INVESTIGATING MEANINGFUL CONSEQUENCES AS AN ATTENUATION STRATEGY FOR THE TRUTH EFFECT AND THE MOSES ILLUSION AS EXAMPLES OF COGNITIVE ILLUSIONS

Inauguraldissertation zur Erlangung des Doktorgrades der Humanwissenschaftlichen

Fakultät der Universität zu Köln nach der Promotionsordnung vom 18.12.2018

vorgelegt von

Felix Speckmann

aus Düsseldorf (Deutschland)

Vorgelegt am 18.09.2021

Diese Dissertation wurde von der Humanwissenschaftlichen Fakultät der Universität zu

Köln im Februar 2022 angenommen.

Table of Contents

Preface

Chapter 2.1 is based on the following publication:

Speckmann, F., & Unkelbach, C. (2021). Moses, money, and multiple-choice: The Moses illusion in a multiple-choice format with high incentives. *Memory & Cognition*, *49*(4), 843–862. https://doi.org/10.3758/s13421-020-01128-z

We developed all experiments together. I programmed all experiments, collected the data, and conducted all data analyses. We wrote the manuscript together.

Chapter 3.1 is based on the following manuscript:

Speckmann, F., & Unkelbach, C. (2021). *Monetary incentives do not reduce the repetition-induced truth effect.* Manuscript submitted for publication (to Psychonomic Bulletin & Review).

We developed the experiment together. The second author programmed the experiment. I collected the data and we conducted the data analyses together. We wrote the manuscript together.

Chapter 4.1 is based on the following manuscript:

Unkelbach, C., & Speckmann, F. (2021). *Believing repeated false information despite knowing better under high incentives.* Manuscript in preparation.

We developed all experiments together. The first author programmed all experiments and I collected the data. The first author conducted all data analyses and we wrote the manuscript together.

Chapter 5.1 is based on the following publication:

Unkelbach, C., & Speckmann, F. (2021). Mere repetition increases belief in factually true COVID-19-related information. Journal of Applied Research in Memory and Cognition, 10(2), 241–247. https://doi.org/10.1016/j.jarmac.2021.02.001

We developed the experiment together. I programmed all experiments and collected the data. The second author conducted all data analyses and we wrote the manuscript together.

Please note that some changes in formatting and citation style were undertaken to fit the layout of the dissertation. No changes were made to the content of the articles and manuscripts.

Acknowledgments

I want to thank my family for always supporting me on my academic path and all other domains of life for as long as I can remember.

I want to thank Christian for being a great supervisor and deepening my fascination with science, and I want to thank Oli for providing me with valuable perspectives from outside my own group.

I want to thank Mareike, Lea, Tobi, and Alex for their scientific and emotional support throughout my work at SoCCCo.

List of Figures

List of Tables

Abstract

Humans make countless decisions every day and inevitably make mistakes. One origin of such errors are cognitive illusions involving flaws in memory, judgment, or perception. On the one hand, they serve to shed light on the functioning of the human cognitive system, but on the other hand, they can also have negative consequences such as stereotyping or misinformation. In this dissertation, I will focus on two specific cognitive illusions and attempt to attenuate them: First, the Moses illusion, in which people read a distorted question with one term replaced by a semantically related but incorrect term, yet fail to notice this distortion and respond to the question as if it were correct. Second, the truth effect, in which people rate repeated judgments as truer than non-repeated judgments. By reviewing the existing literature, I show that among all the investigated possible moderators to weaken the effect, one approach has been largely missing: motivation. In six chapters, I investigate providing monetary incentives for correct responses (for the Moses illusion and the truth effect) and the relevance of statements presented (for the truth effect) as ways to increase motivation and attenuate both cognitive illusions. However, all motivational manipulations result in little attenuation at best (if any). I discuss the possible reasons for the failed attenuation and propose a combined approach for attenuation that includes both a motivational manipulation as well as teaching effective strategies to avoid cognitive illusions.

Deutsche Zusammenfassung

Menschen treffen täglich unzählige Entscheidungen und machen dabei zwangsläufig Fehler. Eine Ursache für solche Fehler sind kognitive Illusionen, die mit Fehlern im Gedächtnis, im Urteilsvermögen oder in der Wahrnehmung einhergehen. Sie dienen einerseits zum besseren Verständnis des menschlichen kognitiven Systems, andererseits können sie aber auch negative Folgen wie Stereotypisierung oder Fehlinformationen haben. In dieser Dissertation werde ich mich auf zwei spezifische kognitive Illusionen konzentrieren und versuchen, sie abzuschwächen: Erstens die Moses-Illusion, bei der Menschen eine verzerrte Frage lesen, in der ein Begriff durch einen semantisch verwandten, aber falschen Begriff ersetzt wurde, diese Verzerrung aber nicht bemerken und die Frage so beantworten, als ob sie korrekt wäre. Zweitens der Truth-Effekt, bei dem Menschen wiederholte Aussagen als wahrer einstufen als nicht wiederholte Aussagen. Unter Berücksichtigung der existierenden Literatur zeige ich, dass unter all den untersuchten möglichen Moderatoren zur Abschwächung des Effekts ein Ansatz weitgehend fehlt: Motivation. In sechs Kapiteln untersuche ich die Bereitstellung monetärer Anreize für korrekte Antworten (für die Moses-Illusion und den Wahrheitseffekt) und die Relevanz der präsentierten Aussagen (für den Wahrheitseffekt) als Möglichkeiten, Motivation zu erhöhen und beide kognitiven Illusionen abzuschwächen. Alle Motivationsmanipulationen führen jedoch bestenfalls zu einer geringen Abschwächung (wenn überhaupt). Ich erörtere die möglichen Gründe für die fehlgeschlagene Abschwächung und schlage einen kombinierten Ansatz

zur Abschwächung vor, der sowohl eine Motivationsmanipulation als auch die Vermittlung wirksamer Strategien zur Vermeidung kognitiver Täuschungen umfasst.

1 General Introduction

The human cognitive system is highly adaptive and has great computational power, but it is not immune to errors. The term “cognitive illusion” describes a deviation in memory, judgment, or perception from reality. This deviation is systematic, happens involuntarily and differs distinctly from the usual and expected behavior (for an overview, see Pohl, 2004). Well-known cognitive illusions include (among others) illusory correlations, where people mistakenly believe two attributes to be correlated because of their co-occurrence (e.g., Hamilton & Gifford, 1976), the anchoring effect, where the mere mention of a number influences people in the direction of that number in numerical judgments (e.g., Tversky & Kahneman, 1974), or confirmation bias, which describes the tendency to search mainly for information that confirms one’s own hypothesis (e.g., Oswald & Grosjean, 2004). Psychological researchers can investigate the conditions of these cognitive illusions to gain insight into the cognitive system, just as optical illusions can be investigated to better understand the perceptual system. For example, consider the Ames room (e.g., Gehringer & Engel, 1986). This room, when viewed through a peephole, appears to be a regular room with a rectangular layout. However, the back wall opposite to the viewer is diagonal, resulting in the left and right back corners being at different distances to the viewer. Since the viewer is oblivious to this information, their assumptions about the rectangular layout lead them to perceive the sizes of objects differently to the objects’ actual sizes, suggesting that perception involves more than “the simple pickup of information in the ambient optic array” (Gehringer & Engel, 1986, p.185). The optical illusion of the Ames room makes it clear how important perspective is for perception. Just as perception is much more than the mere absorption of optical stimuli, cognitive systems are influenced by processes that are susceptible to faulty inclinations an example for cognitive illusions, illusory correlations are perceived correlations between "two classes of events which, in reality, (a) are not correlated, or (b) are correlated to a lesser extent than reported, or (c) are correlated in the opposite direction from that which is reported" (Chapman, 1967, p. 151) and can lead to stereotypes against groups (Hamilton & Gifford, 1976), among other effects.

However, despite what their name might imply, cognitive illusions are adaptive in most cases. For instance, illusions of control lead people to believe they have more control over their lives than they actually do (Langer, 1975), and such perceived control is associated with positive outcomes including successful coping strategies (Thompson & Spacapan, 1991). Despite their overall adaptive value, cognitive illusions can also have negative real-life consequences. The truth effect (also called repetition-induced truth effect or truth by repetition effect), for example, describes the tendency to believe repeated information more than novel information (Unkelbach, Koch, Silva, et al., 2019). Usually, under the assumption that most of the statements people hear or read are true, this is an ecologically valid cue to judge the veracity of statements (Reber & Unkelbach, 2010). As an example, if you heard the statement "Freddy Mercury was born in Zanzibar" (which is true) from a friend, later from a co-worker, and later again during a trivia quiz on television, you would become increasingly likely to conclude that this statement is true. However, this mechanism is also true for false statements. Thus, the constant repetition of the statement "The 2020 U.S. election was fraudulent" could have led people to believe the statement and could have played a role in the storming of the U.S. Capitol in Washington, D.C. on January 6, 2021. Taken together, research on cognitive illusions is useful because on the one hand it provides insight into how the human cognitive system works and on the other hand it helps develop measures against exploitation of heuristics in human decision making to reduce negative consequences (e.g., fake news; Pennycook et al., 2018).

In this dissertation, I will focus on two specific cognitive illusions: the Moses illusion and the truth effect and what can be done to attenuate them. Research has addressed many aspects of these two illusions: the underlying processes, moderators and consequences. In Chapter 1.1, I will start with a summary of existing research on the truth effect. This includes the original paradigm, a comparison of the different explanations for the effect, and an overview of moderators that have been investigated. Chapter 1.2 will follow the same structure and discuss the original paradigm, explanations, and moderators for the Moses illusion. Chapter 1.3 will compare the truth effect and the Moses illusion and illustrate a gap in the existing literature on moderators for both illusions: motivation. Prior research has not yet examined what happens to these two illusions when participants are confronted with immediate consequences for their judgments. I will address this question in Chapter 2, in which we incentivized participants by giving them money for correct responses (i.e., resisting the truth effect) in a standard truth effect paradigm. Chapter 3 transfers the same research question and design to the Moses illusion. In Chapter 4, we investigated whether the truth effect persists even when the statements are highly relevant statements regarding the COVID-19 pandemic. In Chapter 5, we investigated the impact of advice on our incentivized truth effect paradigm. Lastly Chapter 6 discusses our findings and implications.

1.1 The Truth Effect

Prior research found that repetition of a stimulus can influence its perception, such as the liking of a stimulus (mere exposure; e.g., Zajonc, 1968) and the famousness ratings for names (false fame; Jacoby et al., 1989). Similarly, the truth effect is the phenomenon of people inferring trueness from repetition (e.g., Brashier & Marsh, 2020; Unkelbach, Koch, Silva, et al., 2019). The basic paradigm first used by Hasher et al. (1977) had participants read statements and later rate their validity in three separate sessions that were two weeks apart each. In each of those sessions, participants read 60 statements, 20 of which were shown in every session and 40 of which were only shown in that specific session. This means that 20 of the statements were repeated each session (i.e., “old”) and 120 were novel statements (i.e., “new”). Of all the statements, half were factually true (e.g., “Kentucky was the first state west of the Alleghenies to be settled by pioneers”) and half were factually false (e.g., ”Zachary Taylor was the first president to die in office”). The results showed that participants rated old statements as more valid than new statements. More recent research often uses a variant of this paradigm that only involves one session (e.g., Silva et al., 2017): Participants first complete a presentation phase, where a random sample of a larger group of statements are presented. After this presentation phase, participants enter a judgment phase where they see the complete group of all statements (including the previously presented ones). This allows for differentiating between repeated statements (shown in both phases) and non-repeated statements (only shown in the judgment phase) in just one session. Table 1.1 shows ten example statements from research by Jalbert et al. (2019).

Table 1.1

Ten example statements from Jalbert et al. (2019) in correct and incorrect versions each. Differing terms are italicized.

Correct statements	Incorrect statements
The electron is the *lightest* charged particle found in nature.	The electron is the *heaviest* charged particle found in nature.
In chemistry, a mass spectrometer is used to separate substances into its constituent parts according to *mass*.	In chemistry, a mass spectrometer is used to separate substances into its constituent parts according to *color*.
The highest waterfall in the world is in *Venezuela*.	The highest waterfall in the world is in *Argentina*.
Vesuvius is an active volcano in *Italy*.	Vesuvius is an active volcano in *France*.
Mogul skiing is the navigation of large bumps on the ski slope.	*Slalom* skiing is the navigation of large bumps on the ski slope.
Birling is the sport of *logrolling*.	Birling is the sport of *woodchopping*.
Gnu is another name for *wildebeest*.	Gnu is another name for *antelope*.
The *whale* shark subsists on plankton.	The *white* shark subsists on plankton.
Dough is boiled in the process of making *bagels*.	Dough is boiled in the process of making *croissants*.
Most limes have *more* acid than lemons.	Most limes have *less* acid than lemons.

1.1.1 Explanations for the Truth Effect

Based on their findings, Hasher et al. (1977) concluded that frequency of occurrence is the underlying process for the truth effect. This interpretation makes sense, since repetition (old vs. new) was the only factor they manipulated between statements. However, following research proposed several alternative explanations for the truth effect.

Subjective Recognition. Bacon (1979) argued that the findings could also be due to a different process. Participants could have judged statements as truer when they recognized statements as repeated, regardless of whether the statements were factually repeated or not. Testing this hypothesis required asking participants whether they recognize any statements, thus the author modified the basic paradigm used by Hasher et al. (1997). Instead of three sessions, he only used two sessions, but before rating truth, participants indicated whether they recognized a given statement. This allowed differentiating between frequency of occurrence and subjective recognition of statements. The results showed that repeated statements were judged as truer than novel statements, replicating the basic repetition effect. However, the truth ratings were more sensitive to recognition than to factual repetition, that is, participants rated statements they judged to be old as truer than statements they judged to be new. This suggests that the underlying process of the truth effect is due to perceived repetition rather than actual repetition. If a statement were shown repeatedly but not recognized, it *should not* be rated as truer. Conversely, if a statement were judged as repeated even though it factually was not, it *should* be rated as truer.

Familiarity. Building on the notion that the effect works on the side of the individual (subjective recognition) rather than on the side of the statements (objective frequency), Arkes et al. (1989) and Begg et al. (1992) investigated whether subjective feelings of familiarity cause the truth effect. Begg et al. (1992) combined statements with sources of varying reliability: Statements were either preceded by a male or a female name (e.g., "XXX says that…") and participants in the respective conditions were told that female names would always tell the truth and male names would always lie, or vice versa. If participants correctly recalled the source of a statement, they should rate old statements from a lying source as less true than old statements from an honest source or new statements. Participants reliably but modestly discriminated between the sources and rated old statements coupled with an honest source as truer than old statements coupled with a lying source. However, they also rated new statements as less true than old statements from a lying source. The authors concluded that the only viable explanation for their results were the independent contributions of source recollection and familiarity.

Fluency. A more general phenomenon that is compatible with the role of familiarity in truth judgments is processing fluency. When processing fluency is high, people experience mental processes as easier. Previous research had shown that increased processing fluency leads to increased feelings of familiarity (e.g., Whittlesea et al., 1990). Based on these findings, Reber and Schwarz (1999) investigated if increased fluency leads to higher truth ratings independent of presentation frequency. They presented participants with statements in different colors (some easy to read, such as blue or red, and some hard to read, such as yellow or light blue) to manipulate perceptual fluency and asked participants to indicate whether they thought a given statement was true or false. Importantly and unlike in the previously mentioned experiments, this experiment did not include a prior presentation phase, and thus any differences in truth judgments were due to the color manipulation, not presentation frequency. The authors found that statements presented in highly visible colors were rated as true more often, suggesting that any variable that increases fluency (such as familiarity) also increases ratings of truth.

Lev-Ari and Keysar (2010) showed that fluency manipulations are not limited to visual processing: Participants in their experiment listened to statements spoken by different voices and rated statements spoken by voices with foreign-sounding accents as less true than statements spoken by voices without accent. Importantly, the experiment explained to participants that the people behind the voices were not the source of the statement but a messenger relaying the statement, which should have ruled out possible effects of prejudice.

Further research by Hansen et al. (2008) replicated and extended the findings of Reber and Schwarz (1999) by showing that changes in fluency had more influence on truth ratings than the absolute level fluency, and McGlone and Tofighbakhsh (2000) showed that participants rated rhyming aphorisms (e.g., “Woes unite foes”, p. 426) as more accurate than non-rhyming aphorisms (e.g. “Woes unite enemies”, pp. 426). Importantly, participants who were told to distinguish between poetic quality and content of an aphorism did not use the rhyming fluency cue to judge accuracy. This finding is compatible with research by Unkelbach (2006, 2007) who showed that participants can learn to interpret the fluency cue differentially. In the second experiment in Unkelbach (2007), participants first complete a learning phase, in which they respond to statements and receive feedback. Depending on the condition, either true statements were presented in high contrast (leading to high fluency) and false statements were presented in low contrast (leading to low fluency), vice versa, or there was no correlation between fluency and truth (control condition). After the learning phase, participants completed a test phase, in which they were asked to indicate whether they believed each statement to be true or false. As predicted, participants responded with “true” more often to high-fluency statements only in the condition where high-fluency statements were true. In the condition where high-fluency statements were false, this pattern was reversed. Together with earlier research (Unkelbach, 2006) these findings suggest that fluency does not lead to higher truth judgments per se, but that participants learn to interpret fluency as a cue for correctness (or falseness). This finding is consistent with later research by Scholl et al. (2014), who found that participants rely more on fluency as a cue when they previously learned that relying on fluency led to correct judgments. In the real world, the majority of information one is exposed to is true rather than false, thus the use of fluency as a cue for truth is generally justified (Reber & Unkelbach, 2010).

Referential Theory. Unkelbach and Rom (2017) propose a referential theory as an explanation for the truth effect. They assume that reading information activates memory references related to the information. For example, when reading the statement “Freddy Mercury was born in Zanzibar”, people need memory references (i.e., nodes) for Freddy Mercury, Zanzibar, and being born to understand the statement. Perceived truth of the statement is a function of the coherence of all activated memory nodes. In this case, all humans are born at the beginning of their lives, so Freddy Mercury and being born are coherent memory nodes, and Zanzibar is a real (birth)place, so being born and Zanzibar are also coherent. This leads to the formation of coherent links between all memory nodes.

On the other hand, the statement “Freddy Mercury was born on the moon” is incoherent because the moon is not a birthplace, so being born and moon are incoherent, resulting in less coherent links than in the previous example.

People judge truth as a function of the number of corresponding memory nodes and their coherence. For the statement about Freddy Mercury, both versions have the same number of corresponding memory nodes, but overall coherence is higher in the first version, thus people should rate this version as truer than the second version. Furthermore, upon reading the statement, people already form the respective information networks and coherent links. Compared to a novel statement, for which no coherent links have yet been formed, the repeated statement thus has more coherent links and is judged as truer.

Integrating All Explanations. It is important to note that these explanations are not exclusive; any number of them can independently or jointly lead to the truth effect (Unkelbach, Koch, Silva, et al., 2019). For example, a participant in a typical truth effect paradigm might read the statement “The first windmills were built in Persia” (Unkelbach & Stahl, 2009) during the presentation phase at the beginning of the experiment. When reading the statement again in the judgment phase and asked to judge it as true or false, the participant might recognize the statement from before, leading to higher subjective frequency. At the same time, the activated memory network might lead to increased processing fluency of the statement. Both of these processes, as well as their interaction, would thus lead to increased truth ratings . This multitude of pathways leading to the truth effect may be a reason why the truth effect is so robust.

1.1.2 Robustness of the Truth Effect

The truth effect is an overall robust effect, but several moderating variables influence its strength (for a review, see Dechêne et al., 2010). Since the topic of this dissertation is the attenuation of the truth effect, this chapter discusses several potential moderators.

The Role of Knowledge. Some researchers argue that the truth effect only emerges when statements are not fully known (e.g., Dechêne et al., 2010). If one were absolutely certain about the truth status of a statement (e.g., the name of one's parents), repetition would not increase its perceived truth. Brown and Nix (1996) provided participants with feedback containing the factual truth status of statements after their judgment in the first session. In the second session, participants rated old false statements as less true than new false statements, indicating they learned and remembered the factual truth status. Consequently, they also rated old true statements as truer than new true statements. However, labelling false statements as such only had an effect when the second session was conducted one week after the first session (short retention interval), rather than after a month (long retention interval). Supporting evidence for this constraint of the truth effect also comes from earlier research by Begg et al. (1992) who showed that repeated statements from an explicitly lying source were rated as less true than repeated statements from an honest source.

However, Fazio et al. (2015) demonstrated that participants can neglect their own knowledge when faced with the truth effect. In their first experiment, they adapted a typical truth effect paradigm: Participants first saw 88 statements in the presentation phase, then rated 176 statements in the judgement phase. Additionally, participants then completed a third phase, in which they responded to multiple-choice questions about the previously seen statements (e.g., "What is the largest ocean on Earth? – "Atlantic", "Pacific", "don't know") to assess participants' knowledge. The results showed that participants still relied on repetition as a cue for truth, even when they knew the correctness of a statement. Knowledge did not interact with repetition, leading to an increase of rated truth for both known and unknown statements. The authors conclude that participants show "[k]nowledge neglect, or the failure to appropriately apply stored knowledge" (p. 1000), a phenomenon similar to the Moses illusion (see Chapter 1.2).

Unkelbach and Greifeneder (2018) extended these findings by providing participants with highly valid advice. In their fourth experiment, participants first completed a standard presentation phase. In the judgment phase, however, participants received advice from another person to help judge the truth of a given statement. This was represented by a schematic face with a speech bubble either saying a statement was true or false, combined with the average validity of that person's advice. Importantly, one advisor always gave 50% valid advice (i.e., guessing) and the other advisor gave 100% valid advice (i.e., always correct). The results showed that 100% validity had higher impact on truth ratings than 50% validity. However, the influence of repetition was significant for both validity levels, indicating that people still relied on repetition as a cue for truth despite having highly valid advice. This is compatible with research from Skurnik et al. (2005) who showed that repeatedly marking a statement as false decreased true ratings after 30 minutes but increased true ratings after a 30 day interval. Similarly, Nadarevic and Aßfalg (2017) found that showing participants a warning label before the presentation phase that explicates the truth effect and asks participants to not rely on repetition as a cue did diminish the truth effect, but could not eliminate it entirely.

Further evidence from Fazio et al. (2019) showed that the truth effect occurs irrespective of the plausibility of the statement. If knowledge protects against the truth effect, then highly plausible statements that are generally known to be true (e.g., "Most Americans have ridden in a vehicle of some sort", p. 1707) and highly implausible statements that are generally known to be false (e.g., "The Earth is a perfect square", p. 1707) should be less affected by it. The authors constructed ten groups of statements of decreasing plausibility and presented participants with half of the statements of each group during the presentation phase and all statements during the judgment phase, asking participants to indicate whether each statement was true or false. They computed the truth effect within statements by subtracting the mean “true” ratings of the non-repeated cases of a statement from the “true” ratings of the repeated cases of a statement. This “within-statements truth effect” (rather than “within-persons”) did not vary across different levels of plausibility, suggesting that even supposedly trivial statements are nevertheless affected by the truth effect.

However, research by Brashier et al. (2020) showed that existing knowledge can lead to a reduction of the truth effect when coupled with an accuracy focus. In their experiments, the authors extended the typical presentation phase of a truth effect paradigm by an additional task: Depending on the condition, participants rated all presented statements on how interesting they were or how true they were (accuracy focus). After this, participants completed a judgment phase as usual. The results showed that factually false statements did not receive significantly higher truth ratings when repeated compared to novel only for participants in the accuracy focus condition who possessed the relevant knowledge. These findings suggest that if people have the relevant knowledge and are told to judge statements’ truth as soon as they are presented (i.e., focus on their accuracy), the effect of repetition on rated truth can be diminished.

It seems that while some prior research argued that knowledge alone protects against the truth effect, more recent research suggests that this is not the case. Neither having the relevant knowledge, nor receiving advice from a highly valid source with the relevant knowledge, nor judging highly plausible statements, eliminated the truth effect. However, prior knowledge combined with an instruction to judge statements' truth immediately in the presentation phase did decrease the truth effect.

