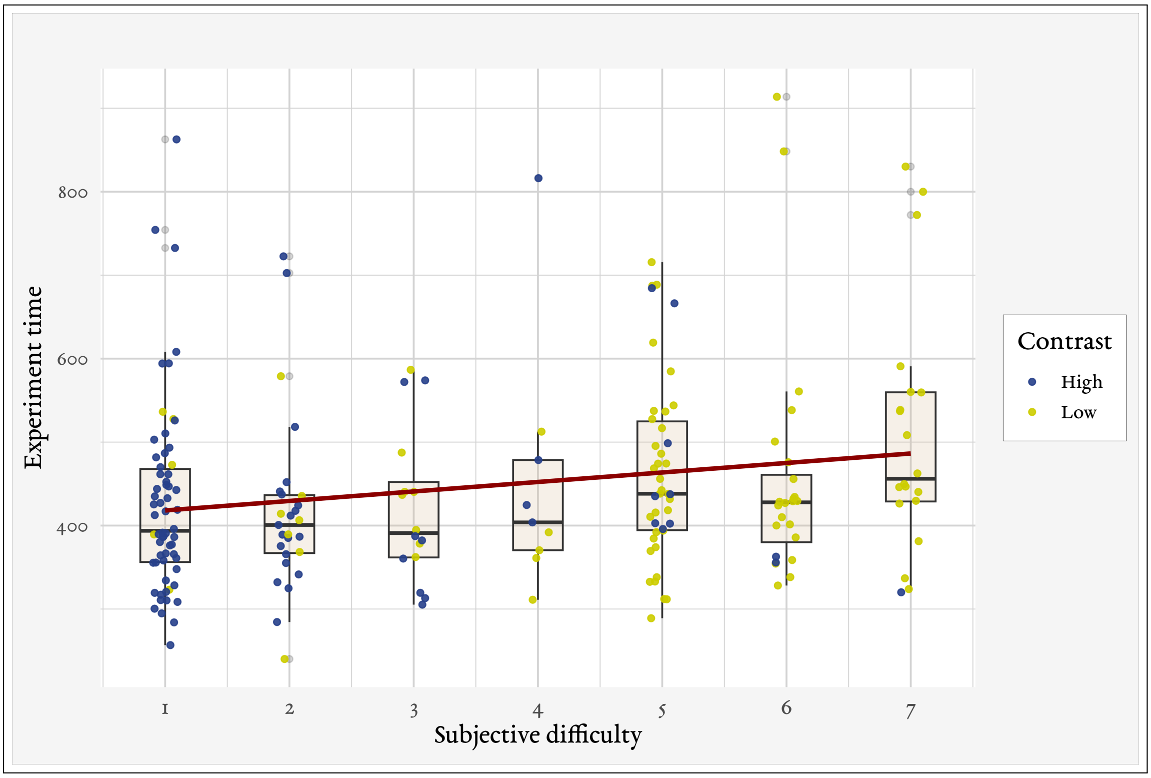

Figure 5.4: Correlation between time and subjective evaluation of reading difficulty.

5.3 Causality ratings

Our main goal was to test if perceptual disfluency, induced by a low contrast of the written text with the background, was able to produce the engagement of system two, leading to a reduction in the illusion of causality induced in a null contingency scenario. The causality ratings are represented in Figure 5.5 (the red dots on the charts indicate the means for each group).

A two-way between-subjects ANOVA with factors contingency (null vs. true) and contrast (HC vs. LC) showed a statistically significant main effect of contingency, $F(1, 196) = 12.98, p < .001, \eta_p^2 = .06$. As expected, the causality ratings were larger in the true contingency condition ($M = 69.95, SD = 15.82$) than in the null contingency condition ($M = 60.82, SD = 18.62$). However, it is worth highlighting the large mean value observed in the null contingency condition, which confirms the presence of a robust illusion of causality. The main effect of contrast was not statistically significant, $F(1, 196) = 0.86, p = .36, \eta_p^2 = .004$, as the means of the causality ratings in the HC and the LC condition were similar to each other (HC: $M = 63.32, SD = 18.25$; LC: $M = 65.62, SD = 17.93$). Crucially, the two-way interaction was not statistically significant, $F(1, 196) = 1.14, p = .29, \eta_p^2 = .005$, which is at odds with the hypothesis that perceptual disfluency induced by low contrast can lead to a decrease in the magnitude of the illusion of causality in the null contingency scenario.

In line with the pre-registered analysis plan (see Section 4.1), we also conducted a classic *one-tailed independent samples t-test* and a one-sided *Bayesian t-test* to test if, in the null contingency condition, the causality ratings in the HC condition were larger than the causality ratings in the LC condition. The results of the classic t-test were not statistically significant $t(118) = -1.32, p = .90, d = -0.24$. It is worth noting that the difference is in a direction opposite to that hypothesized, as the causality ratings in the HC condition ($M = 58.58, SD = 18.51$) were slightly smaller than those in the LC condition ($M = 63.05, SD = 18.64$).

The Bayesian t-test was performed with both R (R Core Team, 2022) and *JASP* (JASP Team, 2023). The results and graphics yielded from *JASP* software, as the main results are redundant with the ones yielded by R, will not be presented, but they can be found on OSF (see Section 4.1). In R we performed the Bayesian t-test using the *BayesFactor* package (Morey & Rouder, 2022).

We used a function to perform a so-called *JZS* t-test (Morey & Rouder, 2022), where the standardized effect size[1] under the alternative hypothesis (H_1) has been characterized by a truncated half-Cauchy prior distribution (see the dotted blue line in Figure 5.6) with a standard width parameter of $\sqrt{2}/2$ (i.e., we assumed a probability of $.50$ that the effect size lay between 0 and ~ -0.707). The standardized effect size under the null hypothesis (H_0) has been characterized by a point-null prior spike distribution.

The computation of *Bayes Factor* (BF) through a *Markov Chain Monte Carlo* ($MCMC$) procedure showed that the observed data were over 11 times more likely under the null hypothesis than under the alternative hypothesis ($BF_{01} = 11.07$), indicating strong evidence for the null hypothesis (see the moon chart in Figure 5.6, which was inspired by the type of graphics produced by *JASP*; JASP Team, 2023). The posterior distribution (see the red curve in Figure 5.6) was simulated through a *MCMC* method (10^5 iterations) using the alternative model as a prior. The posterior distribution for the alternative hypothesis was highly condensed near 0, in line with the conclusion of an absence of a significative disfluency effect ($Mdn = 0.06$, $CI = [-0.265; -0.002]$, see the red box in Figure 5.6).

Based on these results, we can conclude that increased perceptual disfluency, obtained through the presentation of the CLT experiment with LC written stimuli, did not elicit a reduction in the magnitude of the illusion of causality.

[1] *Standardized effect size*:

$$\delta = \frac{\mu_1 - \mu_2}{\sigma} \tag{5.1}$$

Where μ_1 is the mean of the first group, μ_2 is the mean of the second group, and σ is the standard deviation.

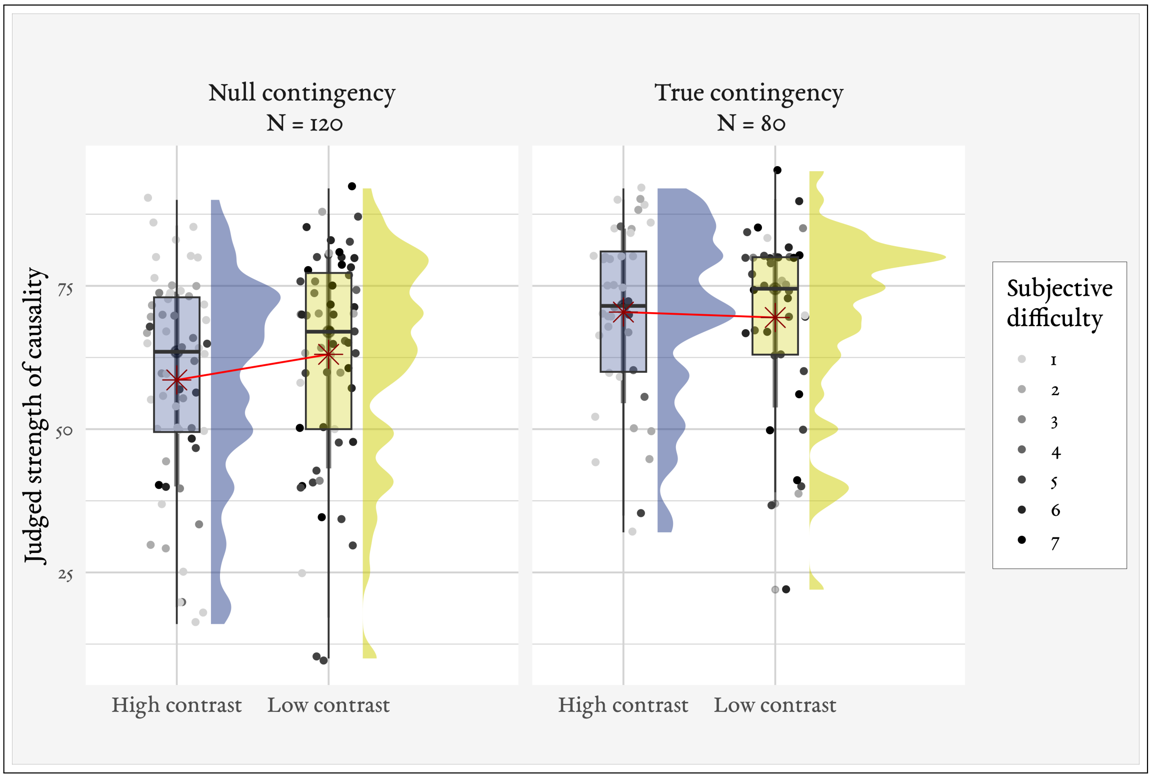

Figure 5.5: Causality ratings.

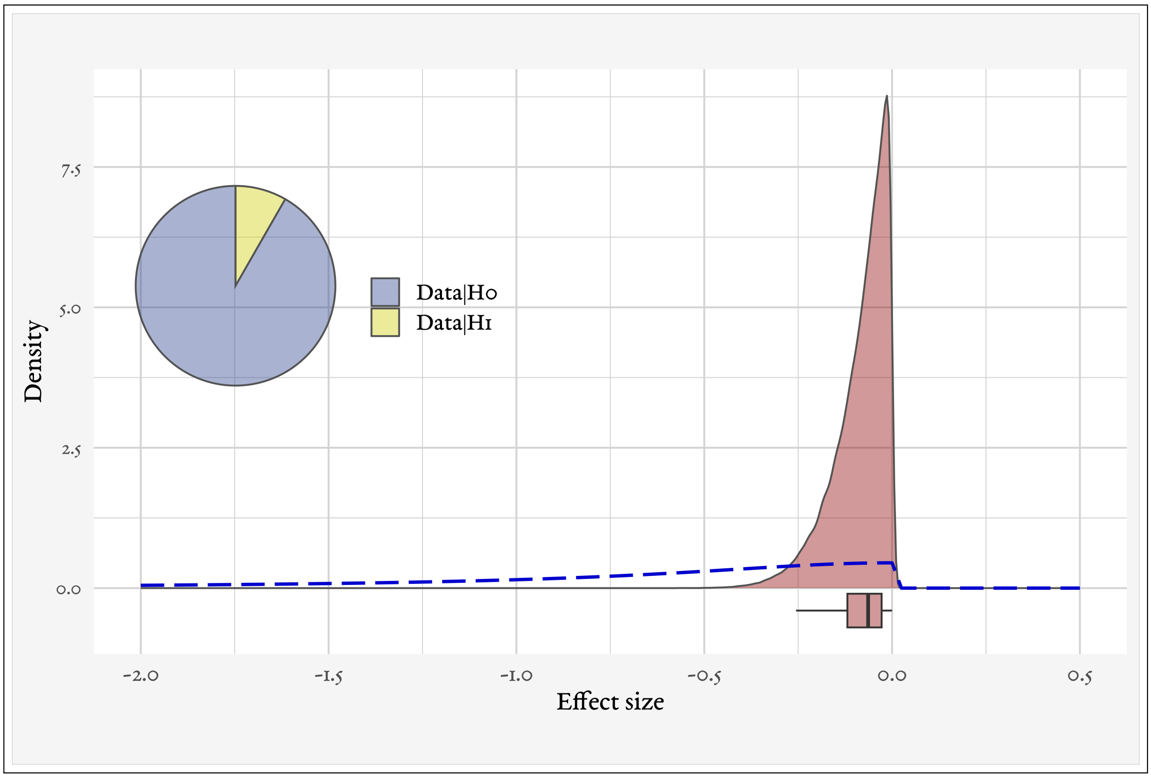

Figure 5.6: Prior and posterior for the alternative hypothesis.