Source Effects. As mentioned above, source credibility influences the truth effect. Begg et al. (1992) found that repeated statements from untruthful sources are rated as less true than statements from a truthful source. However, statements from untruthful sources were still rated as truer than new statements, indicating that the truth effect persists despite explicit labelling of sources as untruthful, again demonstrating the robustness of the effect. Furthermore, when no explicit source is given, attributing a statement to a within-experiment source leads to lower truth ratings than attributing the statement to a source outside of the experiment: Arkes et al. (1989) found that participants rated statements as truer that they subsequently claim to have heard before partaking in the experiment. This finding was replicated by Law et al. (1998) who found that statements attributed to outside sources are rated as truer than statements attributed to the first session of the experiment, but both kinds of repeated statements are rated as truer than new statements. Taken together these findings suggest that information about the source of a statement may moderate the truth effect but does not eliminate it.

Individual Differences. Finally, apart from features of the experimental situation, the strength of the truth effect may be moderated by individual differences (e.g., personality traits). Newman et al. (2020) investigated the role of Need for Cognition (Cacioppo & Petty, 1982). The authors added one extra condition to a typical two-session (three days apart) truth effect paradigm: Half the participants saw a warning prior to the presentation of the statements that stated: "half the statements are true and half the statements are false" (p.

9), the other half did not see any warnings. After the truth ratings in the second phase, participants completed the Need for Cognition scale (Cacioppo & Petty, 1982). The results showed a typical truth effect (higher truth ratings for repeated statements) and participants high on Need for Cognition seemed to show a larger truth effect when no warning was given before the presentation phase. However, an internal meta-analysis across both experiments showed no significant influence of Need for Cognition.

De keersmaecker et al. (2020) investigated additional individual factors that might moderate the truth effect: cognitive ability (i.e., information processing capacity), need for cognitive closure (Kruglanski & Webster, 1996), and cognitive style (i.e., intuitive vs. deliberate; Evans & Stanovich, 2013). Across seven experiments, they used both the two-session (i.e., presentation phase and rating phase five to seven days apart) and the one-session paradigm (i.e., presentation phase and rating phase within the same session) and extended each experiment with the respective measurement tools for the individual factors. They replicated the basic truth effect in each of the seven experiments but found no evidence for an influence of general cognitive ability, need for cognitive closure, preference for intuition, or preference for deliberation.

Overall, the research on potential moderators shows that prior knowledge, source effects, and individual differences have little, if any, effect on the truth effect. Along with the fact that virtually all studies on the truth effect replicated the basic effect, this speaks for the general robustness of the truth effect. However, motivation as a potential moderator has so far not been investigated, and Chapter 1.3 will explain why this could be an effective moderator.

1.2 Moses Illusion

When asked, “How many animals of each kind did Moses take on the ark?”, most people respond “two”, even though they know that it was Noah who took the animals on the ark in the Biblical story. This tendency to overlook a distortion in a sentence and respond as if the sentence were not distorted is known as the Moses illusion (e.g., Park & Reder, 2004). This illusion was first demonstrated by Erickson and Mattson (1981), who presented participants with several questions. For some of these questions, they replaced one term with a semantically related, but different term (i.e., “Moses” instead of “Noah”). Importantly, this term was never the target of the question, as the question asked for the number of animals. Thus, participants can answer the question with or without noticing the distortion. If participants noticed any distortions, they were asked to respond with “can’t say”, but if participants did not notice the illusion, they answered as they would if the question were not faulty (i.e., “two”). I will refer to this type of response as “Moses response” from now on. To reiterate, the correct response to a distorted question (i.e., “How many animals of each kind did Moses take on the ark?”) was “can’t say”, while the correct response to an undistorted question (i.e., “How many animals of each kind did Noah take on the ark?”) depended on the question (in this case, “two”). To ensure that they had the relevant knowledge for responding, participants in the experiments by Erickson and Mattson (1981) also responded to direct questions targeted at the critical term at the end of the experiment (i.e., “Who took two animals of each kind on the ark?”) and only participants who answered those questions correctly were included in the analysis. The results showed that participants responded with Moses responses most of the time (21 out of 26 for the original Moses question), providing the first evidence for this illusion.

Based on the research by Erickson and Mattson (1981), a typical paradigm investigating the Moses illusion involves three important steps: informing participants that some questions may be distorted, presenting participants with both distorted and undistorted questions, and confirming if participants had the relevant knowledge after the response phase.

1.2.1 Explanations for the Moses Illusion

Several potential explanations exist for this illusion, differing in their compatibility with existing evidence (Park & Reder, 2004). In the following, I will briefly discuss all of them with particular focus on the explanation that is compatible with the most evidence.

Cooperative Communication. Grice (1975) suggests that cooperative communication should not include trick questions, such as the distorted questions in a Moses illusion paradigm. Thus, when participants see a distorted question, they might assume that this distortion was unintentionally inserted by the experimenters and willfully ignore the distortion to provide the response that would be correct if the question were not distorted. This would imply that participants do notice the illusion but decide to ignore it. However, later research by Reder and Kusbit (1991) provided evidence against this explanation. The authors implemented two different conditions into the basic Moses illusion paradigm. Participants in the “gist” condition should respond to all questions while ignoring possible distortions. This means that even if they noticed a distortion in a question, they should respond as if the question were undistorted. Participants in the “literal” condition, on the other hand, should respond with “can’t say” to distorted questions, as in the basic paradigm. Assuming that participants in both conditions notice the distortion, it should be easier to ignore the distortion (the “gist” condition) than to take it into account (the “literal” condition), resulting in a greater (or equal) number of errors in the “literal” condition. However, the results showed that participants in the “gist” condition committed less errors than participants in the “literal” condition, suggesting that participants failed to notice the distortions and thus making responding in the “literal” condition easier, and contradicting the cooperative communication explanation.

Imperfect Encoding. If a cooperative communication setting is not the reason for the Moses illusion, then it is likely that the perception or processing of the question is flawed. The imperfect encoding explanation suggests that, when reading a distorted question, participants encode it in its undistorted form (i.e., “Noah”). This would result in the distorted term never being encoded in the participants’ mental representation of the question and thus, participants would not be able to detect the distortions.

However, to control for this explanation, Erickson and Mattson (1981) had participants read each of the questions aloud before responding to them. Failing to encode a question correctly should result in participants reading the question in its incorrect (i.e., undistorted) form. Yet, participants read the questions aloud in their correct (i.e., distorted) form, and still gave Moses responses, suggesting that any errors related to perception or processing of the question did not emerge at the encoding stage.

Imperfect Retrieval. When responding to a question, participants need to match the encoded question with relevant memory structures to find an answer. Even if participants correctly encode questions, they could commit errors when retrieving relevant knowledge structures from memory. For example, when responding to the original Moses question, participants could focus on retrieving information relevant to the answer to the question (i.e., how many animals), rather than retrieving all associated information that would enable participants to detect the distortion but would be irrelevant to answer the question (i.e., who took the animals).

Reder and Kusbit (1991) argued that if the Moses illusion were due to imperfect retrieval, improving memory retrieval should diminish the illusion. In Experiment 2, they extended their previous design ("gist" and "literal" conditions) by having participants read information related to the questions (i.e., statements) before the rating phase. The statements were based on the undistorted versions of half of the questions (e.g., "Noah took two animals of each kind on the ark"). The imperfect retrieval explanation would predict that participants give less Moses responses to questions for which they have previously studied the respective statements. However, while questions with previously studied statements did receive more correct responses overall, the general pattern from Experiment 1 was identical: Participants in the "literal" condition committed more errors than participants in the "gist" condition. This pattern is inconsistent with the predictions of the imperfect retrieval explanation, as studying a statement should have made it harder to ignore distortions (relevant to participants in the "gist" condition) and easier to detect distortions (relevant to participants in the "literal" condition), making the imperfect retrieval explanation unlikely.

Partial Matching. Given that participants do not willfully ignore distortions, that they correctly encode the questions, and that they correctly retrieve relevant information from memory, Reder and Kusbit (1991) suggested that participants do not perfectly match the encoded question with their retrieved memory structures. Full matching would require matching each word to memory separately, which takes time and effort. A quicker and easier method would be partial matching, which only requires matching some of the words in a question to memory structures, accepting smaller inconsistencies and treating a question as valid if it is close enough to memory structures. Prior research found that the Moses illusion depends on the relatedness between the distorted and undistorted term. Erickson and Mattson (1981), for example, found that the illusion disappears when Noah was replaced with Nixon, but persists when Noah was replaced with Adam or Abraham (Exp. 3). Kamas et al. (1996) argue that this is because when participants read a question, activation spreads through a semantic network of related memory nodes (see also Kamas & Reder, 1995). The stronger two nodes are connected, the easier activation flows between them, and the more nodes are shared between two terms, the stronger they are connected, making distortion detection more difficult. For example, Noah and Moses from the original paradigm (or Noah and Abraham from Erickson and Mattson (1981)) share many common nodes. They are usually depicted as old men, their stories involve water, and they appear in the Old Testament. If each of those features is modelled as a node in the semantic network, the resulting connection between Noah and Moses is strong. Replacing Moses with Nixon, however, leads to only few shared nodes and a weaker connection of the distorted and undistorted term. The likelihood of detecting a distortion should be a direct function of the connectedness of both terms in the semantic network: The distortion becomes increasingly harder to detect, the more the two terms are related.

Considering all available evidence, partial matching appears to be the most likely explanation for the Moses illusion (cf. Speckmann & Unkelbach, 2021b) which is why I will focus on this explanation from here on.

1.2.2 Robustness of the Moses Illusion

Similar to the truth effect, the Moses illusion is a robust phenomenon, which has reliably been replicated many times (for an overview, see Speckmann & Unkelbach, 2021b).

Nevertheless, previous research has found some moderators influencing the illusion, which I will discuss in the following.

Question Features. Bredart and Modolo (1988) adapted the basic paradigm by using statements instead of questions and varying whether the statement focuses on the distorted term or not. In the first condition, for example, participants would see statements such as “It was Moses who took two animals of each kind on the ark” (focus on the distorted term), whereas participants in the second condition would see statements such as “It was two animals of each kind that Moses took on the ark” (focus on the number of animals). The results indicated that such a focus on the distorted term led to fewer Moses responses. However, a conceptual replication by Kamas et al. (1996) used a different focus operationalization and found more complicated results. In their study, they used bold font for some of the words in the presented statements: either for the (un)distorted term (i.e., “MOSES” or “NOAH”), or the response part (i.e., “TWO”). Which terms were focalized was identical to Bredart and Modolo (1988). Kamas et al. (1996) found a decrease in Moses responses as well. But, since this decrease in “can’t say” responses happened in both conditions, the authors further conclude that the focalization led to a shift in response bias rather than actual sensitivity for distortions.

Situational Factors. While processing fluency is not a critical part of the partial matching explanation, Song and Schwarz (2008) argued that it may nevertheless influence distortion detection. In their experiment, the authors varied between conditions how well readable the questions were: In one condition participants received questions in easy-to-read font (high fluency), in the other condition they received questions in hard-to-read font (low fluency). Participants in the hard-to-read condition gave more Moses responses.

However, the authors used only two different questions per experiment (one distorted, one undistorted) and did not manipulate distortion within questions. This procedure limits the generalizability of the results (Judd et al., 2012) and cannot rule out the possibility of a response bias shift (cf. Kamas et al., 1996). Moreover, a more recent replication used all of the original materials but failed to replicate the original effect (Levordashka & Lakens, 2015).

In a similar vein, Lee et al. (2015) used fish odor in order to elicit feelings of suspicion to attenuate the Moses illusion. The authors prepared their experimental booths with either water (control condition) or fish oil (“fishy smell”) before presenting participants with a distorted and an undistorted question. Participants in the “fishy smell” condition gave fewer Moses responses than participants in the control condition, but as this experiment also consisted of only two questions and no within-questions manipulation, the same limitations as in Song and Schwarz (2008) apply (i.e., generalizability, possible bias shift), and the results are rather questionable.

Individual Differences. The partial matching explanation of the Moses illusion posits that memory structures influence illusion strength. A direct and testable consequence of this assumption would be that people fall for the illusion to a differing extent based on their individual memory capacities. Hannon and Daneman (2001) extended the basic paradigm by assessing two individual measures related to memory: knowledge access (i.e., how well participants can access prior knowledge in long-term memory) and working memory span (i.e., participants’ processing and storage capacity while reading). They measured knowledge access by presenting participants with statements to judge as true or false that require prior knowledge to answer (e.g., “A ROBIN lives in Canada, whereas a PENGUIN typically doesn’t”, p. 452). To measure working memory span, they had participants read aloud sentences that become increasingly longer and, at the end of a set, asked participants for the last word of each sentence. All participants also responded to a series of distorted and undistorted questions. The results showed that 36% of sensitivity for distortions was explained by the combination of these two individual measures, suggesting that individual differences in memory capacity can influence susceptibility to the Moses illusion.

Since memory structures seem to play an important role, Umanath et al. (2014) reasoned that older age might influence the Moses illusion as well, since older people have more lifetime memories than younger people. They presented two groups of participants (older adults vs. younger adults) with distorted and undistorted questions and added an additional phase at the end of the experiment, in which they asked participants open ended questions about the (un)distorted term (e.g., “Who was it that took two animals of each kind on the Ark?”) to check if participants erroneously adapted their knowledge structures to match the distortion. The results showed that older adults made more errors in the first phase (distortion detection) but were less suggestible in the second phase, that is, they gave less Moses responses when asked directly about the distorted term. This pattern can be explained by partial matching: The stronger prior knowledge of older adults might lead to stronger semantic networks, which trigger a partial match earlier than the weaker semantic networks of younger adults, but stronger prior knowledge is also likely to be more resistant against change (leading to less errors when asked directly).

Overall, the general robustness of the Moses illusion is not influenced by many moderators, but similar to the truth effect, the role of motivation has not yet been investigated (see Chapter 1.3.1 for why this might be promising). When controlling for the difference between a general shift in response bias (i.e., responding “can’t say” to all questions regardless of distortion) and genuine sensitivity, the only remaining significant moderators seem to be individual differences related to memory capacity, in line with the partial matching explanation.

1.3 Comparing the Truth Effect and the Moses Illusion

Despite their ostensible differences, the truth effect and the Moses illusion have much in common. First, both occur because people use heuristic cues to form a judgment about a question or statement. In case of the truth effect, processing fluency can serve as a cue for truth. When participants experience the processing of a statement as easier, they more likely rate a statement as true. In case of the Moses illusion, participants do not engage in full matching between the question and their relevant knowledge structures. Instead, they likely judge heuristically if the question is close enough to their memory structures and proceed to answer the question without further scrutiny if the fit is high enough. Second, the semantic network model proposed by Kamas et al. (1996) to explain the Moses illusion is similar to the referential model proposed by Unkelbach and Rom (2017) to explain the truth effect. In the referential model, incoming information activates the corresponding memory nodes and coherent links are formed between these memory nodes. The average coherence between nodes is then used as a cue to judge truth of statements. Previously presented statements already have an established information network and thus, more coherently linked references. Similarly, in the semantic network model, some nodes are assumed to already have shared connections (e.g., “Moses” and “Noah”) and the more shared connections the nodes pertaining to the distorted and undistorted terms have, the more likely a person will accept a distorted question. In terms of the referential model, people more likely fail to notice distortions when the respective memory network has more coherent nodes.

Given their similarity, it is perhaps unsurprising that both illusions share a general robustness against moderators. Both for the truth effect and the Moses illusion, moderators targeting different parts of the illusion (e.g., prior knowledge, fluency, etc.) showed small, if any, effects. However, one possible moderator that has so far not received much attention, is motivation.

1.3.1 Motivation as a Possible Moderator

Most previously mentioned moderators had in common that they target the process involved in the illusion. However, it should also be possible to target the participants directly. If one views correct judgments in both illusions as task performance, it makes sense to explore the role of motivation. Research on achievement motivation has shown that participants with high achievement motivation outperform those with low achievement motivation (Lowell, 1952) both in learning tasks (i.e., when the task is novel to all participants) and effort tasks (i.e., simple math problems). The effort task involved participants solving arithmetic addition problems and was later replicated by Biernat (1989), suggesting that a high achievement motive leads to higher performance on tasks requiring high mental concentration (Brunstein & Heckhausen, 2018). Both the truth effect and Moses illusion do not require high mental concentration per se, but the conscious decision to invest more cognitive effort could lead to a reduction in heuristic decision making and thus to fewer errors.

In a similar vein, research in organizational psychology has investigated the role of task motivation. For example, Locke (1968) argued that setting goals leads to increased task motivation, which in turn leads to increases in attention, effort, persistence, and strategy development (Locke et al., 1981). Similarly, Terborg (1976) asked participants to work on text material about electricity at their own pace and complete a test on the material at the end of the experiment. He divided participants into two conditions: One group would receive compensation in form of an hourly pay, and the other groups would receive monetary compensation based on how quickly participants worked through the material. Participants receiving the hourly pay invested significantly less effort (measured as the percentage of time spent on the task, rather than breaks) than participants who received performance-contingent compensation. Taken together, these findings suggest that it is possible to motivate people to increase their effort and, subsequently, their performance (see also Heyman & Ariely, 2004, Exp. 2, for a more recent demonstration).

Increased effort may also reduce errors in both the truth effect and Moses illusion paradigms. Specifically, higher effort intensity (Bonner & Sprinkle, 2002) as a measure of the amount of cognitive resources directed at a task, could lead participants to avoid the heuristic decision strategies that likely lead to the illusions. For example, when confronted with a Moses question, a participant with high effort intensity may engage in less partial matching, that is, they may compare more of the words in a sentence with their memory structures. While this still would not qualify as complete matching, it could raise the probability of detecting a distortion. Similarly, higher effort intensity in a truth paradigm could lead to more precise recall of a statement's source or reduce the reliance on intuition (i.e., fluency), which could lead to a reduction in subjective truth. Assuming that increased motivation leads to an attenuation of the Moses illusion and truth effect, it makes sense to investigate ways to increase motivation, such as direct consequences for the participants.

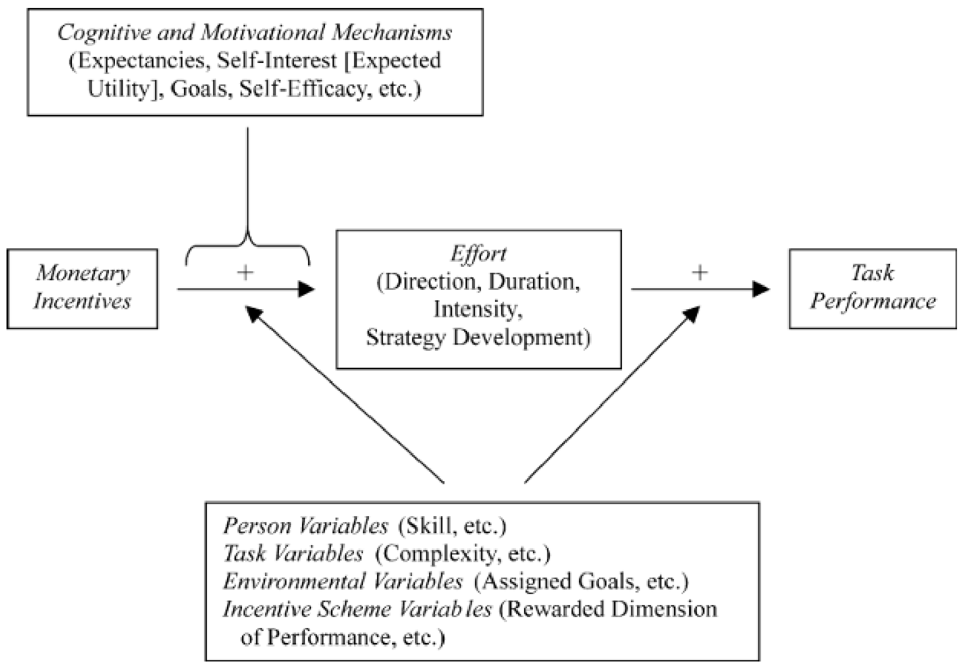
1.3.2 Meaningful Consequences

Most research on the truth effect and Moses illusion has used paradigms that do not involve direct consequences for participants' decisions. For example, in a typical truth effect study, participants first see a set of statements in a presentation phase. In the following judgment phase, they then decide for a larger set of statements (including the previously seen statements) whether each individual statement is true or false. However, the correctness of this decision is usually ignored, and participants receive no feedback on their accuracy. To raise task motivation, one could add consequences to each decision. Chapters 2-4 and 5 illustrate two types of consequences: direct short-term monetary consequences, and potential indirect health consequences.

Monetary incentives are commonly assumed to be an effective way to increase motivation (e.g., Cerasoli et al., 2014). Bonner and Sprinkle (2002) propose a general three-part framework in which monetary incentives influence effort, which in turn influences task performance. Cognitive and motivational mechanisms such as expectancy and goals would mediate the link between incentives and effort, and person variables (e.g., skill) and task variables (e.g., complexity) would moderate both the link between incentives and effort, as well as the link between effort and performance (Figure 1.1).

Figure 1.1

Conceptual framework for the effects of performance-contingent monetary incentives on effort and task performance.



Note: From "The effects of monetary incentives on effort and task performance: Theories, evidence, and a framework for research", by S. E. Bonner & G. B. Sprinkle, 2002, *Accounting, Organizations and Society*, 27(4–5), p. 303–345. Copyright 2002 by Elsevier Science Ltd.

Recently, Enke et al. (2021) investigated the relationship between incentives and performance for several well-known cognitive biases. They implemented three incentive conditions: High incentives, standard incentives, and no incentives. To feasibly offer high incentives, they conducted their research in Kenya, where purchasing power parity (a measure for prices at different locations) is much lower than in the United States. Due to this, participants in the low incentive conditions received $23.50 at purchasing power parity in the United States and participants in the high incentive condition could earn a bonus of up to $2,350 at purchasing power parity in the United States. Participants worked on questions related to two different biases and responded to two questions for each. The authors found no substantial differences between conditions for the base rate neglect task (e.g., Gigerenzer & Hoffrage, 1995), nor the Wason selection task (e.g., Cosmides, 1989), nor the anchoring task (e.g., Tversky & Kahneman, 1974). Only in the cognitive reflection test (Frederick, 2005) did high incentives slightly improve performance. In this task, participants respond to questions that have an intuitive (but incorrect) answer and a correct answer. The correct answer to the question "A bat and a ball cost [$1.10] in total. The bat costs [10 cents] more than the ball. How much does the ball cost?", p. 12), for example, is 50 cents, but the intuitive answer is 10 cents. Participants gave more correct responses in the high incentives condition, implying that they decided against the intuitive response. This reduced reliance on intuition could also influence the Moses illusion and the truth effect, as both are likely due to heuristic processing, which has been equated to intuition (e.g., Gilovich et al., 2002). More precisely, for the Moses illusion, relying on partial matching to judge a question as "close enough" is similar to intuitive decision making and for the truth effect, intuition is related to fluency (Topolinski & Strack, 2009).

Of course, monetary incentives are not the only way to increase motivation. Decisions involving grave real-life consequences (e.g., related to safety or health) are likely to elicit high motivation to make the correct choice (cf. McCaul et al., 2006). For example, choosing the correct medication to treat an illness is a decision most people will be highly motivated to make correctly.

In the following, I will investigate whether either monetary incentives or potential real-life consequences influence the links between incentives, effort, and performance and determine if they could serve as functional tools to attenuate the Moses illusion and truth effect.

2 The Moses Illusion With Monetary Incentives

As delineated before, people fall for the Moses illusion due to partial matching, that is, they forego complete matching every word in a sentence with to the corresponding memory structure and instead rely on a heuristic "close enough" approach that is influenced by the connectedness of words in a semantic network. In the following paper, we investigated whether incentivizing participants' responses with money reduced the amount of Moses responses they gave. Chapter 2.1 is based on the following publication:

Speckmann, F., & Unkelbach, C. (2021). Moses, money, and multiple-choice: The Moses illusion in a multiple-choice format with high incentives. *Memory & Cognition*, *49*(4), 843–862. https://doi.org/10.3758/s13421-020-01128-z

Please note that some changes in citation style and formatting were undertaken to keep the layout of this dissertation consistent. No changes were made to the content of the article.

2.1 Moses, Money, and Multiple-Choice: The Moses Illusion in a Multiple-Choice Format With High Incentives

Abstract

When people answer the question "How many animals of each kind did Moses take on the Ark?", they often respond with "two", although Moses does not appear in the biblical story of the Ark. We investigated this "Moses illusion" in a multiple-choice format and tested the influence of monetary incentives on the illusion's strength. Thereby, we addressed the role of a cooperative communication context for the illusion's emergence, as well as the role of participants' motivation. In four experiments (total N = 914), we found that the Moses illusion persists in a multiple-choice format. As the multiple-choice format realizes a cooperative context in which the correct answer is always available, we exclude a cooperative context explanation for the illusion. Monetary incentives reduced the strength of the illusion. However, the reduction was numerically and statistically small. We thereby show that the illusion is not due to violations of cooperative communications, and not due to a lack of motivation. The multiple-choice approach will facilitate further research on the Moses illusion and the data provide additional evidence for the Moses illusion's empirical robustness and constrain its theoretical explanations.

Keywords: Moses illusion, cognitive illusions, incentivized responding, response biases

Moses, Money, and Multiple-Choice:

The Moses illusion in a multiple-choice format with high incentives

"How many animals of each kind did Moses take on the Ark?" Most people spontaneously respond "two". This erroneous response often occurs even when people know that according to the Bible, it was Noah and not Moses who took two kinds of each animal onto the Ark. Hence, the question cannot be answered, but people readily answer it nonetheless. Since its seminal examination by Erickson and Mattson (1981), this "Moses" illusion has become a robust classic.

Here, we investigate the illusion in a multiple-choice format in which the correct response ("can't be answered") is available and we make responses relevant because people may win or lose money for correct and incorrect responses. Thereby, we answer two research questions regarding the Moses illusion. First, people may believe that questions should be answerable in principle, following a norm of cooperative communication. As Grice (1975) delineated, people expect communication to be cooperative; that is, respondents expect that they should be able to answer a question. However, if the illusion persists in a multiple-choice format, then it is unlikely due to norms of cooperative communication. People might respond "two" because the response fits to the setup of animals and the ark in a cooperative communication setting. Yet, providing the correct "can't be answered" response as an option in the response set creates a cooperative communication setting; that is, the solution is present and available. Second, respondents may simply not care for the correct response and provide a response that at least partially fits with the question. However, if the illusion persists with incentivized responses, then it is unlikely due to participants' superficiality and lack of motivation. Conversely, participants might not show the illusion if responses have real monetary consequences. Rather, they might retrieve the fact that Noah, not Moses, is the Biblical character who built the Ark.

In the remainder, we provide a short overview of research on the Moses illusion. Based on this overview, we delineate our two research questions in more detail. Finally, we present four experimental studies that address these questions empirically.

Previous research on the Moses illusion.

To keep the terminology consistent across authors and experiments, we refer to "Moses" questions ("How many animals did *Moses*…") as "distorted" questions and to the question from which they are derived as "undistorted" questions ("How many animals did *Noah*…"). We refer to erroneous responses to Moses questions (e.g., "two" to the Moses question) as "Moses" answers. We refer to the term that is changed between distorted and undistorted questions as the "critical" term (e.g. "Moses" vs. "Noah"). We refer to the presence of the illusion when participants provide significantly more Moses responses as can be expected by chance.

The Moses illusion. Erickson and Mattson (1981) first demonstrated what is now well-known as the Moses illusion. They provided participants with a set of questions and informed them that some questions might be distorted and unanswerable. Across four "Moses" questions, participants provided 52.3% Moses answers (i.e., about 2 out of 4 on average), despite possessing the relevant knowledge. In Study 2, the authors shifted the focus of the question to the question's critical term by moving it to the sentence's beginning (i.e., "Moses took two animals of each kind on the Ark") and used a "true" or "false" judgment format. This reduced the illusion but did not eliminate it; participants still provided 26.5% Moses answers, despite possessing the relevant knowledge. In Study 3, the question's critical term varied (e.g., "Moses", "Adam", "Abraham", or "Nixon", for the Moses question). The critical terms differed in their phonological and semantic similarity to the undistorted terms, and higher semantic relation resulted more participants falling for the illusion. Names without semantic relations (e.g., "Nixon took two animals…") eliminated the illusion (see also van Oostendorp and Kok, 1990). For all remaining cases, participants still provided 52.3% Moses answers, despite possessing the relevant knowledge.

Further research substantiated these findings; the illusion partially depends on the direction of focus (Bredart & Modolo, 1988), higher semantic relatedness between the names used in the questions (e.g., Noah and Moses vs. Noah and Adam) increases the illusion (Van Oostendorp & De Mul, 1990), and phonetic relatedness (e.g., identical number of syllables, identical first vowel) increases the illusion (Shafto & MacKay, 2000). In addition, lower processing fluency decreases the illusion (Song & Schwarz, 2008), expertise decreases the illusion (Cantor & Marsh, 2017), and olfactory cues metaphorically related to suspicion decrease the illusion (Lee et al., 2015). In addition, there seem to be interindividual differences in access to long-term memory knowledge as well as short-term memory capacity that influence the illusion (Hannon & Daneman, 2001).

Next, we address two explanations that fit with the presented evidence in more detail and which are most relevant for the present research question.

The cooperative communication explanation. Reder and Kusbit (1991) investigated cooperative response behavior (Exp. 1) as an explanation. They argued that participants might notice the distortion but choose to ignore it to cooperate with the experimenter. To test this explanation, they used a "literal" and a "gist" condition. The literal condition instructed participants to answer questions literally, requiring the answer "can't say" to

Moses questions. The gist condition instructed participants to ignore minor inconsistencies, requiring the answer “two” to the Moses question.

Accordingly, if participants register the distortion and choose to ignore it, the literal condition should be easier, as the gist condition requires the step to ignore the distortion. However, in terms of participants’ errors and response latencies, the gist condition appeared to be easier. They thus concluded that the illusion does not appear due to participants cooperative behavior.

The partial-matching hypothesis. The partial-matching assumes that people only match certain parts of sentences to relevant memory structures when answering the questions (Reder & Cleeremans, 1990). In their study, Reder and Cleeremans also used a literal and a gist condition (see above). They argued that the gist condition is more suited for a partial matching strategy whereas the literal condition is more suited for a complete matching strategy. Based on participants’ superior performance in the gist condition, the authors concluded that the illusion follows from partial-matching. This argument is supported by the finding that the more terms within the distorted question match with the undistorted question, the stronger the illusion becomes. They further argued that the default processing mode aims for reduction effort (p. 248), and therefore, a partial match suffices to trigger a Moses response (see also Kamas et al., 1996; Park & Reder, 2004).

The present research

We investigated two aspects of the Moses illusion that follow from the two presented explanations. First, we wanted to examine again the cooperative communication explanation. As discussed, Song and Schwarz (2008), as well as Lee and colleagues (2015), showed that under circumstances that prompt a critical mindset (i.e., low processing fluency) and calling the cooperative context into question (i.e., “fishy” smells), the illusion is reduced. In fact, one might consider that the cooperative communication mind-set feeds into the partial-matching hypothesis. In a cooperative setting, a partial match might suffice, while in a competitive setting people should consider all the presented information. In addition, the conclusion by Reder and Kusbit (1991; Exp. 1) rests on the assumption that people fully remember and comply with the instructions. In other words, the better performance in the gist condition follows because it matches participants’ standard low effort mode, which presumes a cooperative context. However, as stated, manipulating the context to be less cooperative reduces the illusion.

To address the cooperative context explanation more directly, we investigated the Moses illusion in a multiple-choice format (Exp. 1 and 2). If the illusion persists if the correct answer for distorted questions is present on each trial using the multiple-choice format, this would exclude a cooperative communication explanation of the illusion. As the correct response for distorted questions is present on each trial, a cooperative context is directly realized. If the typical Moses illusion is due to a violation of the maxim of cooperation (Grice, 1975), the effect should disappear when participants are repeatedly reminded through the response option “can’t say” that the question itself might be wrong. In addition, choosing one of four answer should be similar effortful for each answer. The exception would be a response option that asks for combinations of other options (e.g., “none of the above”), which is not recommended for multiple-choice tests (see Butler, 2018) and not implemented here.

Beyond the theoretical implications, the multiple-choice format has the practical advantages that it facilitates data collection. In previous experiments, participants wrote down their response or responded verbally. Both cases required additional coding. The multiple-choice format eliminates this step and allows for standardized coding via computer-scripts.

Second, building on Experiments 1 and 2, we investigate the role of motivation by monetary incentives (Exp. 3 and 4). As delineated, in particular the partial matching hypothesis (Kamas et al., 1996; Reder & Cleeremans, 1990) relies on the default of low-effort processing and assumes that participants do not thoroughly process the presented questions. This is in line with the classic “cognitive miser” hypothesis (Kurzban et al., 2013; Zipf, 1949). However, if answers have monetary consequences, participants should be motivated to read and process the questions more carefully (e.g., Terborg & Miller, 1978), thereby reducing the illusion. Monetary incentives are usually a good tool to increase motivation in quantitative tasks (e.g., Cerasoli et al., 2014). In other words, if the Moses illusion depends on a lack of motivation, then incentivizing responses should reduce the illusion. Conversely, if incentives do not influence the illusion, one may conclude that people are unable rather than unmotivated to avoid the illusion.

In the following, we report two experiments that investigate the Moses illusion in a multiple-choice format, and building on these, two further experiments that incentivize participants’ responses in the multiple-choice format. Experiment 1 presents an initial test if the Moses illusion persists with a multiple-choice format. Experiment 2 replicates Experiment 1 and varies procedural aspects of the multiple-choice format. The multiple-choice format realizes a cooperative context, as the correct answer is available on every trial. Experiment 3 and 4 then use this multiple-choice format to investigate the influence of motivation by monetary incentives on the Moses illusion. By incentivizing responses, we aim to increase participants' motivation to process the questions thoroughly. If the illusion persists nevertheless, we will have evidence that it does not depend on people's shallow processing of the questions, in particular when the multiple-choice format makes the correct answers readily available.

We report how we determined our sample sizes, all data exclusions (if any), all manipulations, all measures, and all studies we did in this research line so far. In addition, we pre-registered Experiments 2, 3, and 4; these preregistrations, as well as data and materials for all four studies can be found at: https://osf.io/8dzkt/?view_only = 24b06bb6dd364d66913521dad3c58836.

Experiment 1: Moses and Multiple-Choice I

Experiment 1 investigated if the Moses illusion persists in a multiple-choice format across a series of distorted ("Moses") and undistorted ("Noah") questions. If the illusion follows from violations of cooperative communication, the multiple-choice format should substantially reduce the illusion, as the correct response ("can't say") is available on every trial.

Method

Materials. We used 40 questions in 2 versions each: a distorted version (i.e., "How many animals of each kind did Moses take on the ark?") and an undistorted version (i.e., "How many animals of each kind did Noah take on the ark?"). We used the 26 questions and their respective answers from Park and Reder (2004) and replaced 14 questions because of apparent cultural knowledge differences between the US and Germany. We thus generated 14 new questions and answers for a final set of 40 questions. Appendix A provides the list of distorted and undistorted questions. We tested people's knowledge of these 40 questions' critical term in the student population that serves as the main source of participants for our lab with open-ended questions (i.e., “Which biblical figure took two animals of each kind on the Ark?”). Six coders rated the responses from 120 participants; Three coders rated the first half of the questions and three different coders rated the second half. Participants answered the open-ended questions correctly 69.7% of the time.[1]

Participants and design. We had no assumptions about the relevant effect size. Based on the recommendations by Ledgerwood (2015), we gathered data of 100 students who participated in exchange for a bar of chocolate and were recruited on campus to participate in a lab-based study. Previous studies typically used between 20 and 40 participants per condition. Two participants did not complete the study, leading to a final sample of 98 (M_{age} = 23.10 years, *SD* = 6.28; 43 female, 55 male).

We manipulated within-participants question type (“distorted” vs. “undistorted”); half of the questions appeared as distorted questions and half appeared as undistorted questions. Each question only appeared as either distorted or undistorted. Each question had four different response options. The first response option (e.g., “two”) could be correct or incorrect, depending on the question type (i.e., “How many animals of each kind did Noah take on the Ark?” vs. “How many animals of each kind did Moses take on the Ark?”). The second response option was always incorrect (e.g., “three” to the question “How many animals of each kind did Noah take on the Ark?”). The third response option was “can’t say”, which could be correct or incorrect, depending on the question type (e.g., incorrect in response to the question “How many animals of each kind did Noah take on the Ark?”, but correct in response to the question “How many animals of each kind did Moses take on the Ark?”). The fourth response option was “don’t know”, which counted as neither correct nor incorrect and effectively allowed participants to skip a question for lack of relevant knowledge. The order of responses options 3 and 4 was swapped for Experiment 1; for consistency, we report all results using the order that Experiments 2-4 employed.

Procedure. The data was collected together with another study that investigated the influence of females wearing a hijab on participants' responses. After finishing this unrelated study, participants continued with the present experiment within a Qualtrics survey. The survey informed participants that they had to answer 40 multiple-choice trivia questions and that they could skip a question by choosing the "don't know" response. It also explained that unanswerable questions could appear in which case the correct response would be "can't say". Specifically, the instructions read "It is possible that some questions appear that cannot be answered. In that case, please select the response option can't say. If you don't know the answer to a question, please select the response option I don't know". The questions appeared individually in a new randomized order for each participant. Participants had to choose one the of the responses to proceed. After answering the 40 questions, participants answered two more questions pertaining to the previous task and indicate if they had suspicions about the aim of the study. Afterwards, the experimenter thanked and debriefed participants.

Results

Table 2.1 presents participants' mean response frequencies to distorted questions (i.e., "How many animals of each kind did Moses take on the ark?") and to undistorted questions (i.e., "How many animals of each kind did Noah take on the ark?"). As Table 2.1 shows, participants provided erroneous "Moses" responses almost half the time (M = 9.74 out of 20 questions). In addition, participants provided substantially more "can't say" responses for distorted (*M* = 7.88) compared to undistorted (*M* = 2.12) questions, suggesting that participants understood the task.

To provide an inferential statistical test, we coded Moses responses as 1 and all other response types as 0 before adding up all values to compute the total number of Moses responses for each participant. We then compared the number of Moses responses (responses to distorted questions as if the question were undistorted, i.e., “Moses”) to the number expected at chance level of 1/3 (6.67). We chose 1/3 instead of 1/4 because the fourth response option is a “skip” rather than an actual answer and 1/3 is a more conservative comparison. We used a Welch-test to compare the average amount of Moses responses to this number and found a significant difference, $M = 9.73$, 95%CI [8.88, 10.59], $t(97) = 7.10$, $p < .001$, $d = 0.71$.

To check whether the multiple-choice format reduced the illusion, we compared our rate of Moses responses (49%) to the rates reported in Reder and Kusbit (1991). For our comparison, we chose the literal condition, which is identical to our setup. Their Moses response rates were all lower (33% in Experiment 1, 35% in Experiments 2 and 3, 32% in Experiment 4, all taken from the literal task condition), which suggests that our multiple-choice format did not diminish, but even fostered the illusion. Of course, the present data and the data by Reder and Kusbit differ on many aspects, such as the time of collections, the participant sample, and the stimuli; thus, the numerical difference cannot be attributed to the multiple-choice format, but it is important to note that the multiple-choice format did not produce completely different results.

Table 2.1

Experiment 1's mean frequencies of the four different response options across 40 questions as a function of question type (distorted vs. undistorted). Standard deviations in parentheses.

	Response			
Question type	1	2	3	4
Undistorted	15.17 (3.22)	1.23 (0.97)	2.12 (2.98)	1.48 (1.70)
Distorted	9.74 (4.27)	0.62 (1.11)	7.88 (4.74)	1.77 (1.73)

Note. Response 1 represents the "Moses" response for distorted questions and the correct response for undistorted questions. Response 2 represents the false alternative. Response 3 represents the "can't say" option (correct for distorted questions), and response 4 represents the "don't know" option (i.e., lack of knowledge for the topic).

Discussion

Experiment 1 showed a Moses illusion with a multiple-choice format. The constant presence of the "can't say" option should have reminded participants that some of the questions are distorted questions and the effort involved should be similar for all response options. Participants nonetheless selected the Moses response in almost 50% of the distorted cases. Thus, the Moses illusion also appears in a multiple-choice format, which provides strong support that the illusion is not due to participants misunderstanding of a cooperative communication setting.

There are three limitations with regards to conclusions on the Moses illusion in a multiple-choice format, though. First, participants might have learned that the first response option is the relevant one and therefore might have preferentially selected this response option. In other words, because participants always saw the four response options in the same constant order, and the first response is correct in 50% of the cases, they might have inferred that the first response is always the correct response. Response behavior based on such an inference might mimic a Moses illusion in a multiple-choice format. The underlying reason for choosing the “Moses” responses would thus not be a Moses illusion proper, but a learned response bias from the undistorted questions.

Second, we did not check if participants have, in principle, the relevant knowledge. We included the “don’t know” response, but this might not be an accurate representation of what participants know because there are no consequences to guessing blindly.

Third, and finally, participants might have confused the “can’t say” with the “don’t know” options. Although “can’t say” is the response often requested in research on the Moses illusion for distorted items with open formats (e.g., Reder & Kusbit, 1991), it might not be ideally suited for a multiple-choice format.

Experiment 2 addresses these three concerns by comparing a fixed order of response options with a shuffled order, by checking for participants knowledge in the same experiment, and by adjusting the “can’t say” response option to specify its meaning more precisely.

Experiment 2: Moses and Shuffled Multiple-Choice

Experiment 1 showed that in principle, a Moses illusion is also apparent in a multiple-choice format. To address the three limitations discussed above, Experiment 2 included three changes. First, to address the problem that Experiment 1 might have introduced a response bias for the first option, Experiment 2 manipulated between participants whether the order of response options was constant or shuffled anew for each question. Second, to address whether the illusion in the multiple-choice format depends on lack of knowledge about the questions being presented, Experiment 2 asked participants direct, open-ended questions at the end about each question's critical term (e.g., "Which biblical figure took two animals of each kind on the Ark?"). Third, we changed the "can't say" option to "This question can't be answered in this form". Different from Experiment 1, we collected the data online using Amazon's Mechanical Turk platform.[2]

Method

Materials. We translated our newly created questions to English, but replaced questions culturally specific to Germany with additional questions from Park and Reder (31 questions from Park and Reder, 9 of our own questions translated to English). The whole set of questions for Experiment 2 is available in Appendix B.

Participants and design. We again manipulated question type ("distorted" vs. "undistorted") within participants. We manipulated response option order ("shuffled" vs. "fixed") between participants. We aimed for the same sample size as Experiment 1. However, as we manipulated the randomized presentation of the options and the fixed presentation between participants, we aimed to collect data from 200 participants in total (100 per cell; see preregistration). We collected data from 205 Amazon Mechanical Turk workers who participated for $2.85. As preregistered, we excluded three participants because they indicated low concentration during the study, and two apparent bots, leaving

200 participants in the final sample for analysis (M_{age} = 40.0 years, *SD* = 11.3; 82 female, 115 male, 3 prefer not to say).

Procedure. Participants were redirected to the Qualtrics survey from the online platform Amazon Mechanical Turk. The survey randomly assigned participants to one of two conditions. The fixed condition replicated Experiment 1. In the shuffled condition, response options order was randomized anew for each question. Question order was also randomized in both conditions.

In both conditions, participants first read and agreed to an informed consent form and the instructions informed them about the procedure of the experiment and asked them to not look up any answers online. Different from Experiment 1, both conditions used the response option "This question can't be answered in this form" instead of the option "can't say". The instructions stressed the difference between "This question can't be answered in this form" and "Don't know" response options and in what cases they should be used. Specifically, the instructions read "Some of these questions are impossible to answer. In that case, the correct response option is This question can't be answered in this form. If you don't know the answer to a question, please select the response option Don't know".

Upon completing the 40 multiple-choice questions, participants responded to 40 corresponding open-ended format questions (e.g., for the typical Moses question, participants responded to "Which biblical figure took two animals of each kind on the Ark?"), checking if participants have the relevant knowledge to answer the multiple-choice questions correctly. Finally, participants provided demographic information and indicated how concentrated they were during the study on a scale from one to six.

Results

We excluded 192 responses because their respective response times were more than three SDs above the mean for that specific participant (i.e., 2.4% of all responses)[3]. Please note that this also served as our exclusion criterion for potential cheating (i.e., searching for correct answers online), as searching for an answer online should increase response latencies relatively within participants, and if a given participant searches for all answers, relatively to other participants (see also Footnote 3). Table 2.2 presents the mean frequencies of different response types to the multiple-choice questions.

Table 2.2

Experiment 2's mean frequencies of the four different response options across all 40 questions as a function of question type (distorted vs. undistorted) and response order (constant vs. shuffled). Standard deviations in parentheses.

		Response			
Response order	Question type	1	2	3	4
Fixed	Undistorted	17.22 (3.07)	0.73 (2.14)	0.52 (0.87)	1.14 (1.90)
	Distorted	10.89 (4.66)	0.74 (1.90)	6.12 (5.08)	1.63 (2.40)
Shuffled	Undistorted	17.09 (3.02)	0.53 (1.40)	0.40 (0.75)	1.47 (2.02)
	Distorted	10.15 (4.59)	0.72 (1.60)	6.80 (5.78)	1.94 (2.43)

Note. Response 1 represents the "Moses" response (or correct response for undistorted questions), response 2 the false alternative, response 3 represents the "can't say" option (correct for distorted questions), and response 4 represents the "don't know" option (i.e., lack of knowledge for the topic. Individual rows do not add up to 20 because of excluded responses. Please not that in the shuffled condition, response number does not indicate question order.

Two important points are visible from Table 2.2's descriptive statistics. First, overall, we replicated the Moses illusion with an online American sample using the multiple-choice format; participants provided erroneous Moses responses more than half the time (52.6%). As in Experiment 1, we compared the average amount of Moses responses to the amount based on the chance level of 1/3 (6.67). A Welch-test showed a significant difference, M = 10.52, 95% CI [9.87, 11.16], $t(199) = 11.75$, $p < .001$, $d = 0.83$ between the frequency of Moses responses and chance level.

Second, there is only a small difference between the fixed and the shuffled conditions. The preregistered Welch-test between the constant (M = 10.89) and shuffled (M = 10.15) conditions was not significant, $t(197.74) = 1.13$, $p = .259$. To provide statistical evidence beyond the null results that the illusion strength is equivalent between conditions, we followed up the Welch-test with an equivalence test (Lakens et al., 2018). Before data collection, we used the TOSTER package in R (Lakens, 2017) to run a power analysis based on our sample size of 100 per cell and an alpha level of $\alpha = .05$ and 90% power. This resulted in lower and upper equivalence bounds of $\Delta_L = -.47$ and $\Delta_U = .47$, so we preregistered an interval containing an effect of $|d| < .47$ to be equivalent. The equivalence test was significant, 90% CI [-.034, 1.82], $t(197.74) = -2.19$, $p = .015$. Based on the equivalence test and the null-hypothesis test combined, we can conclude that the observed effect is statistically not different from zero and statistically equivalent to zero.

In addition, participants provided on average substantially more “This question can’t be answered in this form” responses for distorted (M = 0.46) compared to undistorted (M = 6.46) questions, suggesting again that participants understood the task.

Moses illusion as a function of knowledge. Six independent coders (three for each half of the questions) coded the open-ended answers for correctness (0 = incorrect, 1 = correct). Their interrater reliability was very high (Fleiss’ kappa = .95). If coders disagreed on a response’s correctness, we used the value that the majority of the coders agreed upon.