6

Second experiment: an extension

6.1 Building upon the first experiment

We designed our first experiment as a generalization attempt of the findings reported by Díaz-Lago and Matute (2019a). The primary objective was to provide empirical support for the illusion-reduction hypothesis (see Section 3.4) through a targeted manipulation of the perceptual attributes of stimuli within a CLT paradigm. In particular, our first experiment sought to investigate whether altering perceptual fluency via contrast manipulation could influence the strength of the illusion. While the experiment successfully induced variations in task (dis)fluency through contrast manipulation (see Section 5.2), the anticipated effects on the magnitude of the illusion did not materialize (see Section 5.3).

The null results suggest a nuanced relationship between fluency manipulations and the cognitive processes underlying the illusion. Not all fluency manipulations, it appears, are capable of modulating the illusion's strength. This outcome underscores the necessity of a critical reassessment of the broader hypothesis that cognitive disfluency invariably triggers a shift towards a more deliberative and effortful cognitive processing mode (i.e., system two).

6.2 Conditions

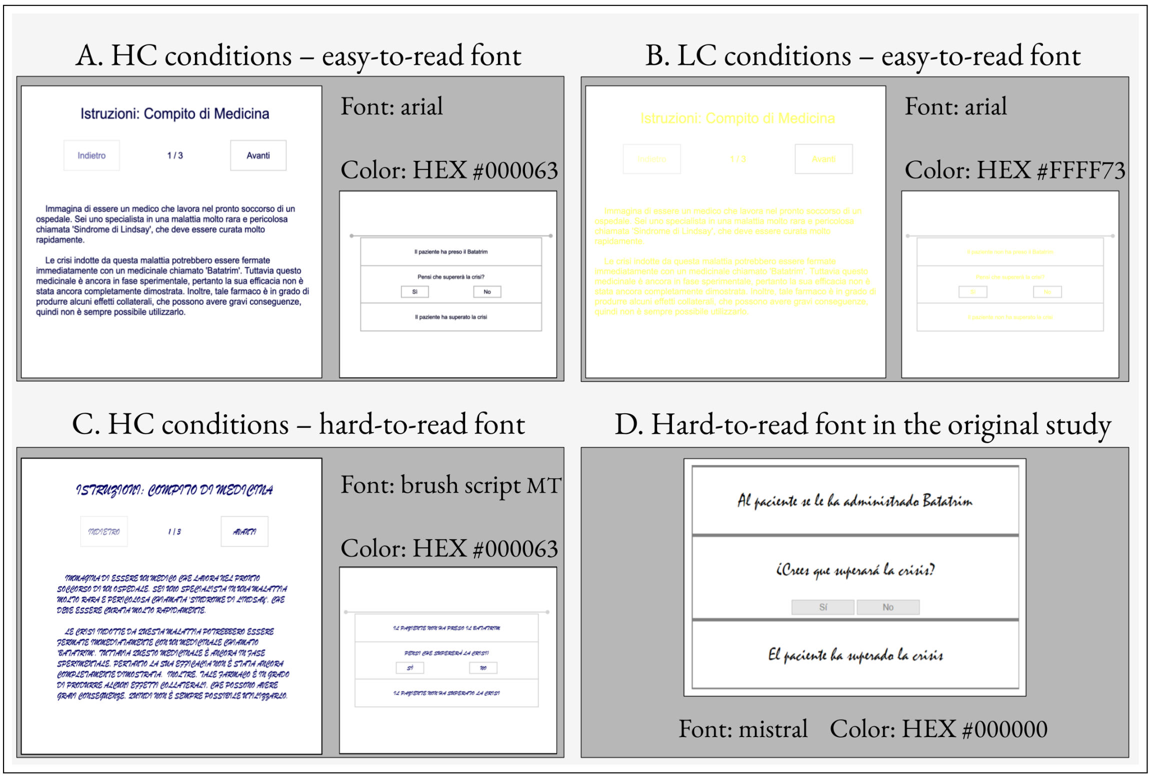

Figure 6.1: Comparisons between previous and latter conditions.

To achieve a more nuanced understanding of how perceptual disfluency influences causality bias, we designed experiment two to test whether the original effect observed by Díaz-Lago and Matute (2019a) could be replicated in a similar experimental setting. Our aim was to further explore the impact of font type on the magnitude of causality bias by conceptually replicating their previous font-manipulation study.

Based on the results from the reading difficulty question discussed in Section 5.2, it is important to note that both the HC true contingency condition and the HC null contingency condition from experiment one were characterized by the use of a font that was easy-to-read (see Figure 6.1 – Panel *A. HC conditions – easy-to-read font*). Specifically, in these conditions that employed a fluent font, participants generally rated the task on a 7-point Likert scale as reasonably easy, with a mean difficulty rating of 1.96 ($SD = 1.46$; see Figures 5.1 and 5.2).

For experiment two, we expanded upon this setup (i.e., the two HC conditions) by introducing two new participant groups who were presented with the experimental materials in a hard-to-read font (i.e., an *uppercase brush script MT*, hereafter referred to as *brush*; see Figure 6.1 – Panel *C. HC conditions – hard-to-read font*). Thus, we conducted a comparison between the already collected data from the HC conditions and the fresh data from two new groups of participants. These new groups were exposed to either a true contingency condition ($\Delta P = 0.60$) or a null contingency condition ($\Delta P = 0$), as in the previous experiment.

The experimental design for experiment two was a 2 (*font*: *arial*, easy-to-read vs. *brush*, hard-to-read) × 2 (*contingency*: *true contingency* vs. *null contingency*) between-subjects factorial design. The task presented in the brush font can be seen in Figure 6.1 (Panel *C. HC conditions – hard-to-read font*). Notice the difference with the LC conditions that we created in our first experiment, which are shown in Figure 6.1 (Panel *B. LC conditions – easy-to-read font*).

If perceptual disfluency induced by the use of a hard-to-read font reduces the strength of the causality illusion, we hypothesize, in the same fashion as our first hypothesis (see Section 3.4), that causality ratings will be lower in the disfluent condition compared to the fluent condition, specifically within the null contingency condition. Conversely, no significant effect of font type is expected in the true contingency condition, which, again, constitutes our control group.

It is important to clarify a technical detail: the brush font used in experiment two differs from the *mistral* font employed in Díaz-Lago and Matute (2019a)' study (see Figure 6.1 – Panel *D. hard-to-read font in the original study*). While visually similar, brush was chosen for its practical suitability in online studies, as it does not require participants to download and install the font on their devices. More information about this font can be found on the *W3schools* website (https://www.w3schools.com/css/cssfontwebsafe.asp). The brush font is designed to mimic handwriting, making it visually comparable to mistral. Moreover, it has been previously utilized as a hard-to-read font to induce cognitive bias reduction (e.g., Song & Schwarz, 2008). Based on empirical observations, we opted for an uppercase variant of the brush font to ensure increased reading difficulty in the online task. To maintain experimental rigor, we strictly adhered to the procedures established in experiment one, ensuring consistency across both experiments to facilitate direct comparisons between their results.

We pre-registered our new hypotheses and the methodological details for experiment two on OSF at the following link: https://osf.io/4tdcy7. Additionally, the experimental code, raw data, and scripts used for the primary analyses are available on OSF at the same link used for the previous experiment: https://osf.io/c26qa.

6.3 Participants

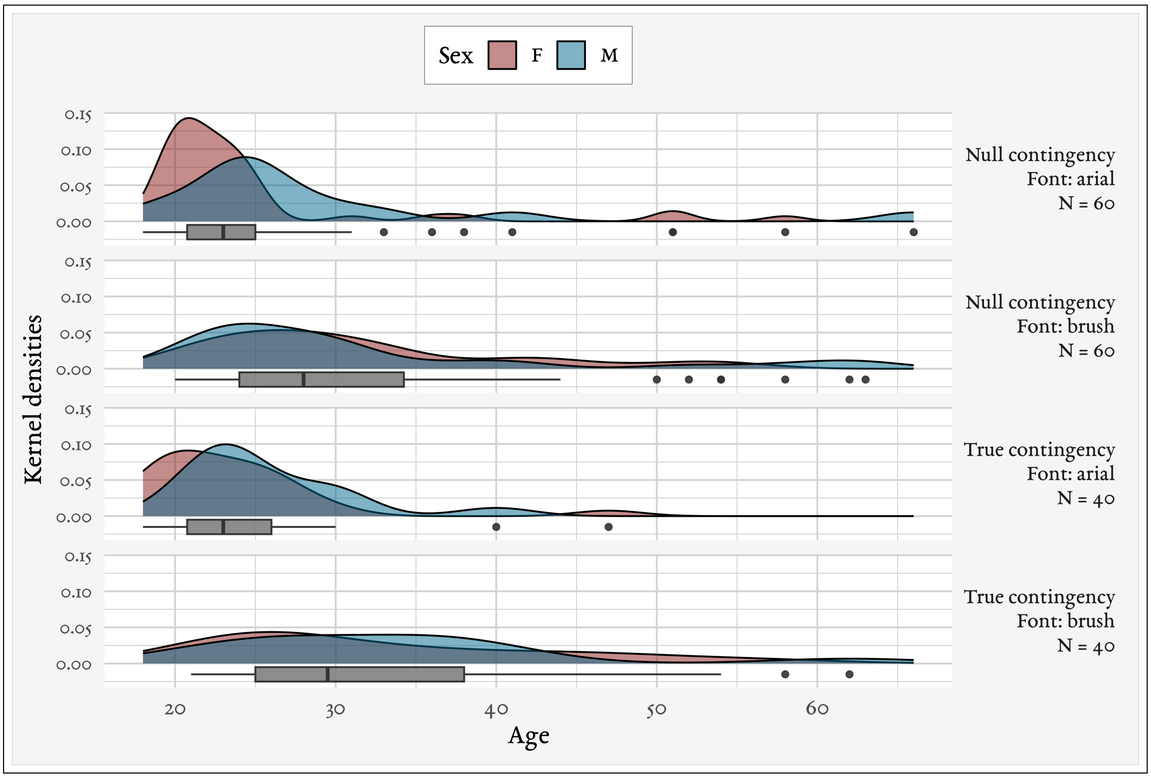

Figure 6.2: Participants' age and biological sex distribution.

The sample size for experiment two was determined using the same criteria as those applied in experiment one (see Section 4.2). In the previous experiment, 100 participants were tested under the easy-to-read arial font condition (i.e., HC condition) – with 60 participants in the null contingency condition and 40 participants in the true contingency condition. To maintain consistency and comparability, we recruited for the second experiment a new sample of 100 participants, who were randomly assigned to either the null contingency condition (60 participants) or the true contingency condition (40 participants). Indeed, the critical comparison maintained the same power achieved for experiment one (see Section 4.2).

Participants, all of whom were native Italian speakers, were recruited via the online platform *Prolific* (https://www.prolific.co/). Our sample comprised 60 females and 40 males, with an average age of 32.19 years ($SD = 10.67$). No participants were excluded from the analysis, adhering to the same exclusion criteria applied in experiment one (see Section 4.2).

To assess demographic balance across the experimental groups (see Figure 6.2), a Pearson's chi-square test of independence was conducted. The results indicated no significant differences in the distribution of biological sex across the four experimental groups, $\chi^2(3) = 6.78, p = 0.08$. However, a Kruskal-Wallis test revealed significant age differences between the groups, $\chi^2(3) = 42.27, p < 0.001$, with the new sample's average age being notably higher than that of participants in the HC conditions of experiment one ($M = 25.21$ years, $SD = 8.21$). Despite these demographic differences, previous research on the illusion of causality consistently shows that this cognitive bias manifests independently of confounding factors (see Section 3.1). Moreover, no substantial evidence in the literature suggests a demographic distinction in the occurrence of the causality illusion. To confirm this, we combined data from all 300 participants across both experiments and conducted a *Bayes factor general linear model analysis* using the *BayesFactor R* package (Morey & Rouder, 2022). This analysis aimed to evaluate the potential influence of age and sex on causality judgments. The results supported the null model, which had at least 7.37 times greater support compared to alternative models that included age, sex, or both as predictors. We included sex as a predictor because the new sample exhibited a more balanced distribution of sexes (60 females and 40 males) compared to the HC conditions of experiment one (70 females and 30 males), although the *p-value* for the chi-square test was slightly above the threshold of statistical significance. These findings suggest that despite the demographic variations between the samples, the comparison between the hard-to-read and easy-to-read font conditions remains methodologically robust. The detailed results can be accessed on OSF at the following link: https://osf.io/c26qa.