Similar to our student sample, participants knew the critical term for a given question in 73.2% of cases.

To test if knowledge influences illusion strength, we excluded all responses for which the coding indicated lack of knowledge of the topic. First, to test if the Moses illusion persists in this dataset with exclusions based on knowledge of the topic, we again ran a Welch test comparing the percentage of Moses responses against the chance level of 1/3. We used ratios instead of frequencies in this case to account for the different number of exclusions per participant. We found a significant difference, M = 0.52, 95% CI [0.48, 0.56], t(199) = 9.57, p < .001, d = 0.68. Second, to test if the exclusions influenced the pattern between conditions, we repeated the Welch test and equivalence test between conditions in the dataset with knowledge exclusions. Again, the frequency of Moses responses did not differ between shuffled (M = 6.0) and fixed (M = 6.7) conditions, t(191.64) = 1.28, p = .203, and the equivalence test was significant, 90% CI [-0.19, .1.49], t(191.64) = -2.04, p = 0.02. Third, we directly compared illusion strength in terms of ratios in the dataset without knowledge exclusions (M = 0.54) with illusion strength in the dataset with knowledge exclusions (M = 0.52). A paired Welch test showed that the 0.02 percent reduction was significant, t(199) = 2.02, p = .044, d = .14. However, importantly, even in the dataset with exclusions, participants still gave Moses responses more than half of the time.

Discussion

Experiment 2 successfully replicated the Moses illusion using a multiple-choice design. Experiment 2 thereby addresses three concerns about Experiment 1. First, the non-significant difference between the fixed and shuffled conditions with considerable power, together with the significant equivalence tests make an explanation of the illusion in terms of a response bias for the first response unlikely. In addition, also the change from a laboratory to an online setting had no major effects.

Second, the illusion was also strongly present for a strict version of the illusion, in which we only considered responses for which participants provided the correct response in corresponding open-ended questions. Excluding responses for which participants could not recall the critical term led to a significant reduction in illusion strength, but this reduction was very small (i.e., from 0.54 to 0.52, in terms of ratios). Thus, we are confident that we are addressing a true illusion, and not another form of guessing bias. This is also relevant for Experiment 1, where we found a similar percentage of knowledge for the critical terms. It is also important to keep in mind that the open-ended knowledge check for the critical term is the most conservative test, as participants might recognize that it was not Noah who took two animals on the Ark, but they might not be able to produce the correct term “Moses” in a free recall format.

Third, the more explicit labeling of the correct response to distorted question as “This question can’t be answered in this form” did also not produce strong changes. The illusion strength was highly similar to Experiment 1, given the change in settings.

Together, Experiment 1 and 2’s results make an explanation of the Moses illusion in terms of Grice’s (1975) maxim of cooperation unlikely. The multiple-choice format presents the correct responses for both distorted and undistorted questions at each trial, fully realizing a cooperative communication setting, that is, the questions have the correct answers available on each trial. The present data thereby substantiates and extends the conclusions by Reder and Kusbit (1991). The experiments also show that a multiple-choice version seems suitable to capture the Moses illusion, which represents a strong practical facilitation of research on this interesting illusion.

Having the feasibility of the multiple-choice format established, we used this format to investigate the role of motivation by monetary incentives on the illusion’s strength.

Experiment 3: Moses and Money

Experiment 3 made responses for participants relevant by providing monetary incentives for each response. If the illusion follows from participants superficial processing of the questions, then incentives should decrease the strength of the illusion. Experiment 3 used Experiment 1’s fixed multiple-choice format. In addition to the within variation of question type (i.e., distorted vs. undistorted), we implemented three between-participants incentive conditions. A “no incentives” condition replicated Experiment 1 besides the differential compensation (i.e., Experiment 3 offered payment, while Experiment 1 offered a chocolate bar). A “low incentives” condition awarded 15 cents per correct response and subtracted 15 cents per incorrect response. Given the 40 questions, participants could thus earn up to 6 Euro in the “low incentives” condition. A “high incentives” condition awarded 30 cents per correct response and subtracted 30 cents per incorrect response. Given the 40 questions, participants could thus earn up to 12 Euro in the “high incentives” condition. In comparison to Experiment 1, the penalty for guessing should deter participants from responding anything other than “don’t know” if they do not have the relevant knowledge or if they feel unsure, which makes this response option an approximation of a knowledge check.

If the Moses illusion is due to participants’ lack of motivation and the following superficial processing of the questions, we would expect a main effect of condition on erroneous “Moses” responses: Motivation and subsequently, attention and depth of question processing, should increase with incentives. Thus, we would expect a linear trend, with high incentives leading to more correct responses than low incentives and low incentives leading to more correct responses than no incentives. Based on this reasoning, we pre-registered a linear trend from the no incentives to the high incentives condition. We did not specify a significant difference between the high incentives and the low incentives conditions. However, there should be no quadratic trend (i.e., less correct responses in the high incentives condition compared to the low incentives condition).

Method

Materials. We used the same questions and responses as in Experiment 1. We adjusted one question due to an ambiguity in the question. Originally, we asked “What is the name of the prize awarded in Sweden for significant contributions in the fields of science *and peace*?” (undistorted). We changed this to “What is the name of the prize awarded in Sweden for significant contributions in the field of science?” (undistorted) because the Nobel Peace Prize is awarded in Oslo, Norway[4].

Participants and design. Based on the sample size of Experiment 1 (n = 100), we pre-registered a sequential analysis based on Lakens (2014) to reduce the cost of the experiment. We pre-registered to gather data of 150 participants (50 per condition) and then stop collecting data if we find the predicted linear trend from the no incentives condition to the high incentives condition. If this was not the case, we planned to collect data the full 100 participants per condition with an adjusted *p*-value of $p < .0294$ (see Lakens, 2014, p. 703). At 150 participants, the pre-registered analyses showed no effect and so we continued data collection with the goal of 300 participants. Ultimately, 318 students participated for base payment of 4 Euro plus the incentives in the incentivized conditions. We excluded three participants who indicated concentration of less than 3 on a scale of 1-6 during the experiment, leading to the final sample of 315 participants (M_{age} = 23.17 years, *SD* = 5.87; 187 female, 123 male, 2 other, 3 prefer not to say). We recruited all participants on a university campus for participation in a lab-based study.

The computer program randomly assigned participants to one of three incentives conditions. The total amount could not go below zero and participants received their final score in cents in addition to a flat payment of 4€. Again, participants could earn up to 12 € in the high incentives condition and up to 6 € in the low incentives condition in addition to their flat payment. Participants in the no incentives condition did not gain or lose money during the experiment. All participants expected to receive 4€ during recruitment, before we randomly assigned them to conditions. In addition to collecting responses to the questions, the program also collected response times for each question from showing the question and the participant clicking the “next” button. The question and response presentation were identical to Experiment 1.

Procedure. Experimenters welcomed participants in the lab and seated them in a cubicle in front of a computer and launched a Python program that led participants through the rest of the experiment. Participants read and agreed to an informed consent form and the program informed them about the procedure of the experiment. It asked them to turn off their smartphones to deter cheating and explained the incentive system. The response options were similar to Experiment 1. The instructions specifically mentioned the “can’t say” and “don’t know” response options and explained that the former was the correct response to non-answerable questions while the latter did not affect the point total and could be used to skip a question. In comparison to Experiment 1, this should help avoid confusion in regards to meaning of the response options. After the 40 questions, participants answered demographic questions and how concentrated they were during the study before the experimenter thanked, debriefed, and paid them.

Results

Table 2.3 shows the percentage of different response types to distorted questions and undistorted questions as a function of no, low, and high incentives. Overall, the results replicated Experiment 1. As the table shows, across conditions, participants showed a substantial Moses illusion. They provided on average Moses responses for 8.85 out of 20 questions.

Table 2.3

Experiment 3's mean frequencies of the four different response options across all 40 questions as a function of question type (distorted vs. undistorted) and incentives (high vs. low vs. none). Standard deviations in parentheses.

Incentives	Question type	Response			
		1	2	3	4
High	Undistorted	15.64 (3.16)	0.16 (0.40)	2.22 (2.11)	1.97 (2.24)
	Distorted	8.05 (4.80)	0.78 (1.00)	8.93 (5.31)	2.24 (2.12)
Low	Undistorted	15.85 (2.85)	0.38 (0.75)	2.19 (1.98)	1.58 (1.84)
	Distorted	8.93 (4.43)	0.58 (0.82)	8.61 (4.95)	1.88 (1.95)
None	Undistorted	15.65 (3.05)	0.49 (1.02)	2.15 (2.47)	1.7 (1.66)
	Distorted	9.58 (4.30)	0.61 (1.19)	7.95 (4.98)	1.87 (1.96)

Note. Response 1 represents the "Moses" response for distorted questions and the correct response for undistorted questions. Response 2 represents the false alternative. Response 3 represents the "can't say" option (correct for distorted questions), and response 4 represents the "don't know" option (i.e., lack of knowledge for the topic).

Confirmatory analyses of incentive effects. We computed illusion strength identically to Experiment 1. We then checked whether the basic Moses effect persisted by comparing the mean number of Moses responses to the number based on the chance level of 1/3 (6.67). A Welch-test showed a significant difference, M = 8.85, t(314) = 8.53, p < .001, d = 0.48, between the frequency of Moses responses and chance level.

To analyze the mean number of Moses responses as a function of incentives, we submitted these data to a one-way ANOVA with condition (incentives: high vs. low vs. no) as the sole between-participants factor. There was no significant main effect for condition, F(2,312) = 3.01, p = .051, η_p^2 = .019. However, as pre-registered, the linear contrast from high to no incentives was significant, t(312) = 2.44, p = .015, d = 0.28. The quadratic trend was not significant, t(312) = -0.23, p = .820, d = -0.03.

To analyze the mean number of correct responses, we coded correct responses as 1 and all other response types as 0 before adding up all values to compute the mean number of correct responses. We pre-registered a main effect of condition and a linear trend from the no incentives condition to the high incentives condition with participants in the high incentives condition giving the most correct responses. The data was submitted to a one-way ANOVA with incentives (high vs. low vs. none) as the between factor. Contrary to our pre-registered hypotheses, the main effect for condition on correct responses was not significant, F(2,312) = 1.10, p = .335, η_p^2 = .007; and neither the linear trend, t(312) = -1.36, p = .176, d = -0.15, nor the quadratic trend were significant, t(312) = -0.60, p = .551, d = -0.07.

Exploratory analyses. As an exploratory analysis, we also compared the average proportion of type 4 responses (skips) between conditions. From the distribution of response types, it seems that participants skipped questions more often in the high incentives condition compared to the other two. This would make sense, because the high incentive condition also has the highest losses for incorrect questions. However, the main effect for condition was not significant, F(2,312) = 1.45, p = .236, η_p^2 = .009, and neither the linear trend, $t(312) = -1.34$, $p = .180$, $d = -0.15$, nor the quadratic trend were significant, $t(312) = 1.05$, p = .295, $d = 0.12$.

The supplements also provide and exploratory analyses of the response times.

Discussion

We again replicated the basic Moses effect using multiple-choice questions. In addition, incentives did influence the frequency of Moses responses. We found the predicted linear trend from high to no incentives; participants in the high incentives condition provided less Moses responses compared to the no incentives condition. However, the effect was much weaker than anticipated. In fact, a potential additional payment of € 12 (around $14) reduced the illusion only by 16% and necessitated 300 participants to show it statistically (see sequential analysis).

Based on feedback from our lab meetings, one reason could be again the wording of the different response options. Specifically, as we used the response option from Experiment 1, participants may construe the phrase “Can’t say” as “I can’t answer this” which would be very close in meaning to “I don’t know”. Even though we made sure to explain the different response options in the instructions, this could nevertheless have influenced participants’ responses in the incentivized version, diminishing the potential incentive influence. We thus aimed to replicate the surprising result from Experiment 3 (i.e., the small incentive effect) with the changed response option from Experiment 2.

Experiment 4: Moses and Money replicated?

We aimed to replicate the basic Moses effect and the linear influence of incentives on the illusion’s strength.

Method

Materials, participants, design, and procedure. We changed the response option “Can’t say” to “This question can’t be answered in this form”, which is also the response option used in Experiment 2. Otherwise, the materials were identical to the ones used in Experiment 3. As for Experiment 3, we pre-registered a sample of 300 participants. 298 students participated for payment (M_{age} = 22.19 years, *SD* = 4.78; 146 female, 148 male, 4 other) and were again recruited on a university campus. The design and procedures were identical to Experiment 3.

Results

Table 2.4 shows the percentage of different response types to distorted questions and undistorted questions as a function of no, low, and high incentives. Across conditions, participants again showed a Moses illusion in a multiple-choice format with incentives. Out of 20 distorted questions, they on average provided Moses responses for 7.74 questions. This is an approximate average 5% drop (or one question) in comparison to Experiment 3, which could be a direct result of the change of phrasing for the “can’t say” option.

Confirmatory analyses of incentive effects. A Welch-test between average number of Moses responses per participant and the number expected from chance level (6.67) was again significant, *M* = 7.77, *t*(297) = 4.46, $p < .001$, *d* = 0.26.

We computed illusion strength as in Experiment 3. We submitted these data to a one-way ANOVA with incentive condition (high vs. low vs. no) as the between factor. There was no incentive condition main effect, *F*(2,295) = 1.07, *p* = .346, η_p^2 = .007, and different from Experiment 3, the linear contrast between high, low, and no incentives was also not significant, *t*(295) = 1.26, *p* = .208, *d* = 0.15; neither was the quadratic trend, *t*(295) = 0.73, *p*

= .467, *d* = 0.08. Thus, while the effect is numerically in the expected direction from the high (*M* = 7.49) to the no incentives (*M* = 8.24) condition, we did not replicate the influence of incentives on the strength of the Moses illusion.

Table 2.4

Experiment 4's mean frequencies of the four different response options across all 40 questions as a function of question type (distorted vs. undistorted) and incentives (high vs. low vs. none). Standard deviations in parentheses.

		Response			
Incentives	Question type	1	2	3	4
High	Undistorted	15.31 (2.87)	0.35 (1.1)	2.18 (1.92)	2.15 (2.11)
	Distorted	7.49 (4.25)	0.66 (1.05)	9.33 (5.02)	2.52 (2.34)
Low	Undistorted	15.19 (2.96)	0.33 (0.73)	2.21 (2)	2.26 (2.51)
	Distorted	7.49 (4.17)	0.46 (0.76)	9.33 (4.95)	2.71 (2.84)
None	Undistorted	15.76 (2.87)	0.41 (0.74)	2.29 (2.47)	1.54 (1.67)
	Distorted	8.24 (4.07)	0.54 (0.85)	9.58 (4.92)	1.64 (1.76)

Note. Response 1 represents the "Moses" response for distorted questions and the correct response for undistorted questions. Response 2 represents the false alternative. Response 3 represents the "can't say" option (correct for distorted questions), and response 4 represents the "don't know" option (i.e., lack of knowledge for the topic).

To analyze the mean number of correct responses, we coded correct responses as 1 and all other response types as 0 before adding up all values to compute the mean number of correct responses. We submitted this data to a one-way ANOVA with incentives (high vs. low vs. none) as the between factor. As in Experiment 3, contrary to our predictions, the incentives main effect was not significant, $F(2,295) = 0.62$, $p = .538$, $\eta_p^2 = .004$. The linear trend was also not significant, $t(295) = 0.88$, $p = .380$, $d = 0.10$, and the quadratic trend was also not significant, $t(295) = 0.68$, $p = .495$, $d = 0.08$.

Exploratory analyses. As an exploratory analysis identical to that of Experiment 3, we also compared the average proportion of type 4 answers (skips) between conditions. From the distribution of response types, it seems that participants skipped questions more often in the high incentives condition compared to the other two. This would make sense since the high incentive condition also has the highest losses for incorrect questions. This time, the main effect for condition was significant, $F(2,295) = 5.69$, $p = .004$, $\eta_p^2 = .037$, and the linear trend was also significant, $t(295) = 2.62$, $p = .009$, $d = -0.30$. The analysis of the response times is presented in the supplement.

Discussion

While we replicated the overall Moses effect, we did not replicate the influence of incentives on the illusion; different from Experiment 3, the linear trend between the three incentives conditions was not significant. Also different from Experiment 3, the main effect of incentives on skips was significant and the negative linear trend indicates that participants skipped more questions as incentives increased.

Overall, the average number of correct responses was unaffected by incentives, but the significance pattern regarding illusion strength as a function of incentives was inconsistent between Experiments 3 and 4. While the incentives linearly decreased the frequency of Moses responses in Experiment 3, we found this trend only numerically, but not statistically in Experiment 4. Because of the different result patterns, we analyze the data from Experiments 3 and 4 together. As the difference between significant (Experiment 3) and non-significant (Experiment 4) is by itself not necessarily significant, this analysis allows us to test this significance difference of the linear trend by an interaction via Experiment.

Combining Experiments 3 and 4

Table 2.5 shows the frequency of different response types to distorted questions and undistorted questions separated by conditions. We added experiment as a factor to the previous ANOVA for a 3 (incentives: high vs. low vs. no) x 2 (Experiment: 3 vs. 4) ANOVA with percentage of Moses response as dependent variable. This ANOVA showed a significant incentive main effect, $F(2, 607) = 3.54$, $p = .030$, $\eta_p^2 = .012$. This incentive main effect was due to the expected the linear trend from high to no incentives, $t(607) = 2.64$, $p = .009$, $d = 0.21$.The quadratic trend was not significant, $t(607) = 0.34$, $p = .736$, $d = 0.03$. Importantly, the linear trend did not interact with Experiment (3 vs. 4), $F(1, 607) = 0.82$, $p = .366$. However, there was an Experiment main effect, $F(1, 607) = 9.88$, $p = .002$, $\eta_p^2 = .016$. Participants in Experiment 3 provided more Moses responses ($M = 8.85$, $SD = 4.54$) than participants in Experiment 4 ($M = 7.77$, $SD = 4.16$).

We also analyzed the correct responses with a 3 (incentives: high vs. low vs. no) x 2 (Experiment: 3 vs. 4) ANOVA. The main effect for incentives was not significant, $F(2, 607) = 0.04$, $p = .961$, $\eta_p^2 = .00$, and neither were the main effect for experiment, $F(1, 607) = 2.07$, $p = .151$, $\eta_p^2 = .00$, nor the interaction, $F(2, 607) = 1.64$, $p = .195$, $\eta_p^2 = .00$.

We used the same ANOVA for skips ("don't know") as dependent variable. This analysis shows a significant effect of incentive condition, $F(2, 607) = 4.60$, p = .010, $\eta_p^2 = .01$, with a significant linear trend from the high to the no incentives condition, $t(607) = -2.88$, $p = .004$, $d = -0.23$. The quadratic trend was not significant, $t(607) = -0.96$, $p = .337$, $d = -0.08$, and the linear trend did not interact with Experiment, $F(1, 607) = 1.43$, $p = .232$.

Table 2.5

Experiments 3's and 4's combined mean frequencies of the four different response options across all 40 questions as a function of question type (distorted vs. undistorted) and incentives (high vs. low vs. none). Standard deviations in parentheses.

		Response number			
Incentives	Question type	1	2	3	4
High	Undistorted	15.48 (3.02)	0.26 (0.82)	2.2 (2.01)	2.06 (2.17)
	Distorted	7.78 (4.54)	0.72 (1.03)	9.13 (5.16)	2.37 (2.23)
Low	Undistorted	15.53 (2.92)	0.36 (0.74)	2.2 (1.99)	1.91 (2.21)
	Distorted	8.24 (4.35)	0.52 (0.79)	8.96 (4.95)	2.28 (2.45)
None	Undistorted	15.71 (2.95)	0.45 (0.89)	2.22 (2.46)	1.62 (1.66)
	Distorted	8.92 (4.23)	0.57 (1.04)	8.75 (5.01)	1.75 (1.86)

Note. Response 1 represents the "Moses" response for distorted questions and the correct response for undistorted questions. Response 2 represents the false alternative. Response 3 represents the "can't say" option (correct for distorted questions), and response 4 represents the "don't know" option (i.e., lack of knowledge for the topic).

Combined discussion

The pooled data from Experiment 3 and 4 confirms the pattern from Experiment 3: a significant main effect of incentives on the average proportion of Moses responses with a linear trend from the high incentives condition to the no incentives condition. Combined with the reversed linear trend of skips and the non-significant effect of condition on correct responses, this pattern provides some insights into the underlying mechanisms. While participants gave fewer Moses responses in the high incentives conditions, their overall number of correct responses did not change. This means that participants in the incentive conditions were more careful and chose the skip option more often. This is probably due to our incentivization system, which involved losing points for incorrect responses but not for skips. However, to be sure, the reduction was small and the reduction effect required high statistical power.

General Discussion

The present research had two goals: We wanted to establish a multiple-choice response format for the Moses illusion and thus rule out the possibility that the illusion is due to a cooperative communication setting, and investigate the effect of motivation on the illusion by monetary incentives for correct responses.

The most important result for the multiple-choice format is that it does not eliminate the illusion. If the illusion were due to participants behaving as cooperative communication partners, understanding the distorted questions correctly and then choosing to respond to it as if it were undistorted, then presenting “can’t say” as a response option should have reduced the illusion. The correct response was available on every trial, which should also remind participants that distorted questions exist and the correct response for those questions is “can’t say”. We found a substantial Moses illusion, ruling out the cooperative communication setting explanation (Grice, 1975) and validating the multiple-choice format.

In addition, the multiple-choice format facilitates research on the illusion, as coding participants' responses is no longer necessary, which greatly reduces the resources required for studies on the Moses illusion.

A small caveat with regards to the multiple-choice format is that there is research showing that participants tend to avoid "none of the above" (NOTA) option in multiple choice questions (Blendermann et al., 2020). If one considers the present "can't say" option as a NOTA variant, this avoidance tendency might contribute to the present illusion. However, given the illusion strength present in the data, it seems unlikely that this tendency is fully responsible for the present effects.

Our motivation manipulation by monetary incentives had the hypothesized effect on the Moses illusion. We expected to find that with enough motivation due to monetary incentives, participants would pay more attention and detect distortions more often. Participants in the high incentives condition should provide the least Moses responses and participants in the no incentives condition should provide the most Moses responses. However, the incentives effect on the average number of Moses responses was much smaller and less reliable than we expected. While the pooled data of Experiments 3 and 4 provides confidence in the statistical significance of the incentives effect, the practical significance is negligible. We paid participants about 4 € (about $4.50) on average for them to give one less Moses response.

When looking at the increased skips in the incentivized conditions for the combined data, one could argue that the improvement in correct answers (and thus, payment) is not due to increased sensitivity to the Moses illusion, but rather a general response bias to be more careful out of fear of losing money. This result appears similar to the bias shift observed by Kamas and colleagues (1996), but in their studies, participants' bias shifted towards "can't say" responses to any question regardless of distortion. This is an incorrect response for half of the questions, whereas the shift towards skip responses in the present research is not an incorrect response but rather a decision to avoid risk.

Taken together, the data from a total of 914 participants shows that there is a strong Moses illusion, even when motivation is high and communication context effects are accounted for. Motivating participants with monetary incentives had some effect, but it does not account for a large part of the variance.

Thereby, our results are best compatible with the partial matching hypothesis, with the qualification that the driving force underlying partial matching is not people's tendency to avoid effort. Our high incentives condition should have been enough motivation for a student sample to invest enough effort to detect the distortion. Rather, one must consider a decision threshold model, in which the partially matching information in the question seems to suffice to pass the threshold to elicit the wrong "Moses" responses (Reder & Cleeremans, 1990, p. 250). Thus, the Moses illusion may emerge not because people do not pay attention or because they aim to be cooperative communication partners, but because the cognitive system is sufficiently prompted by the question's content to respond "two" when asked how many animals Moses took on the Ark, even when the stakes are relatively high.

Conclusion

The Moses illusion is a robust phenomenon that we also observed in a multiple-choice format. This implies that the illusion does not follow from respondents' attempts to be cooperative communication partners. The multiple-choice context clearly communicates on every trial that questions might be wrong. The multiple-choice format also opens many new venues for research on this intriguing illusion. The motive to avoid effort seems to play a minor role in the emergence of the illusion, as monetary incentives had a significant, but numerically small effect. This in turn supports explanations of the Moses illusion that rely on cognitive rather than motivational features.

Footnotes

1. This test was conducted to address a reviewer's concerns about the degree to which the questions were common knowledge, which is why it was conducted after Experiment 1 and on a different sample. Experiment 2 addresses the role of knowledge directly.

2. Experiment 2 was factually the last of the four experiments in this research project. It was conducted to address a reviewer's concern about the fixed order in the response format. In addition, we needed to conduct it online, as the CoViD-19 Pandemic prevented laboratory-based data collection at this time and we could not use an incentivized online version, as participants might easily and readily look up the correct response online.

3. We preregistered 3 SDs above the *group* median, instead of the more typically used exclusion criterion of 3 SDs above the individual mean, which is what we implemented after considering that online participants would be unlikely to search for all answers. Both exclusion criteria produce the same results pattern and do not affect any statistical conclusions.

4. As this potential error is randomly distributed across participants, removing the distorted and undistorted versions of this question from Experiment 1 did not change the results.

Acknowledgements

The present research was supported by a grant from the Center for Social and Economic Behavior (C-SEB) of the University of Cologne, awarded to the second author.

Appendix A

Undistorted questions:

R1

Which type of cigarettes was German chancellor Helmut Schmidt known for?

Menthol cigarettes

Filterless cigarettes

R2

Which resource did the USA suspend troops to Iraq for?

Oil

Solar energy

R3

In which movie does Arnold Schwarzenegger travel back in time to save Sarah Connor?

Terminator 2

Rocky 2

R4

With which instrument did Louis Armstrong become famous?

Trumpet

Violin

R5

Which object does Julie Andrews use to fly at the beginning of the movie “Mary Poppins”?

Umbrella

Broom

R6

Gorbachev was the leader of which communist country?

USSR

USA

R7

Margaret Thatcher was the prime minister of which country?

United Kingdom

France

R8

What year did Germany lose World War II?

1945

1918

R9

Which kind of meat is in the Whopper from Burger King?

Beef

Chicken

R10

What color is Dogmatix’s fur, the dog of Asterix and Obelix?

Black and white

Gray and brown

R11

Which season do we associate with the start of football season, the beginning of school and the trees’ leaves turning brown?

Autumn

Winter

R12

Which statue, given to the USA by France, symbolizes freedom for arriving immigrants at New York Harbor?

Statue of Liberty

Christ the Redeemer

R13

Which part of his body did artist Van Gogh allegedly cut off?

Ear

Nose

R14

What musician won multiple Grammys for their Album „Thriller“?

Michael Jackson

Elton John

R15

What follows „To be or not to be” in Hamlet’s famous soliloquy?

„That is the question.“

„Who knows?“

R16

Who is the video game character and Italian plumber who is Nintendo’s mascot?

Mario

Sonic

R17

Which country is known for cuckoo clocks, chocolate and pocket knives?

Switzerland

Italy

R18

Which political position did Adolf Hitler gain under President Paul von Hindenburg?

Chancellor of the Reich

Mayor

R19

Who found the Glass Slipper lost by Cinderella?

The Prince

The Stepmother

R20

What is the name of the kimono-clad courtesans who entertain Japanese men?

Geisha

Samurai

R21

What is the name of Leonardo da Vinci's famous painting of a woman that is displayed in the Louvre in Paris?

Mona Lisa

The Scream

R22

What is the name of the device that tells time by measuring the incidence of sunlight on a dial?

Sundial

Oscillator

R23

Who is the cartoon character known for eating spinach to get stronger?

Popeye

Mickey Mouse

R24

What is the name of the comic about Charlie Brown and his dog Snoopy?

Peanuts

Cashews

R25

Who is the dictator of North Korea?

Kim Jong-Un

Fidel Castro

R26

What is the name of the molten rock that travels down mountains after an eruption?

Lava

Mud

R27

Who is the Roman god of war after whom a famous candy bar is named?

Mars

Snickers

R28

Who is the white-bearded man in a red suit who distributes Christmas presents out of his sleigh?

Santa Claus

Rumpelstiltskin

R29

What is the name of the Mexican dip made from avocados?

Guacamole

Salsa

R30

What is the protagonist's name in Goethe's "Faust, Part One"?

Dr. Heinrich Faust

Romeo

R31

What is the name of the prize awarded in Sweden for significant contributions in the field of science?

Nobel prize

Academy Award

R32

What is the name of the island located in the south of Italy close to the "boot's toe"?

Sicily

Island of Elba

R33

What is the name of the TV show about a young Viking boy who always rubs his nose when trying to figure something out?

Wickie and the strong men (English: Vicky the Viking)

Pedia and the smart men

R34

Who is the architect of the famous Eiffel Tower in Paris?

Gustav Eiffel

Oscar Niemeyer

R35

What was the name of the wall in East-Germany that was torn down in 1989?

Berlin Wall

Great Wall of China

R36

How long did Sleeping Beauty fall asleep for, after poking her finger on a spindle?

100 years

2 days

R37

What is the name of the New Year festival celebrated on the 31st of December?

New Year's Eve

Carnival

R38

How many times did the German football team win the World Cup?

Four times

Never

R39

How many doors does an Advent calendar have?

24

365

R40

How many animals of each kind did Noah take on the Ark?

Two

Three

Distorted questions:

F1

Which type of cigarettes was German chancellor Helmut Kohl known for?

Menthol cigarettes

Filterless cigarettes

F2

Which resource did the USA suspend troops to Iran for?

Oil

Solar energy

F3

In which movie does Sylvester Stallone travel back in time to save Sarah Connor?

Terminator 2

Rocky 2

F4

With which instrument did Lance Armstrong become famous?

Trumpet

Violin

F5

Which object does Audrey Hepburn use to fly at the beginning of the movie “Mary Poppins”?

Umbrella

Broom

F6

Gorbachev was the leader of which capitalist country?

USSR

USA

F7

Margaret Thatcher was the president of which country?

United Kingdom

France

F8

What year did Germany win World War II?

1945

1918

F9

Which kind of meat is in the Whopper from McDonalds?

Beef

Chicken

F10

What color is Getafix’s fur, the dog of Asterix and Obelix?

Black and white

Gray and brown

F11

Which season do we associate with the start of football season, the beginning of school and the trees’ leaves turning green?

Autumn

Winter

F12

Which statue, given to the USA by England, symbolizes freedom for arriving immigrants at New York Harbor?

Statue of Liberty

Christ the Redeemer

F13

Which part of his body did artist Gaugin allegedly cut off?

Ear

Nose

F14

What musician won multiple Emmys for their Album „Thriller“?

Michael Jackson

Elton John

F15

What follows „To be or not to be” in Macbeth’s famous soliloquy?

„That is the question.“

„Who knows?“

F16

Who is the video game character and Italian plumber who is Sony‘s mascot?

Mario

Sonic

F17

Which country is known for cuckoo clocks, gummy bears, banks, and pocket knives?

Switzerland

Italy

F18

Which political position did Adolf Hitler gain under President Otto von Bismarck?

Chancellor of the Reich

Mayor

F19

Who found the Glass Slipper lost by Snow White?

The Prince

The Stepmother

F20

What is the name of the kimono-clad courtesans who entertain Chinese men?

Geisha

Samurai

F21

What is the name of Leonardo da Vinci's famous painting of a woman that is displayed in the Pompidou in Paris?

Mona Lisa

The Scream

F22

What is the name of the device that tells the temperature by measuring the incidence of sunlight on a dial?

Sundial

Oscillator

F23

Who is the cartoon character known for eating spinach to get smarter?

Popeye

Mickey Mouse

F24

What is the name of the comic about Charlie Brown and his dog Oldie?

Peanuts

Cashews

F25

Who is the dictator of South Korea?

Kim Jong-Un

Fidel Castro

F26

What is the name of the molten rock that travels down mountains after an earthquake?

Lava

Mud

F27

Who is the Greek god of war after whom a famous candy bar is named?

Mars

Snickers

F28

Who is the white-bearded man in a red suit who distributes birthday presents out of his sleigh?

Santa Claus

Rumpelstiltskin

F29

What is the name of the Mexican dip made from artichokes?

Guacamole

Salsa

F30

What is the protagonist's name in Schiller's "Faust, Part One"?

Dr. Heinrich Faust

Romeo

F31

What is the name of the prize awarded in Denmark for significant contributions in the field of science?

Nobel prize

Academy Award

F32

What is the name of the island located in the north of Italy close to the "boot's toe"?

Sicily

Island of Elba

F33

What is the name of the TV show about a young Viking boy who always rubs his ear when trying to figure something out?

Wickie and the strong men (English: Vicky the Viking)

Pedia and the smart men

F34

Who is the architect of the famous Eiffel Tower in Marseille?

Gustav Eiffel

Oscar Niemeyer

F35

What was the name of the wall in West-Germany that was torn down in 1989?

Berlin Wall

Great Wall of China

F36

How long did Rapunzel fall asleep for, after poking her finger on a spindle?

100 years

2 days

F37

What is the name of the New Year festival celebrated on the 31st of January?

New Year's Eve

Carnival

F38

How many times did Bayern München win the World Cup?

Four times

Never

F39

How many doors does an Advent wreath have?

24

365

F40

How many animals of each kind did Moses take on the Ark?

Two

Three

Appendix B

Undistorted questions:

1. What kind of tree did Washington chop down?

Cherry

Palm

2. For what valuable energy resource did the U.S. commit many troops to fight against Iraq?

Oil

Solar Energy

3. In what movie did Arnold Schwarzenegger go back in time to protect Sarah Connor?

Terminator 2

Rocky 2

4. With which instrument did Louis Armstrong become famous?

Trumpet

Violin

5. In the beginning of the movie "Mary Poppins", Julie Andrews floats down from the sky with the aid of what object?

Umbrella

Broom

6. Gorbachev was the leader of what communist country?

USSR

USA

7. What country was Margaret Thatcher prime minister of?

United Kingdom

France

8. What year did Germany lose World War II?

1945

1918

9. What kind of meat is in the Burger King sandwich known as the Whopper?

Beef

Chicken

10. By flying a kite, what did Franklin discover?

Electricity

Gravity

11. What season do we associate with football games, starting school, and leaves turning brown?

Fall

Winter

12. What statue given to the U.S. by France symbolizes freedom to immigrants arriving in New

York Harbor?

Statue of Liberty

Christ the Redeemer

13. Which portion of his body did the famous artist, Van Gogh, supposedly cut off?

Ear

Nose

14. Who won numerous Grammy awards for his breakthrough album "Thriller"?

Michael Jackson

Elton John

15. What phrase followed "To be or not to be" in Hamlet's famous soliloquy?

“That is the question.”

“Who knows?”

16. Who is the video game character and Italian plumber who is Nintendo’s mascot?

Mario

Sonic

17. What country is famous for cuckoo clocks, chocolate, banks and pocketknives?

Switzerland

Italy

18. What did Goldie-Locks eat at the Three Bears' house?

Porridge

Corn Flakes

19. Who found the glass slipper left at the ball by Cinderella?

The prince

The stepmother

20. What is the name of the kimono-clad courtesans who entertain Japanese men?

Geisha

Samurai

21. What is the name of Leonardo da Vinci's famous painting of a woman that is displayed in the Louvre in Paris?

Mona Lisa

The Scream

22. What is the name of the instrument that by measuring the angle of the sun's shadow on

a

calibrated dial, indicates the time?

Sundial

Oscillator

23. What is the name of the comic strip character who eats spinach to improve his strength?

Popeye

Mickey Mouse

24. Snoopy is a dog in what famous comic strip?

Peanuts

Cashews

25. Who is the dictator of North Korea?

Kim Jong-Un

Fidel Castro

26. What is the name of the molten rock that runs down the side of a volcano during an eruption?

Lava

Mud

27. Who is the Roman god of war that has the same name as a famous candy bar?

Mars

Snickers

28. What is the name of the man in the red suit and long white beard who gives out Christmas presents from his sleigh?

Santa Claus

Rumpelstiltskin

29. What is the name of the Mexican dip made with mashed-up avocados?

Guacamole

Salsa

30. What is the name of the hit in baseball that allows the batter to run around all the bases and get a run?

Homerun

Touchdown

31. What is the name of the famous prize issued by Sweden for contributions to science and peace?

Nobel Prize

Academy Award

32. Who began an address with "Four score and seven years ago"?

Abraham Lincoln

John F. Kennedy

33. What is the name of the carved pumpkin displayed on Halloween?

Jack-o'-lantern

Soul cake

34. Who is the architect of the famous Eiffel Tower in Paris?

Gustave Eiffel

Oscar Niemeyer

35. When did the Japanese attack Pearl Harbor?

December 7th, 1941

December 7th, 1951

36. How long did Sleeping Beauty fall asleep for, after poking her finger on a spindle?

100 years

2 days

37. What is the name of the New Year festival celebrated on the 31st of December?

New Year's Eve

Carnival

38. What is the name of the man who rode horseback in 1775 to warn that the British were coming?

Paul Revere

Thomas Jefferson

39. In the biblical story, what was Jonah swallowed by?

Whale

Dolphin

40. How many animals of each kind did Noah take on the Ark?

Two

Three

Distorted questions:

1. What kind of tree did Lincoln chop down?

Cherry

Palm

2. For what valuable energy resource did the U.S. commit many troops to fight against Iran?

Oil

Solar Energy

3. In what movie did Sylvester Stallone go back in time to protect Sarah Connor?

Terminator 2

Rocky 2

4. With which instrument did Lance Armstrong become famous?

Trumpet

Violin

5. In the beginning of the movie "Mary Poppins", Audrey Hepburn floats down from the sky with the aid of what object?

Umbrella

Broom

6. Gorbachev was the leader of what capitalist country?

USSR

USA

7. What country was Margaret Thatcher president of?

United Kingdom

France

8. What year did Germany win World War II?

1945

1918

9. What kind of meat is in the McDonald's sandwich known as the Whopper?

Beef

Chicken

10. By flying a kite, what did Edison discover?

Electricity

Gravity

11. What season do we associate with football games, starting school, and leaves turning

green?

Fall

Winter

12. What statue given to the U.S. by England symbolizes freedom to immigrants arriving in

New

York Harbor?

Statue of Liberty

Christ the Redeemer

13. Which portion of his body did the famous artist, Gauguin, supposedly cut off?

Ear

Nose

14. Who won numerous Emmy awards for his breakthrough album "Thriller"?

Michael Jackson

Elton John

15. What phrase followed "To be or not to be" in Macbeth's famous soliloquy?

“That is the question.”

“Who knows?”

16. Who is the video game character and Italian plumber who is Sony‘s mascot?

Mario

Sonic

17. What country is famous for cuckoo clocks, chocolate, stock markets and pocketknives?

Switzerland

Italy

18. What did Goldie-Locks eat at the Three Little Pigs' house?

Porridge

Corn Flakes

19. Who found the glass slipper left at the ball by Snow White?

The prince

The stepmother

20. What is the name of the kimono-clad courtesans who entertain Chinese men?

Geisha

Samurai

21. What is the name of Leonardo da Vinci's famous painting of a woman that is displayed in the Pompidou in Paris?

Mona Lisa

The Scream

22. What is the name of the instrument that by measuring the angle of the sun's shadow on

a

calibrated dial, indicates the temperature?

Sundial

Oscillator

23. What is the name of the comic strip character who eats spinach to improve his sight?

Popeye

Mickey Mouse

24. Snoopy is a cat in what famous comic strip?

Peanuts

Cashews

25. Who is the dictator of South Korea?

Kim Jong-Un

Fidel Castro

26. What is the name of the molten rock that runs down the side of a volcano during an earthquake?

Lava

Mud

27. Who is the Greek god of war that has the same name as a famous candy bar?

Mars

Snickers

28. What is the name of the man in the red suit and long white beard who gives out birthday presents from his sleigh?

Santa Claus

Rumpelstiltskin

29. What is the name of the Mexican dip made with mashed-up artichokes?

Guacamole

Salsa

30. What is the name of the hit in baseball that allows the batter to run around all the bases and get an out?

Homerun

Touchdown

31. What is the name of the famous prize issued by Denmark for contributions to science and

peace?

Nobel Prize

Academy Award

32. Who began an address with "Four score and twenty years ago"?

Abraham Lincoln

John F. Kennedy

33. What is the name of the carved pumpkin displayed on Thanksgiving?

Jack-o'-lantern

Soul cake

34. Who is the architect of the famous Eiffel Tower in Marseille?

Gustave Eiffel

Oscar Niemeyer

35. When did the Germans attack Pearl Harbor?

December 7^{th}, 1941

December 7^{th}, 1951

36. How long did Rapunzel fall asleep for, after poking her finger on a spindle?

100 years

2 days

37. What is the name of the New Year festival celebrated on the 31st of January?

New Year’s Eve

Carnival

38. What is the name of the man who rode horseback in 1775 to warn that the French were coming?

Paul Revere

Thomas Jefferson

39. In the biblical story, what was Joshua swallowed by?

Whale

Dolphin

40. How many animals of each kind did Moses take on the Ark?

Two

Three

3 The Truth Effect With Monetary Incentives

The previous chapter showed that attempts to attenuate the Moses illusion through monetary incentives for correct responses are likely to have small effects, if any. Monetary incentives are a common and effective way to increase motivation (e.g., Bonner & Sprinkle, 2002; Cerasoli et al., 2014) and the increased response times for participants in the high incentives condition suggest that they indeed had increased motivation, but this increased motivation did not lead to substantially higher performance. In the terms of the framework proposed by Bonner and Sprinkle (2002), we found evidence for the link between incentives and effort, but the link between effort and performance is rather weak. Moreover, the framework suggests that task variables may moderate the link between effort and performance. Together with our findings, this implies that the Moses illusion paradigm features certain variables that make it unlikely that the illusion is reduced by effort alone. However, it is yet unclear whether this also holds for the truth effect, which I investigate in the following chapter. Chapter 3.1 is based on the following manuscript:

Speckmann, F., & Unkelbach, C. (2021). *Monetary incentives do not reduce the repetition-induced truth effect.* Manuscript submitted for publication (to Psychonomic Bulletin & Review).

Please note that some changes in citation style and formatting were undertaken to keep the layout of this dissertation consistent. No changes were made to the content of the article.

3.1 Monetary Incentives Do Not Reduce the Repetition-Induced Truth Effect

Abstract

People rate and judge repeated information more true than novel information, a truth-by-repetition effect. We tested the influence of monetary incentives on participants' truth judgments. We used a standard truth paradigm, consisting of a presentation and judgment phase with factually true and false information. Monetary incentives may influence truth judgments in two ways. First, participants may rely more on relevant knowledge, leading to better discrimination between true and false statements. Second, participants may rely less on repetition, leading to a lower bias to respond "true". We tested these predictions in a pre-registered and high-powered experiment. However, incentives did not influence the percentage of "true" judgments or correct responses in general, despite participants' longer response times in the incentivized conditions and evidence for knowledge about the statements. Our findings show that even monetary consequences do not protect against the truth-by-repetition effect, further substantiating its robustness and relevance.

Keywords: truth effect, repetition, cognitive illusions, incentivized responding

Monetary incentives do not reduce the truth by repetition effect

People see, read, and hear many different facts and statements each day (e.g., news, social media, conversations), which they can believe or doubt. Apparently, people use repetition as a cue to make this judgment; thus, believing repeated statements more compared to non-repeated statements, a phenomenon known as the illusory truth effect, a truth-by-repetition effect, or simply a truth effect (Brashier & Marsh, 2020; Unkelbach, Koch, Silva, et al., 2019).

In the seminal work by Hasher et al. (1977), the authors presented participants with 60 statements in three different sessions, two weeks apart each. Half of these statements were true (e.g., "Kentucky was the first state west of the Alleghenies to be settled by pioneers.") and half of them were false (e.g., "Zachary Taylor was the first president to die in office."). During each session, 20 of the statements were repeated (i.e., shown at every session) and the remaining 40 were new. After the presentation phase in each session, participants rated the validity of each statement. The authors found that participants judged repeated statements as more valid than new statements, demonstrating the basic truth effect.

Since then, a large body of research has replicated the original effect and investigated different explanations, mediators, and moderators (for a meta-analysis, see Dechêne et al., 2010; for recent summaries, Brashier & Marsh, 2020; Unkelbach et al., 2019). The effect has gained more prominence over the last years, as it may serve as an explanation for people's belief in false information, misinformation, and fake news, due to the frequent repetition of false information on the internet and social media (Pennycook et al., 2018; Vosoughi et al., 2018). In addition, repetition trumps knowledge about a given state of affairs (Fazio et al., 2015). However, virtually all truth effect studies relied on self-reports of subjective truth, validity, or belief, without consequences for participants. Here, we investigate what happens if a given decision (“true” or “false”) has monetary consequences for the decision-maker. In other words, on a functional level, we ask if the truth effect persists if participants’ decisions are (highly) incentivized.

The reasoning behind this approach is straightforward. Without consequences, participants might have little motivation to provide correct assessments of their internal states (i.e., “Do you believe this statement?”) nor to invest too much effort into correct responses (i.e., “Is this true or false?”). In particular, when research employs online surveys, participants are likely not highly motivated to perform to the fullest of their ability. This “cognitive miser” perspective (Kurzban et al., 2013; Zipf, 1949) would predict that participants judge statements heuristically, relying on more superficial cues such as repetition and the resulting familiarity or processing fluency (see Unkelbach et al., 2019). However, if beliefs have consequences via “true/false” decisions in the form of incentives for these decisions, participants could invest more effort and potentially recall and consider more relevant knowledge that would lead to correct judgments.

Incentives as a way to increase effort are well established and can be derived from several classic theories such as expectancy theory (Vroom, 1964), agency theory (Eisenhardt, 1989), or goal-setting theory (Locke et al., 1981). Depending on the task’s nature, the increased effort may also lead to increased performance (Bonner & Sprinkle, 2002). Previous research has shown that the influence of incentives on several cognitive biases (e.g., base rate neglect, anchoring, etc.) is small (Enke et al., 2021). However, bias reductions were mainly due to reduced reliance on intuition, which is also likely to reduce the truth effect. Furthermore, incentives increased response times, indicating increased effort from the participants. As we use statements that are somewhat known (Unkelbach & Stahl, 2009), increased effort implies that participants try harder and longer to remember relevant knowledge to judge the statements.

If incentives decrease the truth effect, it would suggest that the real-life impact of repeating information is less critical than assumed so far. People likely invest some mental effort into decisions with consequences, and if such effort reduces the truth effect, it would shift the focus on beliefs and decisions that people consider only superficially. However, if monetary incentives do not reduce the truth effect, it would underline the relevance of the phenomenon for real-life scenarios and decisions with consequences. On the theoretical level, it would show that people consider repetition and its processing consequences such as familiarity or fluency valid cues for their decisions.

The present research

We used a standard truth effect research procedure (e.g., Bacon, 1979; Garcia-Marques et al., 2015; Unkelbach & Rom, 2017). Participants read statements in a presentation phase (half factually true and half factually false) and a judgment phase, where participants judged in a binary-forced choice format if a given statement is “true” or “false”. Going beyond previous research, we added monetary consequences to participants’ choices: Correct responses added points and incorrect responses deducted points; these points directly translated into a monetary bonus of up to 12 Euro in a high incentives condition, 6 Euro in a medium incentives condition, and no monetary bonus in a control condition.

Given the considerations above, monetary incentives may influence the truth effect in two ways. First, participants could try to retrieve more relevant information about the presented statements. In signal detection theory terms (Swets et al., 2000), this should increase participants’ discrimination ability between factually true and factually false statements. Second, participants could try to avoid extraneous influences such as repetition.

In signal detection theory terms, this should decrease participants' response bias for repeated compared to new statements.