6.4 Procedure

The procedural framework of experiment two remained largely consistent with that of experiment one, with the exception of few specific modifications. In this experiment, all instructions, patient records, and causality rating tasks were presented in a hard-to-read font (i.e., brush) with a blue color (HEX: #000063; 0.03 height; 17.79 contrast ratio). Participants recruited on *Prolific* platform were subsequently redirected to the *Pavlovia* website to initiate the experiment. Before beginning the experiment, participants were required to review and accept the same online consent form as presented in Section 4.3.

In the following sections, we will present the results of the second experiment, beginning with the analysis of the perceptual fluency data (see Section 6.5), followed by the results from the analysis of the causality rating task (see Section 6.6).

6.5 Manipulation check (perceptual fluency)

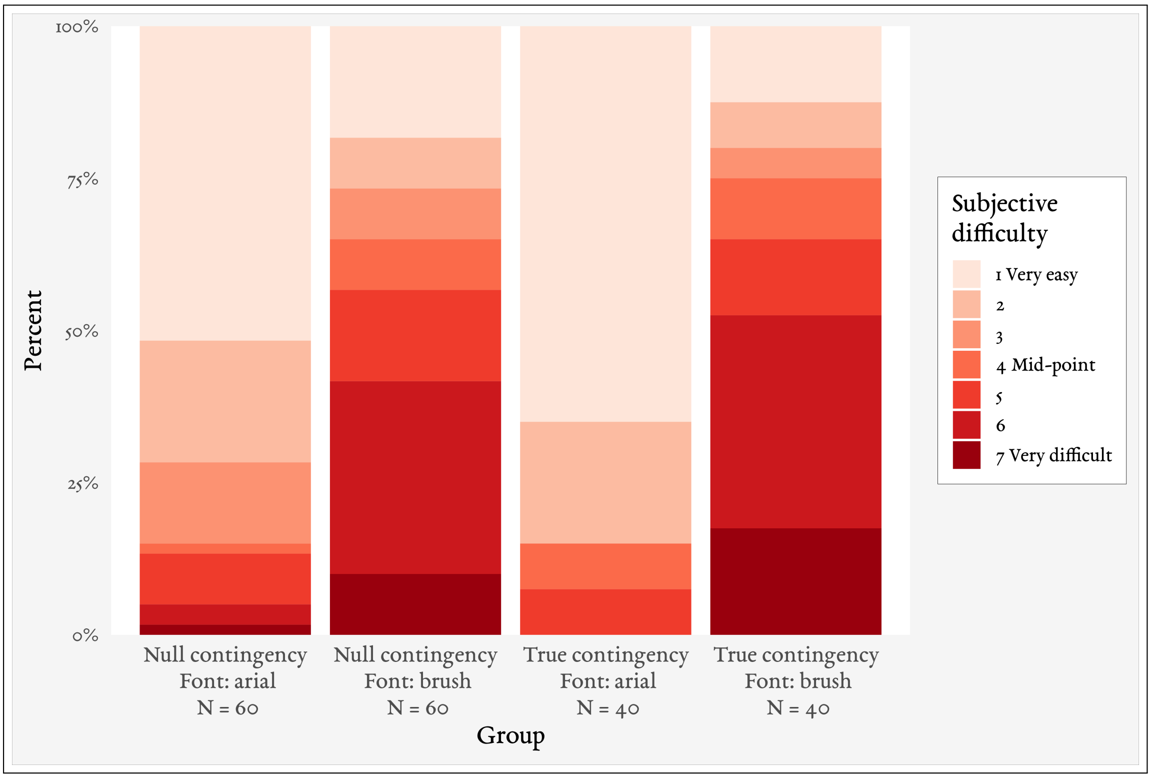

Figure 6.3: Subjective evaluation of reading difficulty expressed as percentages.

As for the subjective measure of disfluency, we considered scores from the single-item question on subjective reading difficulty (see Section 4.3), which are represented in Figure 6.3 as percentages and in Figure 6.4 as discrete distributions.

A two-way between-subjects ANOVA with factors contingency (null vs. true) and font (arial vs. brush) showed a significant main effect of font, $F(1, 196) = 99.84, p < .001, \eta_p^2 = .34$. As expected, reading difficulty was rated higher with the brush font ($M = 4.48, SD = 2.07$) than with the arial font ($M = 1.96, SD = 1.46$). The main effect of contingency and the two-way interaction were not statistically significant [$F(1, 196) = 0.03, p = .85, \eta_p^2 = .0002$; $F(1, 196) = 2.95, p = .08, \eta_p^2 = .014$], suggesting that contingency had no effect on perceived reading difficulty ($M = 3.20, SD = 2.14$ in the null contingency condition and $M = 3.25, SD = 2.27$ in the true contingency condition).

For the objective measure, we considered the total time to complete the entire experimental procedure (see Figure 6.5). Experiment time was analyzed with the same independent variables as the rated reading difficulty. The main effect of font was statistically significant, $F(1, 196) =$

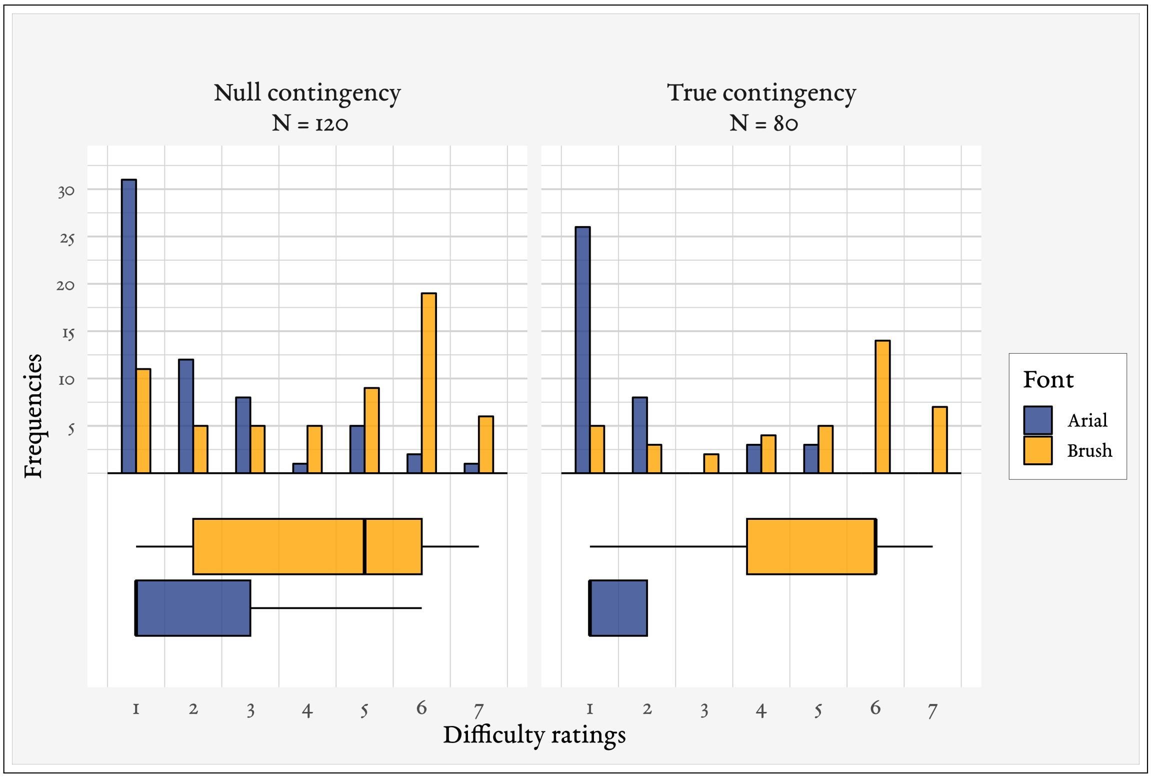

Figure 6.4: Subjective evaluation of reading difficulty expressed as discrete distributions.

$4.97, p = .03, \eta_p^2 = .02$, due to longer experiment time in the brush condition ($M = 376.81$ sec, $SD = 122.37$ sec) than in the arial condition ($M = 341.03$ sec, $SD = 106.26$ sec). The main effect of contingency and the two-way interaction were not statistically significant [$F(1, 196) = 2.13, p = .15, \eta_p^2 = .01$; $F(1, 196) = 3.86, p = .051, \eta_p^2 = .02$], suggesting that contingency had no effect on experiment time ($M = 344.58$ sec, $SD = 97.23$ sec in the true contingency condition and $M = 368.47$ sec, $SD = 126.03$ sec in the null contingency condition). It is worth noting that the outlier shown in the left panel of Figure 6.5 (experiment time longer than 900 sec) was associated with an acceptable Cook's distance (0.11), and the main results did not change even after removing that outlier from the dataset.

Furthermore, the two measures of (dis)fluency (i.e., total experiment time and subjective reading difficulty) were moderately correlated ($r_p = 0.16$), as shown in Figure 6.6, indicating, again, the existence of an association between the two (dis)fluency measures.

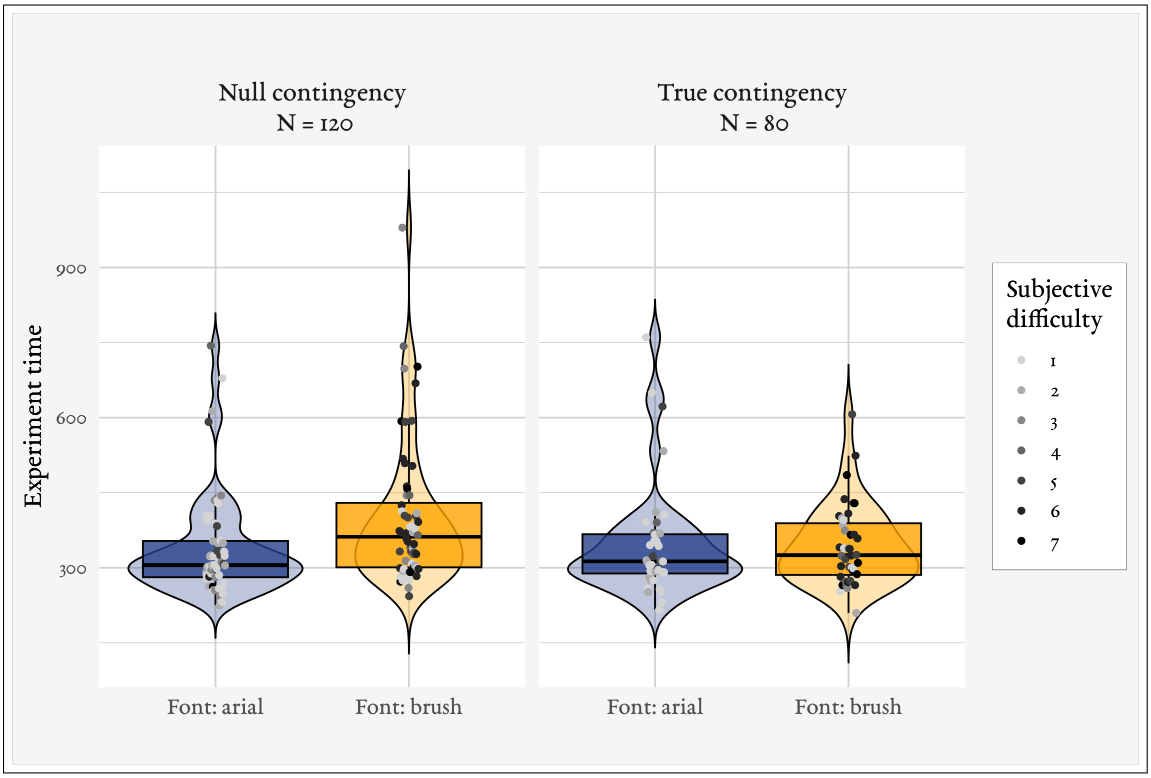

Figure 6.5: Experiment total time.

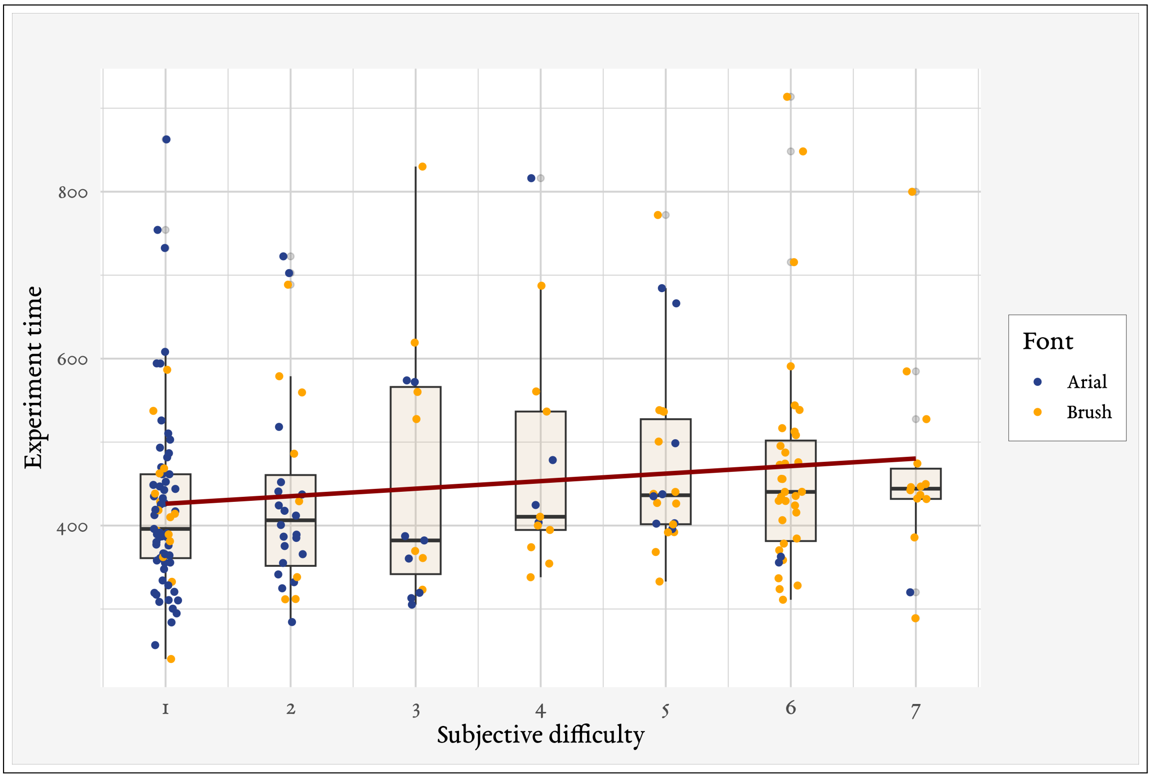

Figure 6.6: Correlation between time and subjective evaluation of reading difficulty.

6.6 Causality ratings

The causality ratings are represented in Figure 6.7 (the red dots on the charts indicate the means for each group). A two-way between-subjects ANOVA with factors contingency (null vs. true) and font (arial vs. brush) showed a statistically significant main effect of contingency, $F(1, 196) = 30.11, p < .001, \eta_p^2 = .13$. As expected, the causality ratings were larger in the true contingency condition ($M = 72.74, SD = 16.72$) than in the null contingency condition ($M = 58.13, SD = 19.44$). The main effect of font was not statistically significant, $F(1, 196) = 0.25, p = .61, \eta_p^2 = .001$, as the means of the causality ratings in the brush and arial font conditions were similar to each other (brush: $M = 64.63, SD = 21.14$; arial: $M = 63.32, SD = 18.25$). The two-way interaction was not statistically significant, $F(1, 196) = 1.08, p = .30, \eta_p^2 = .005$, which is at odds with the hypothesis that perceptual disfluency induced by a hard-to-read font can lead to a decrease in the illusion of causality.

In line with the pre-registered analysis plan, we also conducted a classic one-tailed independent samples t-test and a one-sided Bayesian t-test to test if, in the null contingency condition, the causality ratings in the brush condition were smaller than the causality ratings in the arial condition. The results of the t-test showed that the causality ratings associated with the disfluent brush font ($M = 57.68, SD = 20.48$) were not significantly smaller than those associated with the fluent arial font ($M = 58.58, SD = 18.51$), $t(118) = 0.25, p = .40, d = .05$.

We performed the Bayesian t-test using the *BayesFactor* package (Morey & Rouder, 2022) in R, in the same manner as shown in Section 5.3, using a function to perform the *JZS* t-test using the same alternative and null priors as descibed in Section 5.3. The computation of BF through a *MCMC* procedure showed that the observed data were over 4 times more likely under the null hypothesis than under the alternative hypothesis ($BF_{01} = 4.20$), indicating moderate evidence for the null hypothesis (see the moon chart in Figure 6.8). The posterior distribution for the alternative hypothesis (see the red curve in Figure 6.8) was simulated through a *MCMC* method (10^5 iterations). The posterior distribution was moderately condensed near 0, in line with the conclusion of an absence of a significative effect ($Mdn = -0.13$, $CI = [-0.418; -0.006]$, see the red box in Figure 5.6). Based on these results, we can conclude that increased perceptual disfluency, obtained through the presentation of the CLT experiment with a hard-to-read font, did not elicit a reduction in the the illusion of causality.

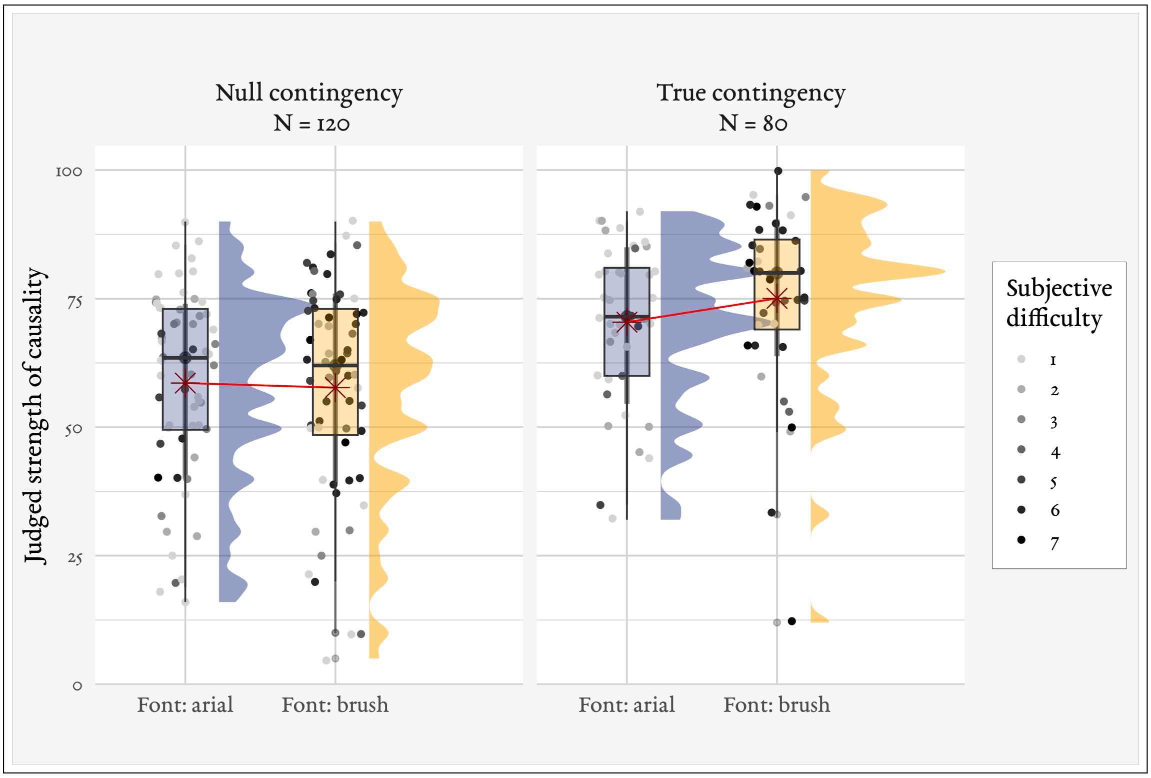

Figure 6.7: Causality ratings.

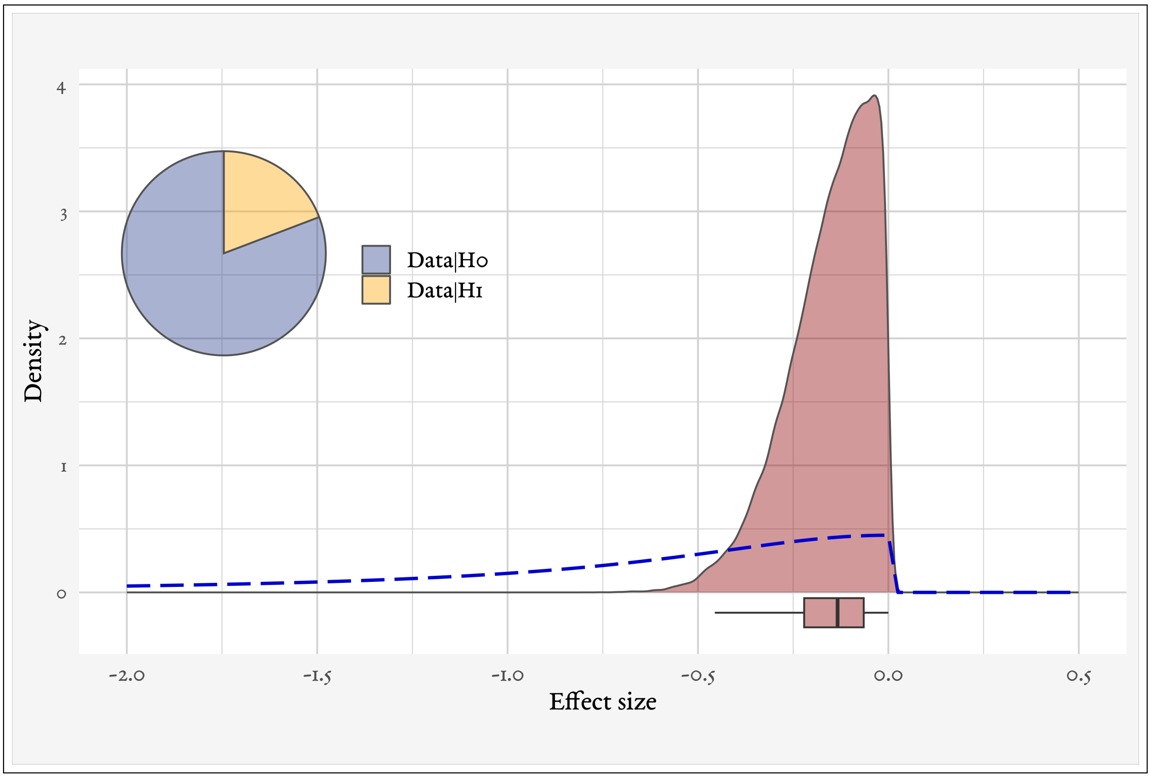

Figure 6.8: Prior and posterior for the alternative hypothesis.

7

Discussion of results

7.1 Evidence of absence

We designed our second experiment as a replication attempt of the findings reported by Díaz-Lago and Matute (2019a). Given that our first experiment did not provide evidence supporting the hypothesis that perceptual disfluency could diminish the illusion of causality (see Section 5.3), we sought to investigate whether altering perceptual fluency through the manipulation of font difficulty – using a hard-to-read font – could exert an influence on the magnitude of the illusion of causality. The results from our second experiment indicate that while the manipulation of font successfully induced variations in task (dis)fluency (see Section 6.5), this alteration did not translate into measurable effects on the strength of the illusion of causality (see Section 6.6). This finding directly contrasts with the expectations set by the original study of Díaz-Lago and Matute (2019a).