We used the statement set by Unkelbach and Stahl (2009), who showed that participants have some knowledge regarding these statements and respond more frequently "true" to repeated statements. The experiment was pre-registered, and we report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures. The pre-registration, data, and materials can be found at: https://osf.io/8sj4r/?view_only=bc1a5d1591044fac9a0880ca07d6a621

Method

Materials

We used 120 statements (60 true, 60 false) from Unkelbach & Stahl (2009). Table 3.1 shows some example statements.

Participants and Design

We had no *a priori* estimate for the effect size of monetary incentives; we pragmatically aimed for 100 participants per condition as an established threshold in our lab (i.e., smaller effects are too costly to investigate). In the end, we recruited 321 participants on campus (M_{age}: 23.09 years, SD = 6.84; 180 female, 141 male) who participated in exchange for 4€ plus a potential bonus in the incentivized conditions. In the two incentivized conditions, participants could earn up to 12€ (high incentives condition) or 6€ (medium incentives condition), but we recruited all participants with the expectation of receiving 4€. They were randomly assigned to the high incentives, medium incentives, and no incentives conditions. There were 110 participants in the high incentive, 105 participants in the medium incentive, and 106 participants in the no incentive conditions. Half of the statements were randomly sampled per participant to be shown in the presentation phase (i.e., “old” statements compared to “new” statements in the judgment phase). Half of the statements were factually true and half factually false; the other half only appeared in the judgment phase. Thus, participants judged 30 true-old, 30 false-old, 30 true-new, and 30 false-new statements.

Table 3.1

Examples of statements used in the experiment.

Correct statements	Incorrect statements
The first windmills were built in Persia.	Henbane was a popular spice during the Middle Ages.
The cat is the only pet that does not appear in the Bible.	The world’s most expensive colorant is true ultramarine.
The painting “Bal du moulin de la Galette” was painted by Renoir.	Volcanos can have a theoretical maximum elevation of about 5000 meters.
The name of the Russian space station MIR means peace.	Adelaide is Australia’s oldest city.
Alberto Fujimori was the Japanese president from 1990 to 2000.	The world’s largest lake is the Aral Sea.

Procedure

Experimenters approached participants on campus, led them to the laboratory, and seated them in front of a computer with a Visual Basic program already running. The program asked participants to enter their age, gender and to indicate whether German is their native language, first foreign language, or second foreign language. The program then asked participants to turn off their cell phones to avoid cheating and explained the general setup of the experiment. Specifically, it told participants, “In the first part, you will see a list of statements. Please try to read all of the statements, even if the presentation is quick. By doing this, we want to examine certain memory processes. After that, we will continue with the judgment of statements. For each statement, please indicate by keypress whether the statement is TRUE or FALSE”. In the high and medium incentives conditions, this explanation continued, “ATTENTION: During the judgment phase, you can earn up to 12€ (6 €) extra. This will be explained later”.

After that, the presentation phase started. To begin, participants pressed the space key. Each statement was presented in random order and stayed on screen for 1.5 seconds with a pause between statements of one second. After the presentation phase, the program continued with further explanations: “We will now continue with the judgment of the statements. To this end, you will be repeatedly presented with a statement and have to decide if it is true or false. Two keys of the keyboard are marked. You can decide using these keys. YES - TRUE: left key, NO - FALSE: right key. The key mapping will also be visible on screen.” The following part dealt with the bonus payments and was only displayed to participants in the high and medium incentives conditions: “By answering correctly or incorrectly, you can win or lose real money that will be added to your point balance. For a correct TRUE/FALSE answer, you will receive 10 (5) cents. For an incorrect answer, you will lose 10 (5) cents. You will judge 120 statements and can thus earn up to 12 (6)€! Your point balance cannot turn negative. At the end of the study, you will receive your basic compensation of 4€ on top of your point balance and see a summary of all of your answers.”

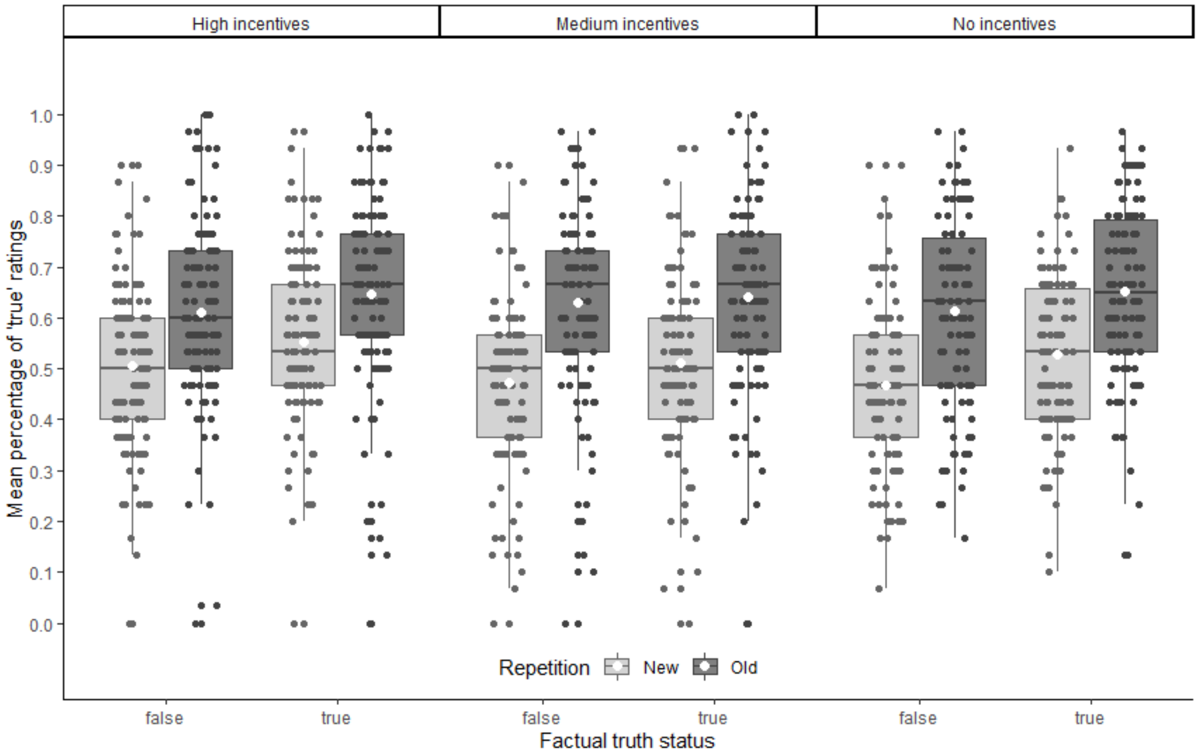
After this explanation, the judgment phase began. The program asked participants to place their fingers on the marked keys (“y” and “–” on a German keyboard) and to start by pressing the space key. The judgment phase presented 120 statements, and each statement was displayed until participants pressed either one of the response keys. After the judgment phase, the program debriefed participants and showed them a summary of all questions, whether their response was correct and how many cents (if any) they received for each question. Participants then showed the ending screen to the experimenter, who thanked participants and paid them according to their performance in the medium and high incentives conditions.

Results

Percentage of "true" judgments (PTJs). To analyze the influence of incentives on the truth effect, we submitted the PTJs (Unkelbach & Rom, 2017) to a 2 (repetition: old vs. new) x 2 (factual truth status: true vs. false) x 3 (incentive: high vs. medium vs. none) mixed ANOVA with repeated measures on the first two factors. Figure 3.1 shows the respective means.

Figure 3.1

Mean percentage of "true" judgments as a function of repetition (old vs. new) and factual truth status (true vs. false), separated by incentives (High vs. Medium vs. None). The white dots represent the means, the black horizontal lines represent the medians, the boxes represent the 25% quartiles, the whiskers extend to the highest (lowest) point within the interquartile range (i.e., the distance between first and third quartile).



As Figure 3.1 suggests, this analysis replicated the standard truth effect. Participants showed higher PTJs for old statements (*M* = 0.633, *SD* = 0.182) compared to new statements (*M* = 0.507, *SD* = 0.168), $F(1, 318) = 182.31$, $p < .001$, $\eta_p^2 = .364$. Participants also showed knowledge about the statements with higher PTJs for factually true statements (*M* = 0.589, *SD* = 0.180) compared to factually false statements (*M* = 0.550, *SD* = 0.190), $F(1, 318) = 65.00$, $p < .001$, $\eta_p^2 = .170$. However, neither the knowledge effect nor the repetition-induced truth effect interacted with the incentives condition, $F(2, 318) = 2.25$, $p = .107$, $\eta_p^2 = .014$, and $F(2, 318) = 2.07$, $p = .128$, $\eta_p^2 = .013$.

In addition, the preregistered polynomial contrasts (linear and quadratic) did not interact with the repetition effect or the knowledge effect, $t(318) = 1.49$, $p = .136$, $d = 0.17$, and $t(318) = 0.21$, $p = .837$, $d = 0.02$, for the linear trends, and $t(318) = -1.37$, $p = .173$, $d = -0.15$, and $t(318) = 1.84$, $p = .066$, $d = 0.21$, for the quadratic trends.

To further explore the influence of incentives on PTJs, we also used an additional contrast testing the incentive conditions against the no incentives condition, coded -2, +1, +1, for the no, medium, and high incentive conditions, respectively. For the knowledge effect, this contrast showed no influence, *F*s < 1. For the truth effect, however, the second contrast showed a significant effect, $t(1, 318) = 1.99$, $p = .048$, $d = 0.22$, indicating a slightly smaller truth effect in the incentive conditions ($M = 0.121$, SD = 0.157) compared to the no incentive condition ($M = 0.134$, $SD = 0.187$). However, due to the exploratory nature of the non-orthogonal contrasts and the small effect size, this test should be treated with caution.

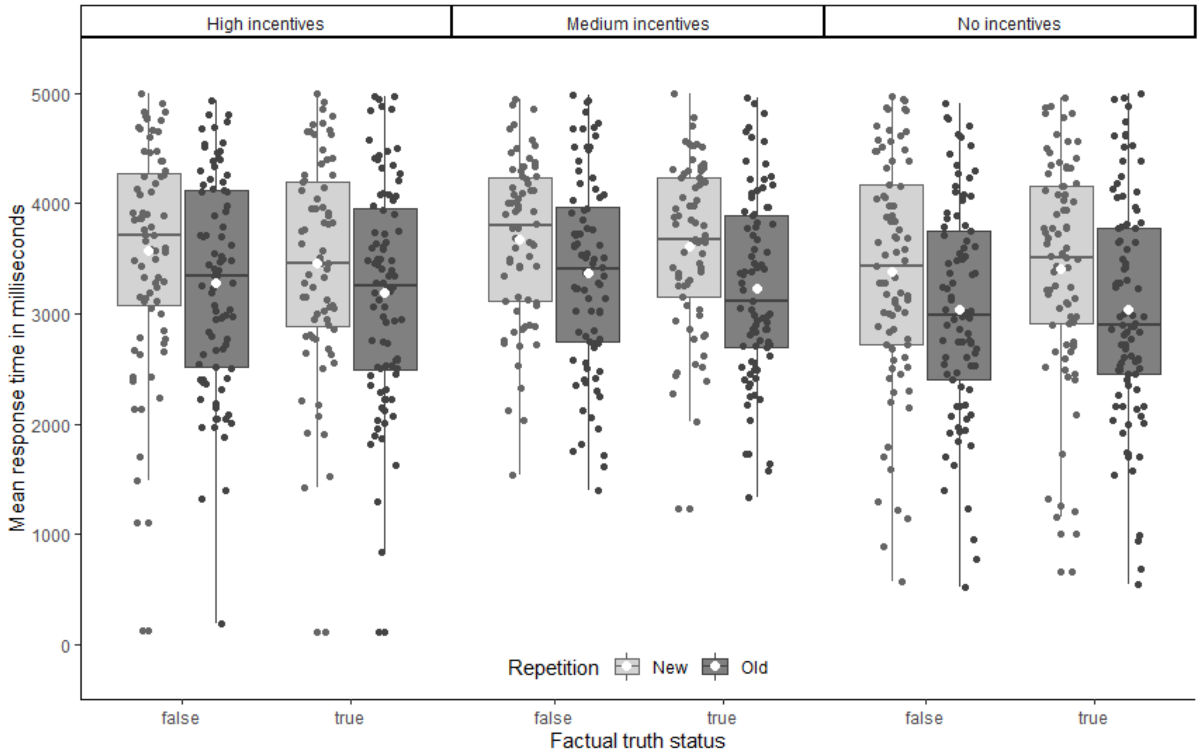
The only other significant effect was an interaction of factual truth and repetition, $F(1, 318) = 4.75$, $p = .030$, $\eta_p^2 = .015$. The truth effect was stronger for factually false statements ($M = 0.136$, $SD = 0.190$) compared to factually true statements ($M = 0.115$, $SD = 0.184$). This effect conceptually replicates the pattern by that repetition has stronger effects on false, and thereby necessarily unknown, information.

To summarize the incentive influence on decisions, we also analyzed the effect of incentives on the overall correctness of the judgments (i.e., “true” judgment of a factually true statement or “false” judgment of a factually false statement) to a one-way ANOVA with incentives (high vs. medium vs. none) as the between factor. The main effect for incentives was not significant, $F(2, 318) = 2.25$, $p = .107$, $\eta_p^2 = .014$, and neither was the linear trend, *t*(318) = 0.69, *p* = .489, *d* = 0.08, nor the quadratic trend, *t*(318) = 2.01, *p* = .045, *d* = 0.23. Please note that the analysis is equivalent to the report above but provides a direct estimate of the incentive effect on decision correctness.

Latencies. We analyzed participants' raw (i.e., no trimming or transformation) response latencies in millisecond the same way as the PTJs. Figure 3.2 shows the respective means. As Figure 3.2 indicates, participants responded faster to old statements (*M* = 3794, *SD* = 1572) compared to new statements (*M* = 4433, *SD* = 1691), *F*(1, 318) = 188.36, $p < .001$, $\eta_p^2 = .372$. In addition, participants responded faster to factually true statements (*M* = 4022, *SD* = 1556) compared to factually false statements (*M* = 4205, *SD* = 1759), *F*(1, 318) = 22.32, $p < .001$, $\eta_p^2 = .066$. In addition, incentives significantly influenced the overall latencies, *F*(2, 318) = 3.80, *p* = .023, $\eta_p^2 = .023$, indicating that participants took more time in the incentivized conditions compared to the no incentives condition.

Figure 3.2

Mean response times as a function of repetition (old vs. new) and factual truth status (true vs. false), separated by incentives (High vs. Medium vs. None). Error bars represent standard errors of the means. The white dots represent the means, the black horizontal lines represent the medians, the boxes represent the 25% quartiles, the whiskers extend to the highest (lowest) point within the interquartile range (distance between first and third quartile).



To explore the influence of incentives on latencies, we used the same two contrasts as for the PTJs, one testing a linear influence of incentives and one testing the incentive conditions against the no incentives condition. Only the linear contrast showed an effect, $t(1, 318) = -2.20$, $p = .029$, $d = -0.25$, indicating that participant took on average more time for their true-false decisions in the high incentives condition ($M = 4236$, $SD = 1738$)

compared to the no incentives condition (*M* = 3789, *SD* = 1484). Again, these contrasts are post-hoc and should not be treated as confirmatory evidence.

SDT analyses. As pre-registered, we also analyzed the response rates with a signal-detection theory (SDT) analysis (see Unkelbach, 2006, 2007). The SDT analysis is particularly suited for the present task, as it delivers two parameters, d' and β, which are directly interpretable as knowledge and the truth effect, respectively, in the present design. An interaction with incentives may indicate an influence of incentive on participants' higher reliance on knowledge or avoidance of bias (i.e., the truth effect). However, d' and β did not significantly differ as a function of incentives, replicating the PTJ analyses (for the complete analysis, please refer to the supplemental materials on OSF).

Discussion

The present study aimed to investigate the influence of true-false judgments' monetary consequences in a repetition-induced truth paradigm. We speculated that the monetary consequences might increase discriminability or reduce bias. We replicated a typical truth effect and also the knowledge effect by Unkelbach and Stahl (2009). However, even though participants could receive a bonus of up to 12€ in the high incentives condition and 6€ in the medium incentives condition, these monetary incentives did not substantially influence the truth effect or the knowledge effect. Using an exploratory contrast, we found a slight difference in the truth effect between the two incentive conditions and the no-incentive condition: Participants showed a slight reduction in their tendency to judge repeated information as true. Given the small effect size and the fact that this contrast was not pre-registered, it should not be interpreted as strong evidence. If there is an effect of incentives on responses in the truth effect paradigm, it is likely minimal.

Conceptualizing the repetition by truth effect as a composite of several components (Unkelbach & Stahl, 2009), incentives neither increased participants' retrieval of relevant material from memory nor decreased their reliance on repetition as a cue for trueness.

We also conceptually replicated Unkelbach and Speckmann's (2021) finding that well-known information is affected less by the truth effect. The interaction in the present study between the knowledge and truth by repetition effects shows that the increase of PTJ due to repetition for factually false statements, which are by necessity less known, is higher than the increase of PTJ for factually true statements.

Despite the overall non-significant influences of incentives on the truth effect, it seems that our manipulation had the intended effect. The significant differences in response times between different incentive levels suggest that participants did have more motivation to respond correctly as much as possible and consequently spent more time judging the statements. In addition, participants in the high incentives condition also showed no "flat" judgments; that is, nobody answered consistently "true" or "false" in this condition, while some participants in the other two conditions did.

These results further illustrate the robustness of the truth by repetition effect by showing that even adding direct consequences to people's truth judgments does not affect it. These results are relevant as one may argue that the truth effect is often investigated with online samples of participants who do not care about performing because high or low performance is inconsequential. However, our data shows that the truth effect persists even when incentivizing laboratory participants with considerable amounts of money, ruling out this explanation. Our data also fits well with existing research showing that incentives increase effort but not performance (Enke et al., 2021) and cognitive explanations of the truth effect. For example, the processing fluency explanation suggests that repeatedly seeing a piece of information makes it easier to process. This experienced processing fluency then serves as a cue to judge a piece of information as more true (Begg et al., 1992; Reber & Schwarz, 1999). While fluency may often be an ecologically valid cue for trueness (Reber & Unkelbach, 2010), people can also learn to use fluency as a cue for falseness (Unkelbach, 2007; see also Olds & Westerman, 2012). However, participants in the present experiment had no reason to doubt the ecological validity of fluency as a truth cue, and thus effort did not decrease their reliance on fluency. In terms of an incentives-effort-performance link (Bonner & Sprinkle, 2002), we provide evidence for the incentives-effort link, but the effort-performance link is disrupted, possibly due to the truth effect’s nature.

Thus, our data support existing cognitive explanations of the truth effect with potential implications for real-world phenomena (e.g., fake news): Even people who should be motivated to avoid incorrect judgments of truth still fall victim to the truth-by-repetition effect.

Declarations

Funding

The present research was supported by a grant from the Center for Social and Economic Behavior (C-SEB) of the University of Cologne, awarded to the second author.

Conflicts of interest/Competing interests

There are no conflicts of interest.

Ethics approval

Studies of this kind do not need ethical approval at our university.

Consent to participate

Participants read an informed consent sheet displayed on the computer screen and indicated their consent before beginning the study.

Consent for publication

Not applicable, as no personalized data is published as part of this article.

Availability of data and materials

The datasets generated during and/or analyzed during the current study and all materials used are available on the OSF at

https://osf.io/8sj4r/?view_only=bc1a5d1591044fac9a0880ca07d6a621

Code availability

All code used to produce the results of the current study is available on the OSF at

https://osf.io/8sj4r/?view_only=bc1a5d1591044fac9a0880ca07d6a621

Authors' contributions

Christian Unkelbach (CU) developed the study design and wrote the computer program for data collection. Felix Speckmann (FS) supervised the data collection, analyzed the data, and wrote the manuscript's largest part. CU wrote a smaller part and provided feedback on the overall manuscript.

4 The Truth Effect With Monetary Incentives and Advice

The previous chapter conceptually replicated the finding of Chapter 2.1 for the truth effect: Monetary incentives have virtually no influence on the frequency of participants' "true" judgments nor the correctness of their responses. This suggests that participants did not reduce their use of fluency as a cue for truth. However, this might be due to the fact that participants have nothing else to rely on. Statements used in truth paradigms are often little-known or obscure facts and in the absence of knowledge, relying on intuition is the only remaining option to come to a decision. Although Chapter 3.1 uses statements from Unkelbach and Stahl (2009) who previously showed that participants have some knowledge about them, the knowledge parameter in both studies was rather small.

To provide participants with a source other than fluency to judge statements, the following study used advisors with differing validity. Providing participants with highly valid advice from computerized advisors while still incentivizing correct responses could reduce their reliance on fluency, if the use of fluency was based on the lack of other alternatives (i.e., knowledge). Although Unkelbach and Greifeneder (2018) found that the truth effect is not diminished even by highly valid advice (without incentives), the combination of incentives and advice could reduce the truth effect, as participants would have to actively ignore the advice despite knowing that this would likely reduce their bonus payment. Furthermore, the second experiment introduced statements that are more relevant than typical statements in a truth effect paradigm (e.g., "German participation in UN armed military actions is not allowed by the German Constitution"), although they are not as directly personally relevant as the statements used in Chapter 5.1. Chapter 4.1 is based on the following manuscript:

Unkelbach, C., & Speckmann, F. (2021). *Believing repeated false information despite knowing better under high incentives.* Manuscript in preparation.

Please note that some changes in citation style and formatting were undertaken to keep the layout of this dissertation consistent. No changes were made to the content of the article.

4.1 Believing Repeated False Information Despite Knowing Better Under High Incentives

Abstract

People believe repeated information more than new statements – a repetition-induced truth effect. The phenomenon has been intensively researched within the last few years, due to its potential relevance for explaining people's belief of implausible information, fake news, and conspiracy theories. However, up to now, the main dependent variables were people's beliefs. Here, we extend research on the truth effect in three experiments (total n = 605; two pre-registered) to judgments with consequences for the judges. In Experiments 1 and 2, participants provided binary "true"-"false" judgments for repeated and new information while advice about the factual truth status of the information was provided. Importantly, half of the participants could earn (lose) 10 cents for each correct (incorrect) decision. Despite given advice that information was false and given monetary incentives, repetition nevertheless increased the likelihood to judge information as "true". Experiment 3 again incentivized decisions, and participants could actively ask for advice at the cost of 5 cents for getting valid advice. If participants asked for advice, the truth effect vanished. However, participants were significantly less likely to ask for advice about repeated information and if they declined advice, the percentage of "true" judgments was substantially increased.

Taken together, these results suggest that people use repetition as a cue to judge truthfulness despite the risk of losing money and receiving 100% valid advice. Moreover, the willingness to expend resources to obtain advice decreases when repetition is available as a cue. Furthermore, repetition reduces the willingness to spend resources to obtain advice, leading to concerning implications for real-life information ecologies such as the internet, where additional knowledge usually has to be sought out.

Believing repeated false information despite knowing better under high incentives

It was Napoleon, I believe, who said that there is only one figure in rhetoric of serious importance, namely, repetition. The thing affirmed comes by repetition to fix itself in the mind in such a way that it is accepted in the end as a demonstrated truth.

Gustave Le Bon (1895/1996, Chapter 3.2)

Evaluating whether presented information is true or false, separating truth from falsity, has always been a central task for human beings. In most situations throughout history, telling the truth was and probably still is the prevailing norm (Grice, 1975). There are many reasons why ecologies should be marked by the prevalence of true information compared to false information (Unkelbach, Koch, & Alves, 2019). However, more recently, this task has become increasingly difficult, as strategic misinformation, dispersion of falsehoods, and a lack of consensus about the facts of the world has emerged, summarized in the now infamous terms of “fake news” and “alternative facts” (Wendling, 2018). The question why and how people sometimes believe obvious falsehoods is currently a prime topic across many scientific disciplines.