In the following sections, we will present a comprehensive analysis of the combined data from both of our experiments (see Section 7.2). Subsequently, we will engage in a thorough discussion of our hypotheses in light of these results (see Section 7.3). This discussion will explore possible explanations for the lack of observed effects, proposing directions for future studies and research.

7.2 Models comparison

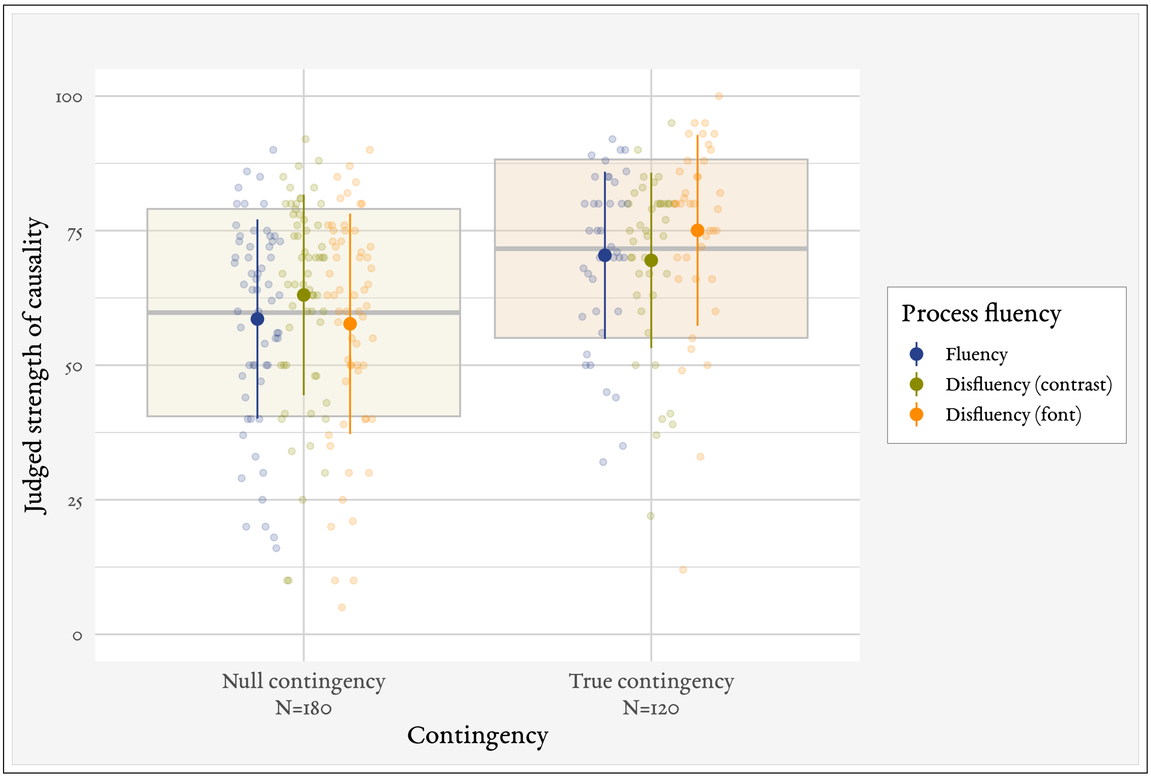

Figure 7.1: Relationship between fluency, contingency, and rated causality.

To gain a deeper understanding of the potential influence of perceptual (dis)fluency on causality judgments, we conducted a comprehensive analysis by merging the datasets from both experiments, resulting in a combined sample size of $N = 300$. This larger dataset allowed us to rigorously assess the relationship between perceptual fluency (i.e., fluency vs. font disfluency vs. contrast disfluency), contingency (i.e., true vs. null), and the perceived strength of causality.

Figure 7.1 offers a visual summary of the interaction between fluency and contingency in shaping participants' causal judgments. In particular, the dot plot shows the judged strength of causality for the six experimental groups. Each dot represents the mean of the corresponding group, and each line represents the standard deviation of the corresponding group. The low-contrast boxes show the overall means (central line) and standard deviations (ends of the boxes) for the null contingency and the true contingency conditions.

We employed four distinct linear models to systematically evaluate the relationship between the dependent variable – judged strength of causality – and the key predictors: contingency (i.e., model 1; $M1$), fluency (i.e., model 2; $M2$), contingency and fluency without interaction (i.e., model 3; $M3$), and contingency and fluency with the interaction effect (i.e., model 4; $M4$). In these models, fluency was treated as a categorical variable with three distinct levels, as depicted in the legend of Figure 7.1.

To quantify the relative evidence for each model, we utilized the BF, calculated via a *MCMC* algorithmic approach, implemented in the $BayesFactor$ package (Morey & Rouder, 2022). Remarkably, the analysis revealed that $M1$, which excluded fluency as a predictor, demonstrated the strongest explanatory power. Specifically, this model was found to be 19 to 3×10^5 times more likely than the models that incorporated fluency as a predictor. This suggests that the inclusion of fluency alongside contingency did not enhance the model's ability to explain the data, thereby weakening the case for an interaction effect between fluency and contingency. To corroborate these findings, we also calculated *Akaike weights* (using the *Akaike information criterion*, *AIC*) and *Bayesian weights* (using the *Bayesian information criterion*, *BIC*) with the *MuMIn* package (Bartoń, 2023). These additional analyses consistently supported the dominance of $M1$ within the evidence, reinforcing the conclusion that this simple model, which considers only contingency without fluency, provides the most robust explanation for the observed data.

Model	*Predictor(s)*	BF_{10}	*AIC weight*	*BIC weight*
M1	Contingency	156163.60	0.69	0.99
M2	Fluency	0.05	< 0.01	< 0.01
M3	Contingency + fluency	7815.36	0.14	< 0.01
M4	Contingency × fluency (interaction)	3690.19	0.17	< 0.01

Table 7.1: Models' descriptions and associated indices.

Table 7.1 presents a summary of the models comparison, highlighting the superiority of $M1$. Figure 7.2 offers two heat maps (one for the AIC weights and one for the BIC weights) with a direct one-on-one comparison between every combination of two models, including $M0$, which constitutes the null model without predictor(s). The weight of a certain model has been divided by the weight of another model, and the result has been log-transformed. On the X-axis are represented the models at the numerator, whereas on the Y-axis are represented the models at the denominator. Each cell value can be interpreted as the relative evidence in favor (highlighted in red) or against (highlighted in blue) the model placed on the X-axis with respect to the model on the Y-axis (this representation was inspired by the manual by Pastore, 2015).

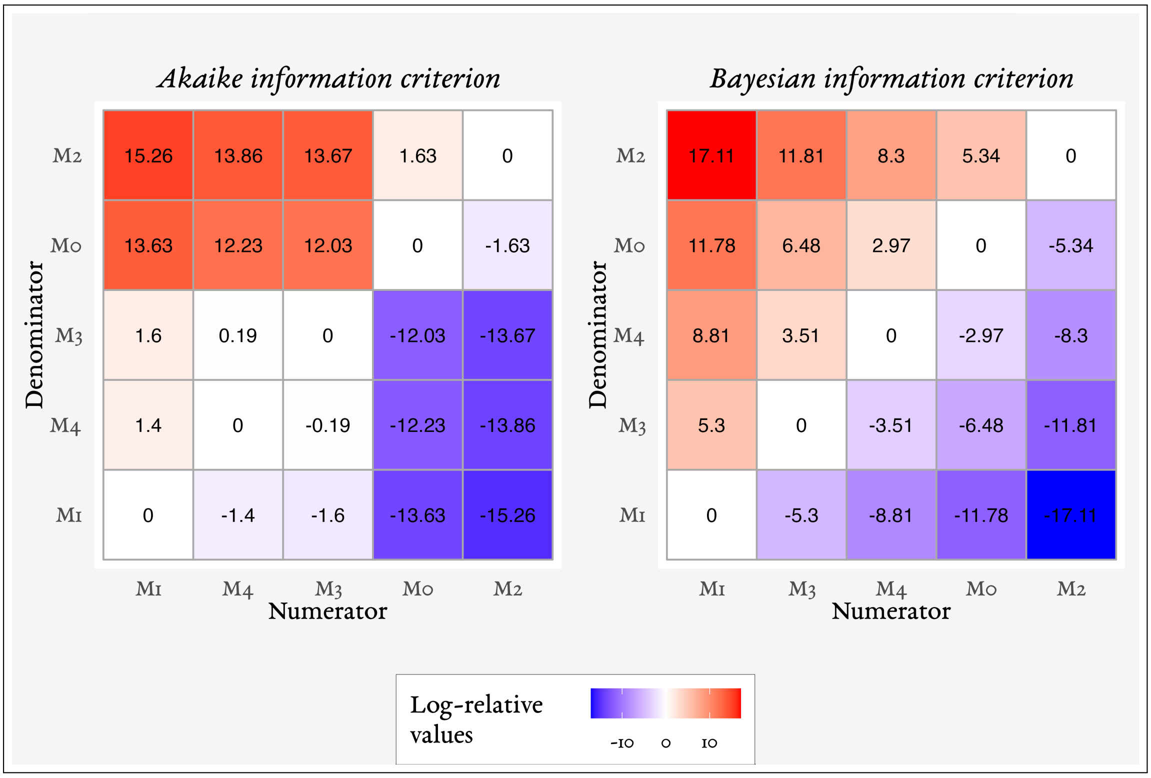

Figure 7.2: Models' (log-)relative evidences.

7.3 Interpretation of results

Our study found null results, supporting the notion that perceptual disfluency does not modulate the illusion of causality, at least in cases where the disfluency manipulation reaches a significant magnitude (see Subsection 7.3.1). These findings also lend support to the broader conclusion that the effects of processing fluency on cognition remain ambiguous, as highlighted by Meyer et al. (2015). Importantly, the sample sizes were determined via an *a priori* power analysis (see Section 4.2), ensuring that the null results cannot be attributed to a lack of statistical power. Furthermore, the absence of effects cannot be ascribed to ineffective experimental manipulations, as both subjective and objective measures confirmed that LC and hard-to-read font conditions were associated with reduced processing fluency compared to HC and easy-to-read font conditions (see Sections 5.2 and 6.5).

In the following Subsections (7.3.1 and 7.3.2), we propose two potential explanations for the null results observed in our study. These hypotheses are tentative and necessitate further empirical investigation to establish their validity.

7.3.1 U-shaped relationship hypothesis

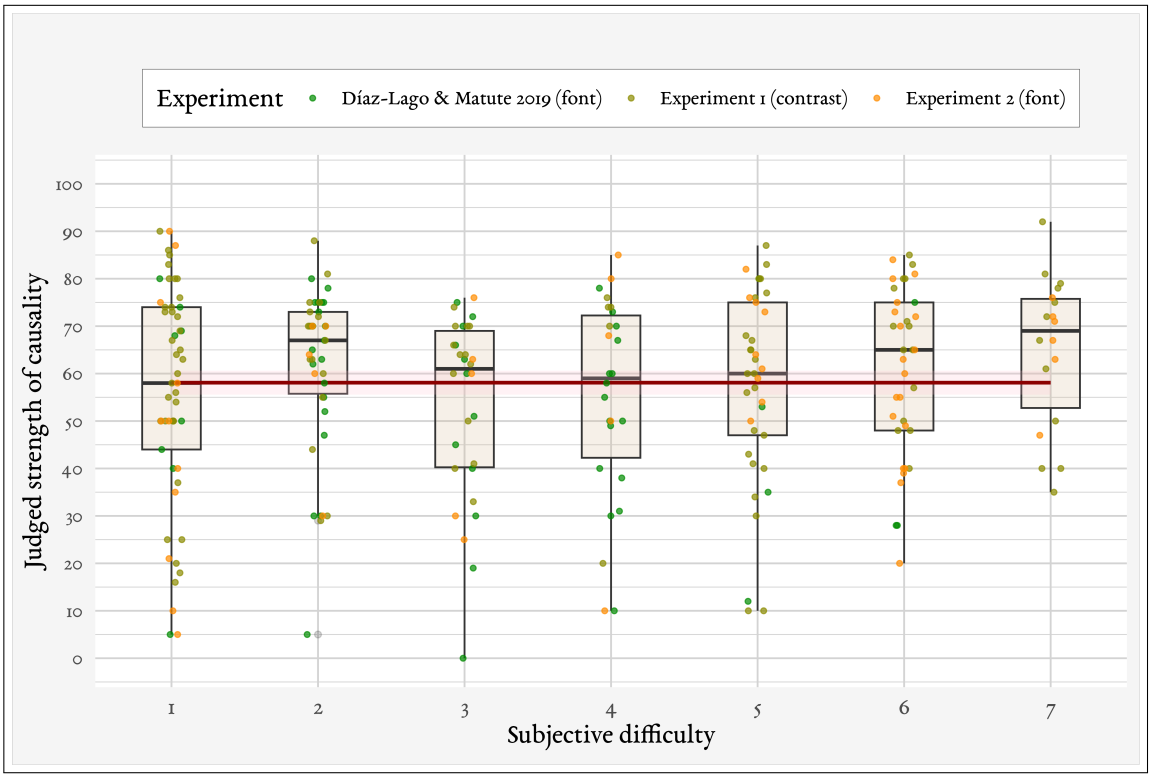

Figure 7.3: Subjective difficulty and causality ratings in null contingency conditions.

Our first hypothesis suggests that the disruption of perceptual fluency may provoke the engagement of a more deliberative and effortful mode of thinking (i.e., system two processes). However, we posit that the relationship between processing fluency and the illusion of causality might follow a non-linear, U-shaped function. In this model, moderate disfluency could enhance performance by engaging system two, but excessive disfluency might overload and saturate the system's capacities, thereby reducing or masking the engagement of system two.

If our hypothesis is accurate, it implies that the level of disfluency induced by LC and hard-to-read font conditions in our study may have exceeded the optimal threshold for system two engagement. Preliminary support for this hypothesis comes from comparing the disfluency measures in our experiments with those in the study by Díaz-Lago and Matute (2019a). In their study, fluency was assessed using three 7-point Likert scales that evaluated the ease of reading, task fluency, and perceived task duration. While no significant effects were found for perceived task duration, font type significantly influenced the ease of reading and task fluency, with a mean difference of approximately 1.4 points between the two font types on the ease of reading scale[1]. In contrast, our experiments observed mean differences of approximately 3 points (experiment one) and 2.5 points (experiment two) in reading difficulty between the disfluent and fluent conditions. These observations align with our proposed hypothesis, suggesting that Díaz-Lago and Matute (2019a) may have observed a mitigated causality bias due to the introduction of low to moderate levels of disfluency. Conversely, our experiments failed to replicate this bias attenuation, possibly due to the imposition of higher levels of disfluency. While this comparison should be approached with caution, the use of a 7-point Likert scale in both studies facilitates such an analysis.

Beyond these qualitative observations, a more rigorous test of the hypothesized U-shaped relationship between processing fluency and the illusion of causality is enabled by the availability of materials and raw data from the study by Díaz-Lago and Matute (2019a) on OSF (https://osf.io/vrukz/). We merged data from our experiments with data from their original study (see Figure 7.3), focusing on the relationship between causality ratings and perceived task difficulty in null contingency conditions[2]. We constructed two statistical models: one with perceived task difficulty as a linear predictor, and the other with perceived task difficulty as a quadratic predictor. To evaluate the strength of evidence supporting each model, we computed the *BF* using a *MCMC* procedure via the *BayesFactor* package (Morey & Rouder, 2022). The results indicated that the null model (see the red line on Figure 7.3) was 9.66 times more likely than the linear model and 4.33 times more likely than the quadratic model, countering the hypothesis of a linear or non-linear relationship between perceived difficulty and causality judgments. It is crucial to note that the experiments considered here were not specifically designed to examine a U-shaped relationship between task difficulty perception and causality ratings. Furthermore, the original study by Díaz-Lago and Matute (2019a) had a smaller sample size (63 participants) compared to the larger sample sizes in our two experiments (180 participants). As a result, a small portion of our merged data was exposed to a less disfluent condition relative to the larger and complementary subset. Therefore, these results should be interpreted with caution. Future studies should address this hypothesis precisely.

[1] In Díaz-Lago and Matute (2019a)' study, the easiness of reading was found to be higher for the easy-to-read font ($M = 5.72$, $SD = 1.11$ in the null contingency condition; $M = 5.62$, $SD = 1.12$ in the true contingency condition) with respect to the hard-to-read font ($M = 4.35$, $SD = 1.20$ in the null contingency condition; $M = 4.28$, $SD = 1.08$ in the true contingency condition).

[2] The *easiness of reading* item (7-point Likert scale) from the study by Díaz-Lago and Matute (2019a) has been inverted in order to express the perceived difficulty of the task instead of the perceived easiness. This makes it directly comparable with the ratings from our two experiments.

7.3.2 Differential activation hypothesis

Our second hypothesis builds upon the absence of any discernible and systematic connection between perceptual fluency and causality bias, as evidenced by the outcomes of our two experiments. The results from our study and Díaz-Lago and Matute (2019a)' study can be interpreted as conflicting evidence, wherein a standard NHST framework could be applied: in our study, the probability of at least one of two critical statistical comparisons correctly rejecting the null hypothesis – thereby detecting a significative effect, if present, given a hypothesized effect size of 0.58 Cohen's d – was greater than 99%. Conversely, the results obtained by Díaz-Lago and Matute (2019a) could represent a false positive with a standard α probability of 5%, making our findings more likely from a statistical standpoint.

This hypothesis also considers the potential influence of the FLE (see Section 3.2) on the illusion of causality (Díaz-Lago & Matute, 2019b). Although the literature on FLE remains inconclusive in explaining this phenomenon (Circi et al., 2021), many bilingual cognitive models emphasize the role of cognitive control mechanisms (Schwieter & Ferreira, 2016). We hypothesize that not all types of processing disfluency are equally effective in activating system two processing. It is possible that system two activation is more closely linked to high-order lexical and semantic processes involved in processing a disfluent FL, while remaining unaffected by manipulations of superficial perceptual features such as contrast or font type. This may explain why presenting the CLT in a disfluent FL reduces causality bias, while presenting the CLT in a hard-to-read format does not. This hypothesis stands in contrast to the first hypothesis (see Section 3.4) and the findings of Díaz-Lago and Matute (2019a), as it suggests that manipulating the perceptual features of information may not reduce the causality bias.

7.3.3 Conclusion

In conclusion, fluency serves as a versatile construct, playing a key role in gauging the perceived difficulty of a task, which is closely tied to the cognitive load across various processes. However, the results from our two experiments raise doubts about the all-encompassing explanatory power of the fluency construct. These findings highlight the need for a more in-depth and nuanced examination of the cognitive mechanisms that influence cognitive biases. Future research should strive to pinpoint, with greater accuracy, the specific manipulations of PF that can effectively reduce the illusion of causality.

Appendix

Chapman and Robbins (1990) demonstrated that when a single event A and a single event B are involved, the RWM (see Subsection 2.6.3) can be reduced to the ΔP rule (see Section 2.4). This demonstration relies on two key assumptions:

- The context X is present during every trial.
- Learning continues until there is no more discrepancy between the actual and expected outcomes (i.e., until $\Delta V_{A(k)}$ converges to 0).

The RWM updates the associative strength $V_{A(k)}$ for each event A (and context X) based on the RWM error-correction rule (see Equation 2.17) expressed in Subsection 2.6.3. In a typical contingency judgment experiment involving one event A and a context X, there are four types of trials, corresponding to the four cells a, b, c, and d, of the 2×2 Table 2.1. The RWM equations for each trial type are as follows:

For a trials

$$\begin{aligned} \Delta V_X &= \alpha_X \beta (1 - (V_X + V_A)) \\ \Delta V_A &= \alpha_A \beta (1 - (V_X + V_A)) \end{aligned} \tag{1}$$

For b trials

$$\begin{aligned} \Delta V_X &= \alpha_X \beta (0 - (V_X + V_A)) \\ \Delta V_A &= \alpha_A \beta (0 - (V_X + V_A)) \end{aligned} \tag{2}$$

For c trials

$$\Delta V_X = \alpha_X \beta (1 - V_X) \tag{3}$$

For d trials

$$\Delta V_X = \alpha_X \beta (0 - V_X) \tag{4}$$

Thus, for each trial type one equation is constructed for both event A and context X present on that trial. The value of λ is set to 1 on trials with event B present and is set to 0 on trials with event B absent. The associative strength V_A is updated during a and b trials, while V_X is updated during all four types of trials. Thus, for any particular block of trials, the average change in V_A is weighted by the relative frequencies of a and b trial types, and the average change in V_X is weighted by the relative frequencies of each trial type:

$$\text{Mean } \Delta V_A = \alpha_A \beta \left[a(1 - (V_X + V_A)) + b(0 - (V_X + V_A))\right] \tag{5}$$

$$\begin{aligned} \text{Mean } \Delta V_X = \alpha_X \beta \, [a\,(1 - (V_X + V_A)) + b\,(0 - (V_X + V_A)) + \\ + c\,(1 - V_X) + d\,(0 - V_X)] \end{aligned} \tag{6}$$

Simplifying:

$$\frac{\text{Mean } \Delta V_A}{\alpha_A \beta} = a - V_A(a+b) - V_X(a+b) \tag{7}$$

$$\frac{\text{Mean } \Delta V_X}{\alpha_X \beta} = a + c - V_A(a+b) - V_X(a+b+c+d) \tag{8}$$

Learning continues until Equations 7 and 8 stabilize at zero. V_A and V_X become constants: although they may fluctuate from trial to trial, their means will maintain a constant value over many blocks of trials. We can set Equation 8 equal to zero and solve for V_X:

$$V_X = \frac{a + c - V_A(a+b)}{a+b+c+d} \tag{9}$$

Setting Equation 7 equal to zero and substituting V_X from Equation 9, we get:

$$0 = a - V_A(a+b) - \frac{(a+b)\,[a + c - V_A(a+b)]}{a+b+c+d} \tag{10}$$

Solving Equation 10 for V_A, Chapman and Robbins (1990) found that V_A yields the ΔP index:

$$V_A = \frac{a}{(a+b)} - \frac{c}{(c+d)} = \Delta P \tag{11}$$

References

Allan, L. G. (1980). A note on measurement of contingency between two binary variables in judgment tasks. *Bulletin of the Psychonomic Society*, *15*(3), 147–149. DOI: 10.3758/bf03334492

Allan, L. G. (1993). Human contingency judgments: Rule based or associative? *Psychological Bulletin*, *114*(3), 435–448. DOI: 10.1037//0033-2909.114.3.435

Allan, L. G., & Jenkins, H. M. (1980). The judgment of contingency and the nature of the response alternatives. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *34*(1), 1–11. DOI: 10.1037/h0081013

Allan, L. G., & Jenkins, H. M. (1983). The effect of representations of binary variables on judgment of influence. *Learning and Motivation*, *14*(4), 381–405. DOI: 10.1016/0023-9690(83)90024-3

Alloy, L. B., & Abramson, L. Y. (1979). Judgment of contingency in depressed and nondepressed students: Sadder but wiser? *Journal of experimental psychology: General*, *108*(4), 441–485. DOI: 10.1037//0096-3445.108.4.441

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and social psychology review*, *13*(3), 219–235. DOI: 10.1177/1088868309341564

Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: metacognitive difficulty activates analytic reasoning. *Journal of experimental psychology: General*, *136*(4), 569–576. DOI: 10.1037/0096-3445.136.4.569

Arkes, H. R. (1981). Impediments to accurate clinical judgment and possible ways to minimize their impact. *Journal of consulting and clinical psychology*, *49*(3), 323–330. DOI: 10.1037//0022-006x.49.3.323

Armstrong, D. M. (2016). *What is a Law of Nature?* Cambridge University Press. DOI: 10.1017/cbo9781316499030

Atkinson, R., & Hilgard, E. (2017). Introduzione alla psicologia (16th ed.). Piccin.