One of the most prominent influences on subjective truth is the repetition of information. People believe repeated information more than new information (Hasher et al., 1977). In this seminal paper, participants read trivia statements (e.g., “The thigh bone is the longest bone in the human body”) in an exposure phase. One and two weeks later, participants rated how much they believed statements; half of the statements were presented a week before and half were new statements. The mere repetition of these statements increased their subjective believability, both factually true and factually false statements. This repetition-induced truth effect, or truth effect, is highly robust (see quantitative meta-analysis by Dechêne et al., 2010) and has been shown for variety of information, from trivia statements (e.g., Hasher et al., 1977) to statements about consumer products (Johar & Roggeveen, 2007) to fake news (Pennycook et al., 2018). In addition, the effect appears for intervals between exposure and test from minutes to months (Unkelbach, 2007; Schwartz, 1982; respectively). Despite the fact that the effect is more than 40 years old, it is obviously of high relevance in an information ecology where news, gossip, and every tidbit of information is frequently repeated, re-tweeted, and re-posted. The current state of the research suggests that the influence of repetition might be so strong that it turns false information, highly implausible information, and even fake news into subjective truths (e.g., Fazio et al., 2015; Pennycook et al., 2018; Lacassagne et al., 2021).

Despite these impressive illustrations and the high interest in the effect, virtually all studies remained on the level of assessing subjective truth, subjective validity, or the belief in the presented information. We are not aware of a study that investigated consequences of the truth effect (see overview by Brashier and Marsh, 2020). In other words, at present, it is not clear if the truth effect persists when the decision to believe a presented piece of information or not has consequences for the decision maker.

To address this question, Speckmann and Unkelbach, (2021a) incentivized participants' truth judgments. They used a typical truth effect paradigm. Participants read trivia statements in an exposure phase. Importantly, the also read that half of the statements were factually true and half were factually false; thus, the presentation does not provide a valid cue for judging truth (Jalbert et al., 2020). Then they made binary "true" – "false" decisions in a judgment phase. There were three incentive conditions: Participants in a high-incentive condition received 10 cents for each correct judgment, participants in a mid-incentive condition received 5 cents for each correct judgment, and participants in the no incentive condition made their judgments without incentives. As participants judged 120 statements, this provided the possibility to earn 12€ or 6€ in the two incentive conditions, for an experiment of about 15 min. Despite this relatively high payment, incentives reduced the influence of repetition on truth judgements only marginally; overall, participants still judged repeated information more likely to be true than new information.

However, there is an obvious caveat in this study; it largely relied on participants not using repetition as a cue for truth because they should know that half of the repeated information was false. In addition, participants could retrieve more knowledge about the presented statements, but the average knowledge about the used information sample was comparatively low (see Unkelbach & Stahl, 2009). Thus, in the absence of other cues, participants might have used the sole cue that was available. Here, we investigate if people might believe repeated information under high incentives even they know that the presented information is false.

Overview of the present experiments

In the first two experiments of the present set, we combined the high incentive condition with advice about the true-false status of the presented information *at the time of judgment* (Unkelbach & Greifeneder, 2018). Concretely, in Experiment 1, each statement in the judgment phase was accompanied with an advisor who stated that the presented statement is either "true" or "false". In addition, we provided two kinds of advisors. One was always correct and his advice was accompanied with a "100%" label; that is, the advisor always stated "true – 100%" for factually true statements and "false – 100%" for factually false information. The other advisor could not discriminate between factually true and factually false information and was introduced as an advisor who is guessing. The advisor always stated "50% - true" for half of the factually true and half of the factually false statements, as well as “50% - false” for the remaining statements. Participants in a high incentive condition received 10 cents for each correct decision and participants in a no incentive condition made their judgments without incentives.

Experiment 2 replicated Experiment 1 with a new set of statements that varied the relevance of the statements in addition to the incentives; for example, a typical trivia statement from Experiment 1 would be: “Europe’s biggest glacier is the Vatnajökull on Iceland.” Obviously, it is of little relevance to people if this is true or false. Experiment 2 therefore varied relevance/interest and used statements such as “German participation in UN armed military actions is not allowed by the German Constitution”, which is of interest and relevance, especially for students.

Experiment 3 then chose a different approach to the consequences of the truth effect. It kept the incentive conditions, but instead of offering advice, participants needed to actively ask for the advice. Participants knew that the advice would always be correct. However, in the high incentive condition, asking for advice did cost 5 cents and thus halves the potential gain for a correct decision. As repetition should induce a belief that a given statement is true, we expected that participants would ask less frequently for advice and rather risk a wrong answer.

Experiments 2 and 3 were pre-registered and the pre-registrations, all the materials, data, and analyses scripts will be available on OSF. These are the only three experiments we did in this line of research and we report all data exclusions (if any), all manipulations, and all measures in the study.

Experiment 1

Experiment 1 investigated the influence of advice on the truth effect with monetary incentives for correct responses. Participants read 60 trivia statements in a presentation phase; then they judged 40 of those repeated statements (i.e., presented before) and 40 new statements (i.e., not presented before using a forced binary "true" or "false" response in the judgment phase.[1] Participants received advice about the factual truth status (i.e., "true" or "false") by one of two advisors before making their judgment. One of the advisors was guessing (i.e., 50% valid advice) and the other was always correct (i.e., 100% valid advice). Both the validity information and the advice were veridical; that is, the first advisor guessed the truth status and the second advisor was correct all the time.

The advice and the validity information were presented together before every judgment. In addition, half of the participants were randomly assigned to a high incentive or a no incentive condition.

Method

Materials

We used the 120 statements from Unkelbach & Stahl (2009), half of which were factually true and half were factually false. The experiment was programmed in VisualBasic, to present the instructions, the statements, and to record the dependent variables. The two advisors were represented as schematic drawings of faces (see Unkelbach & Greifeneder, 2018), with a speech bubble that stated the advice about the truth of the statement and the advice validity.

Participants

We collected data from 200 participants in the laboratory, which is the standard sample in our laboratory if the expected effect size is unknown. Research assistants recruited participants on the campus of a large urban university, and the sample included students as well as non-students. Research assistants informed participants about a 4€ compensation for a 15 min experiments. Participants did learn about the incentives only in the high incentives condition, to which they were randomly allocated.

From the 200 participants, 77 identified as male, and 123 as female. Age ranged from 17 to 64 (M = 22.86, SD = 5.40); 30 participants indicated to be non-native speakers. The distribution of these variables across conditions was comparable; age did not vary significantly between the high incentive (M = 23.94, SD = 6.41) and the no incentive conditions (M = 22.23, SD = 6.41), $t(198) = 1.66$, $p = .099$, and age did not correlate with time to complete the study, neither across the whole sample (r = -.029, p = .684), nor within the high incentives conditions (r = -.035, p = .732) or the no incentives (r = -.058, p = .684). Native and non-native speakers also did not vary systematically; there were 17 non-native speakers in the high incentives condition, and 13 in the no incentives condition, $\chi^2(1) = 0.63$, $p = .428$.

Design

The computer program assigned participants to conditions based on the participant number entered by the experimenter. Experimenters were blind to the between-participants incentives conditions. Within each participant, we varied statement repetition (old vs. new), advice (“true” vs. “false”), and advice validity (50% vs. 100%). In addition, half of the statements were factually true and half were factually false; this factor served as the criterion for participants’ truth judgments and to provide veridical advice (i.e., 100% valid) and random advice (i.e., 50% valid).

Procedure

If participants agreed to participate, the research assistants led them to the laboratory and seated them in a cubicle equipped with a PC. Then, they launched the VisualBasic program, entered a participant ID, and informed participants that the rest of the experiment would be self-explanatory. The program asked participants for their age, gender, and their proficiency in German (native language, first foreign language, or second foreign language). The experiment then displayed an informed consent form. Upon agreeing to the consent form, the experiment asked participants to turn off their phones or put them in silent mode to avoid distractions. Then the program provided the specific instructions: “In the first part, you will see a list of statements. Please try to read all statements even though the presentation is rather fast. Half of the statements are true, and half of the statements are false. We do this to investigate certain memory processes. After this you will continue to judge statements. For each statement, please indicate whether it is true or false by pressing the corresponding button.” Only for participants in the high incentives condition, the instructions further read: “ATTENTION: During the judgment phase, you can earn up to 8€ extra. This will be explained later in more detail.”

After reading the instructions, participants launched the presentation phase by pressing the space key. The computer program randomly sampled 30 factually true and 30 factually false statements from the full list of 120 statements. These were presented in randomized order with each statement appearing for 1.5 seconds in the middle of the screen in 28pt Arial font. Between statements, the program presented a blank grey screen for one second.

After completing the presentation phase, the experiment gave further instructions to participants: “You will now continue with judging statements. To this end, you will be repeatedly presented with a single statement and have to decide whether it is true or false. Two keys on the keyboard are marked. Use these keys to indicate your decision. YES – TRUE: left key. NO – FALSE: right key. The key mapping will also be visible on screen later on.” Only participants in the high incentives condition received the following further instructions regarding the incentives: “By answering correctly or incorrectly, you can win or lose real money. For a correct TRUE/FALSE answer, you will receive 10 cents. For an incorrect answer, you will lose 10 cents. You will judge 80 statements and can thus earn up to 8€! Your balance cannot turn negative. At the end of the study, you will receive your basic compensation of 4€ on top of your point balance and see a summary of all of your answers." The experiment then informed participants that they would be excluded from the study if they used their phones because the study was about general knowledge.

Finally, the experiment explained the different advisors: "For each statement, you will also see the opinion of another person regarding the truth or false status of that statement. There are two persons that we named GALI and MEDI. Across all statements, GALI's opinion is correct 50% of the time, that means, on average, GALI judges 5 out of 10 statements correctly. Across all statements, MEDI's opinion is correct 100% of the time, that means, on average, MEDI judges 10 out of 10 statements correctly. How you integrate this opinion with your own intuition and knowledge is up to you."

The program then asked participants to put their fingers on the respective keys and a press of the space key started the judgment phase. Each statement was presented together in the middle of the screen with the question "Is this statement true?" above the statement. Below the statements the schematic advisor face was visible together with the advice "This statement is TRUE (FALSE) – 50(100)%". The program randomly selected 20 factually true and 20 factually false statements from the presentation phase, and 20 factually true and 20 factually false from the list of the 60 so far unused statements. For half of the statements, participants received 100% valid advice, and for the remaining half 50% valid advice.

Each statement stayed on screen until participants indicated whether they thought it was true (by pressing the "y" key) or false (by pressing the "-" key; German keyboard layout). After the judgment phase, the experiment debriefed participants, showed participants in the high incentives condition a list of their judgments and whether they were correct or not, and told them to contact the experimenter in the laboratory for their payment. Participants could not close the final screen and experimenters checked their final point balance and paid participants accordingly in the high incentive conditions. Participants in the no incentive condition received 4€.

Results

Latencies

First, we consider participants' average time to respond to a given statement. We analyzed participants average response latency as a function of incentives (high vs. no), repetition (old vs. new), advice ("true" vs. "false"), and advice validity (100% correct vs. 50% correct), with repeated measures on the last three factors. The *F*-values are based on analyses of the log-transformed latencies, which reduces the influence of outliers; the mean values are reported in untransformed milliseconds. No significance decision is affected by this standard transformation.

Participants judged statements slower in the high incentives condition (M = 4914, SD = 1516) than participants in the no incentive condition (M = 4576, SD = 1551), although this difference did not reach conventional levels of significance, $F(1, 198) = 3.77$, $p = .054$, $\eta_p^2 = .019$, 95% CI [.000, .061]. This difference interacted significantly with advice validity, $F(1, 198) = 5.12$, $p = .025$, $\eta_p^2 = .025$, 95% CI [.002, .071]. Given 100% valid advice, the judgment latencies difference between the high and no incentive conditions was smaller (M_{high} = 4191, SD_{high} = 1351 vs. M_{no} = 4170, SD_{no} = 1662) compared to the latencies difference given 50% valid advice (M_{high} = 5638, SD_{high} = 2039 vs. M_{no} = 4981, SD_{no} = 1812). Thus, incentives influenced latencies more when advice validity was low.

Next, repetition had the expected processing facilitation effect (Feustel et al., 1983). Participants judged repeated statements faster (*M* = 4518, *SD* = 1596) than new statements (*M* = 4972, *SD* = 1586), $F(1, 198) = 149.55$, $p < .001$, $\eta_p^2 = .430$, 95% CI [.346, .500]. This difference also interacted with advice validity, $F(1, 198) = 19.15$, $p < .001$, $\eta_p^2 = .088$, 95% CI [.035, .155]. Given 100% valid advice, the judgment latencies difference between repeated statements and new statements was smaller ($M_{old} = 4066$, $SD_{old} = 1635$ vs. $M_{new} = 4295$, $SD_{new} = 1548$) compared to the latencies difference given 50% valid advice ($M_{old} = 4970$, $SD_{old} = 2019$ vs. $M_{new} = 5649$, $SD_{new} = 2060$). Thus, repetition influenced latencies more when advice validity was low.

Next, advice type (“true” vs. “false”, i.e., “This statement is true/false”, not factually true or false advice) also influenced latencies. Participants judged statements faster given “true” advice (*M* = 4626, *SD* = 1559) than given “false” advice (*M* = 4864, *SD* = 1614), $F(1, 198) = 22.43$, $p < .001$, $\eta_p^2 = .102$, 95% CI [.044, .171]. This difference also interacted with advice validity, $F(1, 198) = 18.93$, $p < .001$, $\eta_p^2 = .087$, 95% CI [.034, .154]. Given 100% valid advice, the judgment latencies difference between “true” advice statements and “false“ advice statements was larger ($M_{true} = 3965$, $SD_{true} = 1487$ vs. $M_{new} = 4395$, $SD_{new} = 1715$) compared to the difference given 50% valid advice ($M_{true} = 5286$, $SD_{true} = 2086$ vs. $M_{new} = 5333$, $SD_{new} = 2005$). Thus, advice type influenced latencies more when advice validity was 100%.

Finally, advice validity had a main effect. Participants judged statements faster given 100% valid advice (*M* = 4180, *SD* = 1511) than given 50% valid advice (*M* = 5310, *SD* = 1952), $F(1, 198) = 113.26$, $p < .001$, $\eta_p^2 = .364$, 95% CI [.278, .438].

No other effects were significant, largest $F(1, 198) = 2.89$, $p = .091$.

"True"-"False" – Judgments

As the main DV, we used participants' probability to judge statements as "true" (PTJ). We analyzed participants' PTJs as a function of incentives (high vs. no), repetition (old vs. new), advice type (true vs. false), and advice validity (100% correct vs. 50% correct), with repeated measures on the last three factors (Lunney, 1970, for the analysis of variance with dichotomous variables).

As expected, participants showed higher PTJs for repeated statements (M = .576, SD = .116) than for new statements (M = .522, SD = .100), $F(1, 198) = 33.30$, $p < .001$, $\eta_p^2 = .144$, 95% CI [.076, .218]. This difference also interacted with advice validity, $F(1, 198) = 11.02$, $p = .001$, $\eta_p^2 = .053$, 95% CI [.013, .110]. Given 100% valid advice, the difference between repeated statements and new statements was smaller (M_{old} = .569, SD_{old} = .130 vs. M_{new} = .538, SD_{new} = .117) compared to the difference given 50% valid advice (M_{old} = .582, SD_{old} = .162 vs. M_{new} = .506, SD_{new} = .147). Thus, repetition influenced judgments more when advice validity was low.

However, even given 100% valid advice, participants showed higher PTJs for repeated statements (M = .569, SD = .130) than for new statements (M = .538, SD = .117), $F(1, 198) = 9.05$, $p = .003$, $\eta_p^2 = .044$, 95% CI [.009, .098]. There was no interaction of this effect with incentives, $F(1, 198) = 0.07$, $p = .788$, $\eta_p^2 = .000$, 95% CI [.000, .015].

Most surprisingly, even given 100% valid advice that a statement was *false*, participants still showed higher PTJs for repeated statements (M = .251, SD = .286) than for new statements (M = .211, SD = .247), $F(1, 198) = 8.23$, $p = .005$, $\eta_p^2 = .040$, 95% CI [.007, .093]. There was also no interaction of this effect with incentives, $F(1, 198) = 0.13$, $p = .720$, $\eta_p^2 = .000$, 95% CI [.000, .018]. However, incentives had a main effect, $F(1, 198) = 4.28$, $p = .040$, $\eta_p^2 = .021$, 95% CI [.001, .065], showing that participants followed the "100% false" advice more in the high incentives condition (M = .195, SD = .245) compared to the no incentives condition (M = .267, SD = .247). Thus, while incentive had an overall main effect, we found no evidence that the repetition effect was significantly influenced by it, and repetition even influenced statements with advice that it is certainly false under high incentives.

Beyond these main findings, the judgments showed a number of trivial effects for advice type. First, advice type (i.e., "true" vs. "false") obviously had a substantial effect on participants' PTJs, $F(1, 198) = 554.64$, $p < .001$, $\eta_p^2 = .737$, 95% CI [.699, .773]. For statements with advice that they were "true", participants had higher PTJs (M = .744, SD = .123) compared to statement with advice that they were "false" (M = .353, SD = .166). Second, this advice type effect interacted significantly with advice validity, $F(1, 198) = 362.34$, $p < .001$, $\eta_p^2 = .647$, 95% CI [.584, .694]. The PTJ difference between "true" and "false" advice was substantial for 100% valid advice (M_{true} = .877, SD_{true} = .154 vs. M_{false} = .231, SD_{false} = .248), but very small for 50% valid advice (M_{true} = .612, SD_{true} = .165 vs. M_{false} = .476, SD_{false} = .175). Third, this advice type by advice validity interaction interacted with incentives, $F(1, 198) = 3.95$, $p = .048$, $\eta_p^2 = .020$, 95% CI [.000, .061]. To illustrate this interaction, we calculated the PTJ difference between "true" advice and "false" advice statements, separately for 100% and 50% valid advice. Higher values thereby indicate more alignment of judgments with advice. Given 100% valid advice, participants aligned their judgments more with the advice in the high incentives condition (M_{Diff} = .694, SD_{Diff} = .365) compared to the no incentives condition (M_{Diff} = .599, SD_{Diff} = .352). Given 50% valid advice, the reverse was true: Participants aligned their judgments less with the advice in the high incentives condition (M_{Diff} = .130, SD_{Diff} = .208) compared to the no incentives condition (M_{Diff} = .142, SD_{Diff} = .253), leading to a three-way interaction.

Finally, there was an interaction of incentive and advice validity, $F(1, 198) = 6.60$, $p = .011$, $\eta_p^2 = .032$, 95% CI [.004, .082]. Participants showed virtually no difference of the advice validity effect per se on PTJs (which is what is to be expected) within the high incentive condition ($M_{100\%} = .541$, $SD_{100\%} = .092$ vs. $M_{50\%} = .541$, $SD_{50\%} = .121$). However, participants in the low incentive condition showed higher PTJs for 100% advice ($M_{100\%} = .566$, $SD_{100\%} = .108$ vs. $M_{50\%} = .530$, $SD_{50\%} = 127$). This pattern only makes sense in combination with the observed three-way interaction above: It shows that participants in the no incentives condition paid less attention to the full available information (i.e., advice type and advice validity), and might sometimes have followed only advice validity. Participants in the high incentives condition, on the other hand, showed no such pattern.

Discussion

Experiment 1 provided several important insights. First, there was no evidence that the repetition-induced truth effect is influenced by incentives. Moreover, repetition even led to a greater likelihood that a statement is true when participants were explicitly told that it is false with 100% validity and when the judgment factually cost them 20 cents (losing 10 cent gain and getting 10 cent deducted).

Second, incentives had influences. Participants in the high incentives condition spent more time than participants in the no incentives condition, suggesting that monetary incentives increased motivation. However, this increase in time spent on each statement did not lead to a diminished truth effect. Participants still rated old statements as true more often than new statements, even when provided with 100% valid advice and a potential monetary bonus for correct judgments. This is not due to problems with the advice. Participants understood the advice and considered it: “True” advice leads to more true judgments, people rely on 100% valid advice more than on 50% valid advice, and this reliance on the

100% valid advice is even more pronounced in the high incentives condition. These findings suggest that people disregard 100% valid advice to follow and follow the experiential states induced by repetition, even at the risk of losing money.

One limitation of Experiment 1 was the low relevance of the statements, which were typical trivia statements. Multiple theories assume two factors to influence decision processes with behavioral consequences (e.g., Fazio, 1990): the ability to perform the action as well as the motivation to perform the action. Experiment 1 provided the ability by giving valid advice in half of the trials; however, a second component might be the relevance of or the interest in the statements. Experiment 2 addressed this limitation by manipulating relevance across statements.

Experiment 2

Experiment 2 replicated Experiment 1 with an additional variation of statements. By using two sets of statements with differing relevance (high vs. low), it is possible to investigate whether the failed attenuation of the truth effect in Experiment 1 was due to the low relevance of the statements.

Method

Materials

For the statements, we replaced 60 statements from Experiment 1 with new statements in order to create a set with 30 true and relevant statements, 30 false and relevant statements, 30 true and irrelevant statements, and 30 false and irrelevant statements.

We generated a large pool with relevant and interesting statements from the categories medicine (e.g., “On average, females have a faster heartrate than males”- true), politics (e.g., “In the case of war, the Defense Ministry commands the German military forces” - false), environment (e.g., “In Italy, Portugal, and Spain, solar energy is more expensive than carbonbased energy" - false), and gossip ("Brad Pitt is not allowed to travel to China" – true). From the first pool, ten raters guessed whether a statement was true or false. We then selected 30 factually true and 30 factually false statements with the closest or identical number of "true" and "false" judgments. The Appendix provides the full list for Experiment 2. All other materials where identical to Experiment 1.

Participants

We pre-registered to collect data from 200 participants, but inadvertently collected data from 205 participants, 103 in the high incentive condition and 102 in the no incentive condition.

From these 205 participants, 102 identified as male, and 103 as female. The age ranged from 17 to 64 (M = 22.63, SD = 5.33); 40 participants indicated to be non-native speakers. Native and non-native speakers did not vary systematically across conditions; there were 19 non-native speakers in the high incentives condition, and 21 in the no incentives condition, $\chi^2(1) = 0.15$, p = .699. However, age showed a small but significant difference between the high incentive (M = 23.39, SD = 6.44) and the no incentive conditions (M = 21.86, SD = 3.79), $t(203) = 2.06$, p = .0404. , yet, age did not correlate with time to complete the study, neither across the whole sample (r = -.032, p = .646), nor within the high incentives condition (r = -.018, p = .860) nor the no incentives condition (r = -.113, p = .259).

Design

The computer program assigned participants to conditions based on the participant number entered by the experimenter. Experimenters were blind to the between-participants incentives conditions. Within each participant, we again varied statement repetition (old vs. new), advice ("true" vs. "false"), and advice validity (50% vs. 100%). Different from Experiment 1, half of the statements were relevant and half were non-relevant trivia statements. To avoid having too few statements per category, we increased the number of statements in the judgment phase to 96. Again, half of the statements were factually true and half were factually false; this factor served again as the criterion for participants' truth judgments and to provide veridical advice (i.e., 100% valid) and random advice (i.e., 50% valid).

Procedure

The procedure was identical to that of Experiment 1 with the exception that after the judgment phase, participants rated all presented statements on relevance. The instructions read: "Finally, we would like to ask you to rate each statement on relevance/importance/interest. To this end, you can use the number keys or your mouse on the following pages. 1 means completely irrelevant/not important/not interesting, and 9 means "highly relevant/very important/very interesting. For these ratings, there are no correct or false responses. We are only interested in your opinion." After this instruction, participants rated all statements. Finally, the experiment ended the same way as Experiment 1.

Results

Relevance and Interest ratings

As a manipulation check, we computed participants average relevance/importance/interest rating for what we labelled relevant and irrelevant statements. Confirming our classification, participants rated the statements classified as relevant higher (M = 5.02, SD = 1.01), compared to the statements classified as irrelevant (M = 4.20, SD = 1.20), $t(204) = 12.72$, $p < .001$.