Baker, A., Murphy, R. A., & Vallee-Tourangeau, F. (1996). Associative and normative models of causal induction: Reacting to versus understanding cause. *Psychology of Learning and Motivation*, *34*, 1–45. DOI: 10.1016/s0079-7421(08)60557-5

Barberia, I., Blanco, F., Cubillas, C. P., & Matute, H. (2013). Implementation and assessment of an intervention to debias adolescents against causal illusions. *PLoS One*, *8*(8), e71303. DOI: 10.1371/journal.pone.0071303

Barberia, I., Tubau, E., Matute, H., & Rodríguez-Ferreiro, J. (2018). A short educational intervention diminishes causal illusions and specific paranormal beliefs in undergraduates. *PLoS One*, *13*(1), e0191907. DOI: 10.1371/journal.pone.0191907

Barberia, I., Vadillo, M. A., & Rodríguez-Ferreiro, J. (2019). Persistence of causal illusions after extensive training. *Frontiers in Psychology*, *10*, 24. DOI: 10.3389/fpsyg.2019.00024

Bartoń, K. (2023). *MuMIn: Multi-Model Inference* (R Package; Version 1.47.5). Url: https://CRAN.R-project.org/package=MuMIn

Bender, A. (2020). What is causal cognition? *Frontiers in psychology*, *11*, 3. DOI: 10.3389/fpsyg.2020.00003

Benson, B. (2016, September 1). Cognitive bias cheat sheet. Better Humans. Url: https://betterhumans.pub/cognitive-bias-cheat-sheet-55a472476b18

Blaisdell, A. P., Sawa, K., Leising, K. J., & Waldmann, M. R. (2006). Causal reasoning in rats. *Science*, *311*(5763), 1020–1022. DOI: 10.1126/science.1121872

Blanco, F. (2017). Positive and negative implications of the causal illusion. *Consciousness and Cognition*, *50*, 56–68. DOI: 10.1016/j.concog.2016.08.012

Blanco, F., Barberia, I., & Matute, H. (2014). The lack of side effects of an ineffective treatment facilitates the development of a belief in its effectiveness. *PloS one*, *9*(1), e84084. DOI: 10.1371/journal.pone.0084084

Blanco, F., Matute, H., & A. Vadillo, M. (2012). Mediating role of activity level in the depressive realism effect. *Plos One*, *7*(9), e46203. DOI: 10.1371/journal.pone.0046203

Blanco, F., Matute, H., & Vadillo, M. A. (2013). Interactive effects of the probability of the cue and the probability of the outcome on the overestimation of null contingency. *Learning & Behavior*, *41*, 333–340. DOI: 10.3758/s13420-013-0108-8

Broadbent, A. (2024, May 7). Causality. The Internet Encyclopedia of Philosophy. Url: https://iep.utm.edu/causation

Callegher, C., Bertoldo, G., Toffalini, E., Vesely, A., Andreella, A., Pastore, M., & Altoè, G. (2021). PRDA: An R package for prospective and retrospective design analysis. *Journal of Open Source Software*, *6*(58), 2810. DOI: 10.21105/joss.02810

Caverni, J. P., Fabre, J. M., & Gonzalez, M. (1990). *Cognitive Biases: Their Contribution for Understanding Human Cognitive Processes* (1st ed., Vol. 68 – Advances in Psychology book series). North-Holland. DOI: 10.1016/S0166-4115(08)61311-4

Champely, S. (2020). *pwr: Basic Functions for Power Analysis* (R package; Version 1.3-0). Url: https://CRAN.R-project.org/package=pwr

Chapman, G. B., & Robbins, S. J. (1990). Cue interaction in human contingency judgment. *Memory & Cognition*, *18*(5), 537–545. DOI: 10.3758/bf03198486

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological review*, *104*(2), 367–405. DOI: 10.1037//0033-295x.104.2.367

Cipolletti, H., McFarlane, S., & Weissglass, C. (2016). The moral foreign-language effect. *Philosophical Psychology*, *29*(1), 23–40. DOI: 10.1080/09515089.2014.993063

Circi, R., Gatti, D., Russo, V., & Vecchi, T. (2021). The foreign language effect on decision-making: A meta-analysis. *Psychonomic Bulletin & Review*, *28*, 1131–1141. DOI: 10.3758/s13423-020-01871-z

Cornoldi, C., Meneghetti, C., Moè, A., & Zamperlin, C. (2018). *Processi cognitivi, motivazione e apprendimento*. Il Mulino.

Costa, A., Foucart, A., Arnon, I., Aparici, M., & Apesteguia, J. (2014). Piensa twice: On the foreign language effect in decision making. *Cognition*, *130*(2), 236–254. DOI: 10.1016/j.cognition.2013.11.010

Costa, A., Foucart, A., Hayakawa, S., Aparici, M., Apesteguia, J., Heafner, J., & Keysar, B. (2014). Your morals depend on language. *PloS one*, *9*(4), e94842. DOI: 10.1371/journal.pone.0094842

Costa, A., & Sebastián-Gallés, N. (2014). How does the bilingual experience sculpt the brain? *Nature reviews neuroscience*, *15*(5), 336–345. DOI: 10.1038/nrn3709

Craigie, J. (2011). Thinking and feeling: Moral deliberation in a dual-process framework. *Philosophical Psychology*, *24*(1), 53–71. DOI: 10.1080/09515089.2010.533262

Crocker, J. (1981). Judgment of covariation by social perceivers. *Psychological Bulletin*, *90*(2), 272–292. DOI: 10.1037//0033-2909.90.2.272

Dalla Bona, S., & Vicovaro, M. (2024). Does perceptual disfluency affect the illusion of causality? *Quarterly Journal of Experimental Psychology*, *77*(8), 1727–1744. DOI: 10.1177/17470218231220928

De Beni, R., & Borella, E. (2015). *Psicologia dell'invecchiamento e della longevità* (2nd ed.). Il Mulino.

Del Maschio, N., Crespi, F., Peressotti, F., Abutalebi, J., & Sulpizio, S. (2022). Decision-making depends on language: A meta-analysis of the Foreign Language Effect. *Bilingualism: Language and Cognition*, *25*(4), 617–630. DOI: 10.1017/s1366728921001012

Díaz-Lago, M., & Matute, H. (2019a). A hard to read font reduces the causality bias. *Judgment and Decision Making*, *14*(5), 547–554. DOI: 10.1017/s1930297500004848

Díaz-Lago, M., & Matute, H. (2019b). Thinking in a foreign language reduces the causality bias. *Quarterly Journal of Experimental Psychology*, *72*(1), 41–51. DOI: 10.1177/1747021818755326

Dickinson, A., Shanks, D., & Evenden, J. (1984). Judgement of act-outcome contingency: The role of selective attribution. *The Quarterly Journal of Experimental Psychology*, *36*(1), 29–50. DOI: 10.1080/14640748408401502

Evans, J. S. (2011). Dual-process theories of reasoning: Contemporary issues and developmental applications. *Developmental Review*, *31*(2), 86–102. DOI: 10.1016/j.dr.2011.07.007

Evans, J. S. (2012). Dual process theories of deductive reasoning: Facts and fallacies. In K. J. Holyoak & R. G. Morrison (Eds.), *The oxford handbook of thinking and reasoning* (pp. 115–133). Oxford University Press New York. DOI: 10.1093/oxfordhb/9780199734689.013.0008

Fisk, J. E. (2016). Conjunction fallacy. In R. F. Pohl (Ed.), *Cognitive illusions* (2nd ed., pp. 25–43). Routledge. DOI: 10.4324/9781003154730-4

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic perspectives*, *19*(4), 25–42. DOI: 10.1257/089533005775196732

Garfield, J. (2002). The challenge of developing statistical reasoning. *Journal of statistics education*, *10*(3). DOI: 10.1080/10691898.2002.11910676

Geipel, J., Hadjichristidis, C., & Surian, L. (2015). How foreign language shapes moral judgment. *Journal of experimental social psychology*, *59*, 8–17. DOI: 10.1016/j.jesp.2015.02.001

Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond "heuristics and biases". *European review of social psychology*, *2*(1), 83–115.
DOI: 10.1080/14792779143000033

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky. *Psychological Review*, *103*(3), 592–596. DOI: 10.1037//0033-295x.103.3.592

Gigerenzer, G. (2008). Why heuristics work. *Perspectives on psychological science*, *3*(1), 20–29. DOI: 10.1111/j.1745-6916.2008.00058.x

Gigerenzer, G., Hertwig, R., Hoffrage, U., & Sedlmeier, P. (2008). Cognitive illusions reconsidered. In H. R. Plott & V. L. Smith (Eds.), *Handbook of experimental economics results* (pp. 1018–1034). Elsevier. DOI: 10.1016/S1574-0722(07)00109-6

Graf, L. K., Mayer, S., & Landwehr, J. R. (2018). Measuring processing fluency: One versus five items. *Journal of Consumer Psychology*, *28*(3), 393–411. DOI: 10.1002/jcpy.1021

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, *44*(2), 389–400. DOI: 10.1016/j.neuron.2004.09.027

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, *293*(5537), 2105–2108. DOI: 10.1126/science.1062872

Griffiths, O., Shehabi, N., Murphy, R. A., & Le Pelley, M. E. (2019). Superstition predicts perception of illusory control. *British Journal of Psychology*, *110*(3), 499–518. DOI: 10.1111/bjop.12344

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological review*, *108*(4), 814–834.
DOI: 10.1017/cbo9780511814273.055

Hall, R. H., & Hanna, P. (2004). The impact of web page text-background colour combinations on readability, retention, aesthetics and behavioural intention. *Behaviour & information technology*, *23*(3), 183–195. DOI: 10.1080/01449290410001669932

Haselton, M. G., Nettle, D., & Andrews, P. W. (2015). The evolution of cognitive bias. In D. M. Buss (Ed.), *The handbook of evolutionary psychology* (pp. 724–746). Wiley Online Library. DOI: 10.1002/9780470939376.ch25

Hell, W., Fiedler, K., & Gigerenzer, G. (1993). *Kognitive täuschungen: Fehl-leistungen und mechanismen des urteilens, denkens und erinnerns*. Spektrum, Akademischer Verlag.

Holyoak, K. J., & Cheng, P. W. (2011). Causal learning and inference as a rational process: The new synthesis. *Annual review of psychology*, *62*(1), 135–163. DOI: 10.1146/annurev.psych.121208.131634

Huff, D. (1954). *How to lie with statistics*. Norton.

Hume, D. (2000). *A Treatise of Human Nature*. Oxford University Press. (Original work published 1740)

Inhelder, B., & Piaget, J. (2013). *The growth of logical thinking from childhood to adolescence: An essay on the construction of formal operational structures* (Vol. 84). Routledge. DOI: 10.4324/9781315009674 (Original work published 1958)

Irvine, A. D., & Deutsch, H. (2021, October 12). Russell's Paradox. The Stanford encyclopedia of philosophy. Url: https://plato.stanford.edu/archives/spr2021/entries/russell-paradox/

JASP Team. (2023). *JASP* (comp. software; Version 0.17). Url: https://jasp-stats.org/

Jenkins, H. M., & Ward, W. C. (1965). Judgment of contingency between responses and outcomes. *Psychological monographs: General and applied*, *79*(1), 1–17. DOI: 10.1037/h0093874

Kahneman, D. (2012). *Thinking, fast and slow*. Penguin Books.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49–81). DOI: 10.1017/cbo9780511808098.004

Kant, I. (2007). *Critique of Pure Reason*. Palgrave Macmillan London. DOI: 10.1007/978-1-137-10016-0 (Original work published 1855)

Kelley, C. M., & Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of memory and language*, *32*(1), 1–24. DOI: 10.1006/jmla.1993.1001

Keysar, B., Hayakawa, S. L., & An, S. G. (2012). The foreign-language effect: Thinking in a foreign tongue reduces decision biases. *Psychological science*, *23*(6), 661–668. DOI: 10.1177/0956797611432178

Korn, C. W., Ries, J., Schalk, L., Oganian, Y., & Saalbach, H. (2018). A hard-to-read font reduces the framing effect in a large sample. *Psychonomic bulletin & review*, *25*, 696–703. DOI: 10.3758/s13423-017-1395-4

Kruglanski, A. W., & Gigerenzer, G. (2011). Intuitive and deliberate judgments are based on common principles: Correction to Kruglanski and Gigerenzer (2011). *Psychological Review*, *118*(3), 97–109. DOI: 10.1037/a0023709

Langer, E. J. (1975). The illusion of control. *Journal of personality and social psychology*, *32*(2), 311. DOI: 10.1037/0022-3514.32.2.311

Lewis, D. (1983). New work for a theory of universals. *Australasian journal of philosophy*, *61*(4), 343–377. DOI: 10.1080/00048408312341131

Lilienfeld, S. O., Ammirati, R., & Landfield, K. (2009). Giving debiasing away: Can psychological research on correcting cognitive errors promote human welfare? *Perspectives on psychological science*, *4*(4), 390–398. DOI: 10.1111/j.1745-6924.2009.01144.x

Luhmann, C. C., & Ahn, W.-K. (2005). The meaning and computation of causal power: comment on Cheng (1997) and Novick and Cheng (2004). *Psychological review*, *112*(3), 685–707. DOI: 10.1037/0033-295X.112.3.685

Maguire, P., Moser, P., Maguire, R., & Keane, M. T. (2018). Why the conjunction effect is rarely a fallacy: How learning influences uncertainty and the conjunction rule. *Frontiers in psychology*, *9*, 1011. DOI: 10.3389/fpsyg.2018.01011

Malecki, K. M., Keating, J. A., & Safdar, N. (2021). Crisis communication and public perception of COVID-19 risk in the era of social media. *Clinical infectious diseases*, *72*(4), 697–702. DOI: 10.1093/cid/ciaa758

Matute, H., Blanco, F., & Díaz-Lago, M. (2019). Learning mechanisms underlying accurate and biased contingency judgments. *Journal of Experimental Psychology: Animal Learning and Cognition*, *45*(4), 373. DOI: 10.1037/xan0000222

Matute, H., Blanco, F., & Moreno-Fernández, M. M. (2022). Causality bias. In R. Pohl (Ed.), *Cognitive illusions: Intriguing phenomena in thinking, judgment, and memory* (3rd ed., pp. 108–123). Routledge. DOI: 10.4324/9781003154730-9

Matute, H., Blanco, F., Yarritu, I., Díaz-Lago, M., Vadillo, M. A., & Barberia, I. (2015). Illusions of causality: How they bias our everyday thinking and how they could be reduced. *Frontiers in psychology*, *6*, 888. DOI: 10.3389/fpsyg.2015.00888

Matute, H., Yarritu, I., & Vadillo, M. A. (2011). Illusions of causality at the heart of pseudoscience. *British Journal of Psychology*, *102*(3), 392–405. DOI: 10.1348/000712610x532210

Meyer, A., Frederick, S., Burnham, T. C., Guevara Pinto, J. D., Boyer, T. W., Ball, L. J., Pennycook, G., Ackerman, R., Thompson, V. A., & Schuldt, J. P. (2015). Disfluent fonts don't help people solve math problems. *Journal of Experimental Psychology: General*, *144*(2), e16. DOI: 10.1037/xge0000049

Michotte, A. (2017). *The perception of causality*. Routledge. DOI: 10.4324/9781315519050 (Original work published 1963)

Moreno-Fernández, M. M., Blanco, F., & Matute, H. (2021). The tendency to stop collecting information is linked to illusions of causality. *Scientific Reports*, *11*(1), 3942. DOI: 10.1038/s41598-021-82075-w

Morey, R. D., & Rouder, J. N. (2022). *BayesFactor: Computation of Bayes Factors for Common Designs* (R package; Version 0.9.12-4.4). Url: https://CRAN.R-project.org/package=BayesFactor

Morys-Carter, W. L. (2022). Vespr study portal. VESPR. Url: https://moryscarter.com/vespr/portal

Ng, D. W., Lee, J. C., & Lovibond, P. F. (2024). Unidirectional rating scales overestimate the illusory causation phenomenon. *Quarterly Journal of Experimental Psychology*, *77*(3), 551–562. DOI: 10.1177/17470218231175003

Novemsky, N., Dhar, R., Schwarz, N., & Simonson, I. (2007). Preference fluency in choice. *Journal of marketing research*, *44*(3), 347–356. DOI: 10.1509/jmkr.44.3.347

Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in cognitive sciences*, *12*(6), 237–241. DOI: 10.1016/j.tics.2008.02.014

Pastore, M. (2015). *Analisi dei dati in psicologia*. Il Mulino.

Pastore, M., Lionetti, F., & Altoè, G. (2017). When one shape does not fit all: A commentary essay on the use of graphs in psychological research. *Frontiers in psychology*, *8*, 303994. DOI: 10.3389/fpsyg.2017.01666

Pearce, J. M., & Bouton, M. E. (2001). Theories of associative learning in animals. *Annual review of psychology*, *52*(1), 111–139. DOI: 10.1146/annurev.psych.52.1.111

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Hochenberger, R., Sogo, H., Kastman, E., & Lindelov, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior research methods*, *51*, 195–203. DOI: 10.3758/s13428-018-01193-y

Perales, J. C., & Shanks, D. R. (2007). Models of covariation-based causal judgment: A review and synthesis. *Psychonomic bulletin & review*, *14*, 577–596. DOI: 10.3758/bf03196807

Pohl, R. F. (2022). *Cognitive illusions: Intriguing phenomena in thinking, judgment, and memory* (3rd ed.). Routledge. DOI: 10.4324/9781003154730

R Core Team. (2022). *R: A Language and Environment for Statistical Computing* (comp. software; Version 4.41). Url: https://www.R-project.org/

Reber, R., & Schwarz, N. (1999). Effects of perceptual fluency on judgments of truth. *Consciousness and cognition*, *8*(3), 338–342. DOI: 10.1006/ccog.1999.0386

Reber, R., Winkielman, P., & Schwarz, N. (1998). Effects of perceptual fluency on affective judgments. *Psychological science*, *9*(1), 45–48. DOI: 10.1111/1467-9280.00008

Rescorla, R. A., & Wagner, A. (1972). A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. *Classical conditioning, Current research and theory*, *2*, 64–69. Url: https://api.semanticscholar.org/CorpusID:51139715

Roediger, H. L. (1996). Memory illusions. *Journal of memory and Language*, *35*(2), 76–100. DOI: 10.1006/jmla.1996.0005

Roediger, H. L., & McDermott, K. B. (2000). Distortion in memory. In E. Tulving & F. Craik (Eds.), *The oxford handbook of memory*. Oxford University Press. DOI: 10.1093/oso/9780195122657.003.0010

Sauer, M. A., Truelove, S., Gerste, A. K., & Limaye, R. J. (2021). A failure to communicate? How public messaging has strained the COVID-19 response in the United States. *Health security*, *19*(1), 65–74. DOI: 10.1089/hs.2020.0190

Schwarz, N. (2004). Metacognitive experiences in consumer judgment and decision making. *Journal of consumer psychology*, *14*(4), 332–348. DOI: 10.1207/s15327663jcp1404_2

Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H., & Simons, A. (1991). Ease of retrieval as information: Another look at the availability heuristic. *Journal of Personality and Social psychology*, *61*(2), 195–202. DOI: 10.4324/9780203496398-11

Schwieter, J. W., & Ferreira, A. (2016). Effects of cognitive control, lexical robustness, and frequency of codeswitching on language switching. *Cognitive control and consequences of multilingualism*, *2*, 193–216. Url: http://digital.casalini.it/9789027266729

Shanks, D. R., Lopez, F. J., Darby, R. J., & Dickinson, A. (1996). Distinguishing associative and probabilistic contrast theories of human contingency judgment. In D. R. Shanks, D. L. Medin, & K. J. Holyoak (Eds.), *Psychology of learning and motivation* (pp. 265–311). Elsevier. DOI: 10.1016/s0079-7421(08)60563-0

Smedslund, J. (1963). The concept of correlation in adults. *Scandinavian journal of Psychology*, *4*(1), 165–173. DOI: 10.1111/j.1467-9450.1963.tb01324.x

Song, H., & Schwarz, N. (2008). Fluency and the detection of misleading questions: Low processing fluency attenuates the Moses illusion. *Social cognition*, *26*(6), 791–799. DOI: 10.1521/soco.2008.26.6.791

Stanovich, K. E. (1999). *Who is rational?: Studies of individual differences in reasoning*. Psychology Press. DOI: 10.4324/9781410603432

Tentori, K., Bonini, N., & Osherson, D. (2004). The conjunction fallacy: A misunderstanding about conjunction? *Cognitive Science*, *28*(3), 467–477. DOI: 10.1207/s15516709cog2803_8

Thompson, D. V., & Ince, E. C. (2013). When disfluency signals competence: The effect of processing difficulty on perceptions of service agents. *Journal of Marketing Research*, *50*(2), 228–240. DOI: 10.1509/jmr.11.0340

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *Science*, *185*(4157), 1124–1131. DOI: 10.1126/science.185.4157.1124

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453–458. DOI: 10.1007/978-1-4613-2391-4_2

Tversky, A., & Kahneman, D. (1996). On the reality of cognitive illusions. *Psychological Review*, *103*(3), 582–591. DOI: 10.1037//0033-295x.103.3.582

Vadillo, M. A., & Matute, H. (2007). Predictions and causal estimations are not supported by the same associative structure. *Quarterly Journal of Experimental Psychology*, *60*(3), 433–447. DOI: 10.1080/17470210601002520

Vadillo, M. A., Matute, H., & Blanco, F. (2013). Fighting the illusion of control: How to make use of cue competition and alternative explanations. *Universitas Psychologica*, *12*(1), 261–270. Url: http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1657-92672013000100023&lng=en&nrm=is

Vives, M.-L., Aparici, M., & Costa, A. (2018). The limits of the foreign language effect on decision-making: The case of the outcome bias and the representativeness heuristic. *PloS one*, *13*(9), e0203528. DOI: 10.1371/journal.pone.0203528

Waldmann, M. R., Hagmayer, Y., & Blaisdell, A. P. (2006). Beyond the information given: Causal models in learning and reasoning. *Current Directions in Psychological Science*, *15*(6), 307–311. DOI: 10.1111/j.1467-8721.2006.00458.x

Warren, G. W., & Lofstedt, R. (2022). Risk communication and COVID-19 in Europe: lessons for future public health crises. *Journal of Risk Research*, *25*(10), 1161–1175. DOI: 10.1080/13669877.2021.1947874

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly journal of experimental psychology*, *12*(3), 129–140. DOI: 10.1080/17470216008416717

Wason, P. C. (1966). Reasoning. In B. Foss (Ed.), *New horizons in psychology* (pp. 135–151). Penguin Books.

Wasserman, E. (1990). Detecting response-outcome relations: Toward an understanding of the causal texture of the environment. In G. H. Bower (Ed.), *Psychology of learning and motivation* (pp. 27–82). Elsevier. DOI: 10.1016/s0079-7421(08)60051-1

Wasserman, E., Dorner, W., & Kao, S. (1990). Contributions of specific cell information to judgments of interevent contingency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(3), 509–521. DOI: 10.1037//0278-7393.16.3.509

WHO. (2023, December 11). *WHO Coronavirus (COVID-19) Dashboard*. World Health Organization. Url: https://covid19.who.int/

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Yarritu, I., Matute, H., & Luque, D. (2015). The dark side of cognitive illusions: When an illusory belief interferes with the acquisition of evidence-based knowledge. *British Journal of Psychology*, *106*(4), 597–608. DOI: 10.1111/bjop.12119

Zuffi, S., Brambilla, C., Beretta, G., & Scala, P. (2007). Human computer interaction: Legibility and contrast, In *14th international conference on image analysis and processing (iciap 2007)*. IEEE. DOI: 10.1109/iciap.2007.4362786

Thanks

I would like to extend my gratitude to my Supervisor, Professor Michele Vicovaro, for his support, guidance, and dedication throughout my academic journey. His commitment to improving the quality of this work has been invaluable. I am also deeply thankful to Professor Massimo Nucci and Professor Andrea Spoto for their insightful feedback and continuous encouragement during my Master's degree. My sincere appreciation goes to Professor Massimiliano Pastore for his expert guidance in honing my skills in data analysis.

A special thanks to my parents, Sabrina and Emilio, for their love, patience, and unwavering belief in my abilities, which have been fundamental to my academic and personal development. I am also profoundly grateful to my fiancée, Eleonora, for her steadfast support and encouragement throughout this challenging yet rewarding journey. Her understanding and faith in me have been tremendous sources of strength and motivation. Finally, I wish to express my deep thanks to my friends for their camaraderie and support. Their presence has made this journey more enjoyable and less solitary.