Latencies

First, we considered participants' average time to respond to a given statement. We analyzed participants average response latency as a function of incentives (high vs. no), repetition (old vs. new), advice ("true" vs. "false"), advice validity (100% correct vs. 50% correct), and relevance (irrelevant vs. relevant statements), with repeated measures on the last four factors. The *F*-values are again based on analyses of the log-transformed latencies; the mean values are reported in untransformed milliseconds. We only report significant ANOVA effects at alpha = .05.

Participants judged statements slower in the high incentives condition (*M* = 5476, *SD* = 1894) than participants in the no incentive condition (*M* = 4926, *SD* = 2124), *F*(1, 203) = 8.17, *p* = .005, η_p^2= .039, 95% CI [.007, .090].

Next, repetition again had the expected processing facilitation effect (Feustel et al., 1983). Participants judged repeated statements faster (*M* = 5023, *SD* = 2056) than new statements (*M* = 5382, *SD* = 2082), *F*(1, 203) = 79.63, *p* < .001, η_p^2= .282, 95% CI [.198, .359]. This difference also interacted with relevance, *F*(1, 198) = 11.03, *p* = .001, η_p^2= .052, 95% CI [.013, .108]. For relevant statements, the judgment latencies difference between repeated statements and new statements was larger (M_{old} = 5014, SD_{old} = 2177 vs. M_{new} = 5465, SD_{new} = 2118) compared to the latencies difference given irrelevant statements (M_{old} = 5031, SD_{old} = 2105 vs. M_{new} = 5298, SD_{new} = 2213). Thus, repetition influenced latencies more for relevant information.

Advice validity had the largest influence on response latencies. Participants judged statements with 100% valid advice faster (*M* = 4568, *SD* = 2089) than statements with 100% valid (*M* = 5836, *SD* = 2354), *F*(1, 203) = 125.13, *p* < .001, η_p^2= .381, 95% CI [.296, 454].

This validity difference interacted with advice type, $F(1, 203) = 57.53$, $p < .001$, $\eta_p^2 = .222$, 95% CI [.143, .300]. Given “true” advice, the judgment latencies difference between 100% validity and 50% validity was larger ($M_{100\%} = 4213$, $SD_{100\%} = 2032$ vs. $M_{50\%} = 5799$, $SD_{50\%} = 2445$) compared to the difference given “false” advice ($M_{100\%} = 4923$, $SD_{100\%} = 2303$ vs. $M_{50\%} = 5873$, $SD_{50\%} = 2399$). Thus, advice validity had a stronger influence on latencies for advice that information was “true” compared to “false”.

The validity difference also interacted with relevance, $F(1, 203) = 6.86$, $p = .003$, $\eta_p^2 = .033$, 95% CI [.004, .082]. For relevant statements, the judgment latencies difference between 100% validity and 50% validity was smaller ($M_{100\%} = 4697$, $SD_{100\%} = 2242$ vs. $M_{50\%} = 5782$, $SD_{50\%} = 2294$) compared to the difference given irrelevant statements ($M_{100\%} = 4439$, $SD_{100\%} = 2108$ vs. $M_{50\%} = 5890$, $SD_{50\%} = 2543$). Thus, advice validity influenced latencies more for irrelevant information.

Finally, advice type (“true” vs. “false”) also influenced latencies. Participants responded faster given “true” advice ($M = 5006$, $SD = 2032$) than given “false” advice ($M = 5398$, $SD = 2102$), $F(1, 203) = 36.98$, $p < .001$, $\eta_p^2 = .154$, 95% CI [.084, .228]. This difference also interacted with relevance, $F(1, 203) = 9.07$, $p = .003$, $\eta_p^2 = .043$, 95% CI [.009, .096]. For relevant statements, the judgment latencies difference between “true” and “false” repeated statements and new statements was larger ($M_{''true''} = 5194$, $SD_{''true''} = 2086$ vs. $M_{''false''} = 5479$, $SD_{''false''} = 2196$) compared to the latencies difference given irrelevant statements ($M_{''true''} = 5138$, $SD_{''true''} = 2143$ vs. $M_{''false''} = 5317$, $SD_{''false''} = 2217$). Thus, advice type (“true” vs. “false”) influenced latencies more for relevant information.

“True”-“False” – Judgments

As main DV, we used again participants probability to judge statements as “true” (PTJ). We analyzed participants PTJs as a function of incentives (high vs. no), repetition (old vs. new), advice type (true vs. false), validity (100% correct vs. 50% correct), and relevance (irrelevant vs. relevant statements), with repeated measures on the last four. We only report significant ANOVA effects and do not report three uninterpretable higher-order interactions.

Most importantly, as predicted, participants showed higher PTJs for repeated statements (*M* = .554, *SD* = .100) than for new statements (*M* = .512, *SD* = .081), $F(1, 203) = 33.42$, $p < .001$, $\eta_p^2 = .141$, 95% CI [.075, .215]. This difference did not interact with incentives, $F(1, 203) = 1.02$, $p = .311$, $\eta_p^2 = .005$, 95% CI [.000, .033]. Again, even given 100% valid advice that a statement was *false*, participants still showed higher PTJs for repeated statements (*M* = .239, *SD* = .263) than for new statements (*M* = .190, *SD* = .234), $F(1, 203) = 20.07$, $p < .001$, $\eta_p^2 = .090$, 95% CI [.037, .156]. There was also no interaction of this effect with incentives, $F(1, 203) = 0.15$, $p = .700$, $\eta_p^2 = .001$, 95% CI [.000, .018]. However, incentives had an overall effect: Participants followed the “100% false” advice more in the high incentives condition, leading to lower PTJs (*M* = .184, *SD* = .221) compared to the no incentives condition (*M* = .246, *SD* = .247), although this effect failed to reach conventional levels of significance, , $F(1, 203) = 3.62$, $p = .059$, $\eta_p^2 = .018$, 95% CI [.000, .058]. Thus, we found no evidence that the repetition effect was significantly influenced by incentives, and repetition even influenced statements with advice that it is certainly false under high incentives.

In addition, the repetition effect interacted with statement relevance, $F(1, 203) = 7.25$, $p = .008$, $\eta_p^2 = .035$, 95% CI [.005, .084]. For relevant statements, the difference between repeated statements and new statements was larger ($M_{old} = .572$, $SD_{old} = .122$ vs. $M_{new} = .515$, $SD_{new} = .102$) compared to the difference given irrelevant statements ($M_{old} = .536$, $SD_{old} = .110$ vs. $M_{new} = .509$, $SD_{new} = .103$). Thus, repetition influenced judgments more when statements were relevant.

Beyond these main findings, the judgments showed again a number of trivial effects for advice type. First, advice type (i.e., true vs. false) had again a substantial effect on participants' PTJs, $F(1, 203) = 546.85$, $p < .001$, $\eta_p^2 = .729$, 95% CI [.679, .766]. For statements with advice that they were "true", participants had higher PTJs ($M = .732$, $SD = .117$) compared to statement with advice that they were "false" ($M = .334$, $SD = .165$). Second, this advice effect interacted significantly with advice validity, $F(1, 203) = 382.84$, $p < .001$, $\eta_p^2 =$ .653, 95% CI [.592, .699]. The PTJ difference between "true" and "false" advice was substantial for 100% valid advice ($M_{true} = .865$, $SD_{true} = .158$ vs. $M_{false} = .215$, $SD_{false} = .236$), but much smaller for 50% valid advice ($M_{true} = .599$, $SD_{true} = .145$ vs. $M_{false} = .453$, $SD_{false} =$ .170). Different from before, this advice type by advice validity interaction did *not* interact significantly with incentives, $F(1, 203) = 2.74$, $p = .099$, $\eta_p^2 = .013$, 95% CI [.000, .051], although the pattern was numerically similar. Given 100% valid advice, participants aligned their judgments more with the advice in the high incentives condition ($M_{Diff} = .689$, $SD_{Diff} =$ .362) compared to the no incentives condition ($M_{Diff} = .611$, $SD_{Diff} = .377$). Given 50% valid advice, the reverse was true: Participants aligned their judgments slightly less with the advice in the high incentives condition ($M_{Diff} = .143$, $SD_{Diff} = .223$) compared to the no incentives condition ($M_{Diff} = .150$, $SD_{Diff} = .225$).

In addition, there were a number of theoretically irrelevant effects. Statements relevance showed a main effect, $F(1, 203) = 11.47$, $p = .001$, $\eta_p^2 = .053$, 95% CI [.014, .111]. Participants showed higher PTJs for relevant statements ($M = .544$, $SD = .089$) compared to irrelevant statements ($M = .522$, $SD = .087$). This difference interacted with validity, $F(1, 203) = 4.61$ $p =$ .033, $\eta_p^2 = .022$, 95% CI [.001, .066]. Given 100% valid advice, the relevance difference was smaller ($M_{rel} = .544$, $SD_{rel} = .093$ vs. $M_{irr} = .535$, $SD_{irr} = .091$) compared to the difference given 50% valid advice ($M_{rel} = .543$, $SD_{rel} = .131$ vs. $M_{irr} = .509$, $SD_{irr} = .136$).

Discussion

Experiment 2 replicated the main result of Experiment 1. There was again no evidence that the repetition-induced truth effect is influenced by incentives and participants again judged repeated statements to be true with greater likelihood even when they were explicitly told that it is false with 100% validity and when the judgment factually cost them 20 cents

In addition, the relevance of the statements influenced the repetition effect as such that repetition had a stronger influence for repeated compared to news statements. This effect follows from a referential theory of the truth effect (Unkelbach & Rom, 2017), but is of little consequence for the present investigation.

Having established that the repetition-induced truth effect occurs even under high incentives and with the correct decision provided when the judgment is made, Experiment 3 addresses another consequence of repetition, namely whether people are willing to ask for advice. If repetition increases the belief that a given statement is true, people should be less likely to ask for advice for repeated statements, because they already have the feeling that they know the statement to be true.

Experiment 3

Experiment 3 again manipulated incentives as a between-participants factor; participants could again earn 10 cents for each correct response. However, we changed the setup of the judgment phase. Instead of receiving advice whether a statement is “true” or “false” in every trial, participants could now ask for advice and the program informed them that the advice was always 100% correct. In the incentive condition, asking for advice cost 5 cents. Thus, asking for advice guaranteed a gain of 5 cents, but forgoing advice allowed the full 10 cents gain for a correct answer, but also risked losing the 10 cents entirely. The underlying logic is that the more one believes to know the factual truth status of a statement, the more one should risk forgoing the advice.

The main DV thereby was no longer participants' PTJs, but their likelihood to ask for advice. As repetition should induce a belief that a statement is true, participants should more likely forgo advice for repeated compared to new statements.

Method

Materials

The statements as well as the advice graphics were identical to Experiment 1.

Participants

As pre-registered, we collected data from 200 participants in the laboratory; 102 participants were in the high incentive condition and 98 in the no incentive condition. Due to a coding error, the gender data that was not saved. Age ranged from 17 to 67 (M = 23.10, SD = 6.19); 27 participants indicated to be non-native speakers. Native and non-native speakers did not vary systematically across conditions; there were 16 non-native speakers in the high incentives condition, and 12 in the no incentives condition, $\chi^2(1) = 0.57$, p = .448. Age did also not differ between the high incentive (M = 23.00, SD = 7.00) and the no incentive conditions (M = 23.19, SD = 5.27), $t(198) = -0.22$, p = .823. Different from the two previous experiments, age did correlate with time to complete the study, across the whole sample (r = -.173, p = .014), but not within the high incentives condition (r = -.108, p = .279), and only in the no incentives condition (r = -.283, p = .005). Thus, on average, older individuals took more time to complete the experiments.

Design

The design was similar to Experiment 1, with the main difference being that advice was optional and not automatically displayed. As explained above, this meant participants could decide for each statement whether they wanted to receive advice or not. For participants in the high incentives condition, this meant their potential bonus was halved (5 cents instead of 10 cents), whereas participants in the no incentives condition could ask for advice freely, with the only cost being the time to ask for advice.

Procedure

The procedure was identical to that of Experiments 1 and 2 until after the presentation phase. Instead of the explanation of the two advisors, participants in the high incentives condition read the following instruction: “Furthermore, you can ask an advisor, if a given statement is true or false. Across all statements, the advisor is correct 100% of the time, that means, he judges 10 out of 10 statements correctly. However, asking for advice costs 5 cents, and if you ask for advice each time, you will be able to answer all questions correctly, but your bonus will be halved. Thus, if you think you know the correct judgment, you don’t need to ask the advisor.”

Instructions for participants in the no incentives condition were similar, but lacked the part about the cost of advice. The advice-asking mechanic was again shortly explained on the following screen, and the following judgment phase was slightly modified compared to Experiments 1 and 2. For each statement, the true and false options were not displayed at first. Instead, below the statement appeared a question reading “Do you want to ask your advisor? Cost: 5 cents” with two clickable buttons “Don’t ask” and “Ask”. Pressing the “Don’t ask” button made both buttons and the advisor question disappear and the true and false labels in the bottom corner of the screen were displayed. Pressing the “Ask” button additionally led to a picture being displayed below the statement (identical to Experiments 1 and 2) that contained the advice, which was always 100% valid. The rest of the experiment was identical to Experiment 1.

Results

Latencies to ask for advice

First, we consider participants' average time to decide about advice. This is the main latency DV here, as the overall decision time is confounded with the advice decision. We analyzed participants average response latency as a function of incentives (high vs. no) and repetition (repeated vs. new), with repeated measures on the first factor. The *F*-values are again based on analyses of the log-transformed latencies; the mean values are reported in untransformed milliseconds. We only report significant ANOVA effects at alpha = .05. Overall, participants made advice decisions slower in the high incentives condition (*M* = 5180, *SD* = 1852) than participants in the no incentive condition (*M* = 4416, *SD* = 1559), $F(1, 198) = 7.66$, $p = .006$, $\eta_p^2 = .037$, 95% CI [.006, .089]. Next, repetition again had the expected processing facilitation effect (Feustel et al., 1983). Participants made advice decisions faster for repeated statements (*M* = 4611, *SD* = 1701) than for new statements (*M* = 5000, *SD* = 1911), $F(1, 198) = 84.86$, $p < .001$, $\eta_p^2 = .300$, 95% CI [.214, .378]. Repetition and incentives did not interact, $F(1, 198) = 0.08$, $p = .781$, $\eta_p^2 = .000$, 95% CI [.000, .015].

Latencies to make "true-false" judgments

To partially unconfound the time to make true-false judgments from the time to decide about advice, we analyzed the judgment latencies separately for statements for which participants asked for advice and statements they did not ask for advice. We again used the log-transformed latencies and analyzed them with the same ANOVA as the latencies to decide about advice.[2]

Latencies given advice asked. Ten participants had never asked for advice and are excluded from the following analysis. On average, participants took 1724 ms (*SD* = 1114) to judge a statement's truth after asking for advice. Five additional participants had missing values for either repeated or new statements. Neither incentives nor repetition influenced these latencies, largest $F(1, 184) = 0.53$, $p = .468$. Thus, both the repetition and incentive effects on latencies are captured in the decision to ask for advice.

Latencies given advice not asked

Seven participants always asked for advice and are excluded from the following analysis. On average, participants took 1612 ms ($SD = 4083$) to judge a statement's truth after deciding not to ask for advice. Given participants did not ask for advice, participants judged statements slower in the high incentives condition ($M = 2072$, $SD = 5612$) than participants in the no incentive condition ($M = 1124$, $SD = 784$), $F(1, 191) = 13.08$, $p < .001$, $\eta_p^2 = .064$, 95% CI [.019, .126]. No other effects were significant, largest $F(1, 191) = 0.42$, $p = .519$. Thus, given that participants did not take the advice option, they spent more time considering their judgment in the high compared to the no incentive condition.

Advice Decisions

The main DV was participants' average tendency to ask for advice for repeated and new statements. We calculated this tendency as the probability to ask for advice given repeated and new statements, and analyzed them using an incentives (high vs. no) by repetition (repeated vs. new) ANOVA, with repeated measures on the second factor.

Most importantly, as predicted, participants were less likely to ask for advice given repeated statements ($M = .466$, $SD = .288$) compared to new statements ($M = .520$, $SD = .295$), $F(1, 198) = 32.35$, $p < .001$, $\eta_p^2 = .140$, 95% CI [.073, .215]. Interestingly, this repetition difference was numerically smaller in the high incentives condition ($M_{old} = .440.$, $SD_{old} = .309$ vs. $M_{new} = .477.$, $SD_{new} = .321$) compared to the no incentives condition ($M_{old} = .494.$, $SD_{old} = .265$ vs. $M_{new} = .564.$, $SD_{new} = .261$), although the interaction did not reach conventional levels of significance, $F(1, 198) = 3.22$, $p = .074$, $\eta_p^2 = .016$, 95% CI [.000, .056].

In addition, people in the high incentives condition numerically asked for less advice (*M* = .458, *SD* = .307) compared to the no incentives condition (*M* = .529, *SD* = .255), although this effect was also not significant, $F(1, 198) = 3.12$, $p = .074$, $\eta_p^2 = .016$, 95% CI [.000, .055]. This might follow because participants have the costs of the advice present when making their advice decisions.

Truth Judgments

Similar to the latencies, we analyzed the decisions dependent on whether participants asked advice or not.

Judgments given advice asked. Again, ten participants never asked for advice; they were not included in the present analysis. Interestingly, after asking for advice, participants almost perfectly followed the advice and judged statements as true given "true" advice (M = .975, SD = .098) and false given "false" advice (M = .025, SD = .070). These extreme probabilities violated the assumptions for an ANOVA analysis as there is almost no variation. Contrarily, repetition played no longer a role; participants had the same probability to responded "true" for repeated (*M* = .504, *SD* = .113) as for new statements (*M* = .497, *SD* = .103), $F(1, 184) = 1.32$, $p = .253$, $\eta_p^2 = .007$, 95% CI [.000, .040]. No other effect was significant, largest $F = 1.32$.

Judgments given advice not asked. Again, seven participants always asked for advice and did not enter the present analysis. As the advice asking data showed, participants asked less for advice for repeated information. We hypothesized this pattern assuming participants do not ask for advice because they believe they know that repeated statements are true. This assumption is confirmed by participants' overall high PTJs (*M* = .648, *SD* = .175) for statements, for which they do not ask for advice, which is substantially higher compared to all other values. However, even within this elevated value, participants still had higher PTJs for repeated (*M* = .673, *SD* = .201) compared to new statements (*M* = .607, *SD* = .206), *F*(1,

191) = 16.73, $p < .001$, $\eta_p^2 = .081$, 95% CI [.029, .147]. No other effect was significant, largest $F = 1.64$.

Discussion

Experiment 3 changed the availability of advice compared to the previous two experiments. Rather than advice being readily available for each statement and deciding whether to rely on it for each statement, participants had to decide whether they wanted to receive advice in the first place. Consequently, it makes sense to investigate statements with requested advice and statements without separately. Importantly, Experiment 3 managed to reduce the truth effect: Given that participants request advice, repetition no longer plays a role for the probability of a "true" judgment. This means that once participants choose to receive advice (a costly option for half the participants), they also consequently follow that advice. However, participants ask for advice less often for repeated statements, suggesting that repetition induced an illusion of knowledge: People believed repeated information to be true more often, and this belief was strong enough to risk 20 cents in following this belief.

General Discussion

People believe repeated information more than new information. This repetition-induced truth effect has a pervasive influence on people's judgments of truth. In a world where both true and false information is presented repeatedly more often than not, this effect might have serious consequences. However, research on the truth effect has so far not investigated truth judgments with consequences. Here, we investigated the influence of such consequences in the form of incentives. To avoid the problem that people might not know what is factually true or false, we also manipulated whether people received 100% valid advice about a given statement's factual truth status. To summarize the results of the first two experiments most poignantly, on average, repetition still significantly increased participants likelihood to judge a given statement as “true” when they were given 100% valid advice that a given statement is true and when the incorrect judgment incurred costs of 20 cents. In other words, people believed false information despite knowing better under high incentives due to the mere repetition of this information.

Experiment 3 went one step further and investigated a situation in which advice about factual truth was not always provided but needed to be actively sought. This experiment thereby reflects a more ecologically valid setting, namely the situation when people encounter information and decided to believe it or not, or decide to search for more information. Given repeated information, people were less likely to request advice, and given they did not request advice, they had markedly increased levels of “true” judgments, indicating people were more likely to forgo advice when they believed a statement to be true. However, if people decided to look for advice, that is, when they were really unsure, they followed the advice almost unanimously.

A possible explanation

Most explanations of the truth effect (Bacon, 1979, Begg et al., 1992; Reber & Schwarz, 1999; Unkelbach & Rom, 2017) refer to a processing experience that is differentially elicited by processing repeated compared to new information, termed remembering (Mandler, 1980), recognition, familiarity, or fluency. Unkelbach and Greifeneder (2013) suggested a Brunswikian lens model framework (Brunswik, 1952; Hammond, 1955) for judgments based on such processing experiences. Accordingly, truth is a distal criterion that people cannot assess directly. Rather, they need to rely on proximal cues at the time of judgments, and these cues are integrated in e linear fashion (Karelaia & Hogarth, 2008). In the present case, we provided an informational cue (i.e., advice) that varied in validity and direction, and an experiential cue (i.e., repetition), that only varied as being present or absent. In a lens model, the cue utilization does not need to match the factual cue validity. Thus, as both types of cues are always present, in a probabilistic fashion, the experiential cue will sometimes dominate the informational cue. In other words, just because the statement "feels" true, people will ignore the 100% valid informational advice that it is false. Importantly, this is not a dual-process model; there is no need to assume that people sometimes judge truth "rationally" and follow the given 100% valid advice, and sometimes they judge truth "intuitively" and follow their feeling due to repetition. Rather, people always use both cues in a probabilistic fashion, and as the experiential cue (i.e., repetition) only varies between present and absent, its influence is also comparable for both types of advice, 100% valid or 50% valid. We believe that such a lens model approach to judging truth provides a clear framework to study the important topic of judging truth in the future.

In addition, people are typically justified to use repetition as a cue for truth in the environment, as long as true information is more likely to be repeated than false information (see Reber & Unkelbach, 2010). In other words, people might have learned that using repetition as a cue to truth leads to more correct decisions than not. There is data, though, showing that in online contexts false information spreads faster than factually true information (Vosoughi et al., 2018). Nevertheless, overall, factually true information should be repeated more often than false information; this follows because there are many ways in which information can be false, but there is typically only one way to be correct (see Alves et al., 2017; Unkelbach et al., 2020). This asymmetry should overall lead to a prevalence of repeated true information (Unkelbach, Koch, & Alves, 2019), as long as there is no bias to strategically repeat the same false information.

Implications

Together, the three experiments underscore the relevance of the truth effect to understand many current irrational beliefs. In short, simply because information is repeated, some people will come to believe it, and as our relevance data suggest, the effect might even be amplified for information that is relevant to and interesting for people. Thus, one might understand beliefs in seemingly irrational statements, fake news, conspiracy theories or even statements that are directly contradicted by 100% valid information (i.e., "This was the largest audience to ever witness an inauguration" – Sean Spicer, Jan 22nd, 2017) as a probabilistic error that information feels true although one knows that it is not.

One might find it disheartening that this probabilistic error persisted even when people judged statements that were irrelevant for them and even when wrong judgments had monetary consequences. In other words, one might not downplay the effect with the notion that it will be eliminated or reduced when it really counts. While we observed some reduction, it was overall neglectable. The positive aspect from our data come from Experiment 3. When people decided to ask for advice, or translated to a typical setting, check information using a reliable source, they followed the advice. Nevertheless, the data clearly showed a bias to ask less frequently for advice for repeated information, just because the information was already repeated.

Finally, we must concede that we probably underestimate the truth effect in the present design, as we informed people that repetition is not a valid cue for truth (i.e., the presented information was half false). Jalbert et al. (2020) showed that the truth effect is substantially larger when this instruction is omitted. However, we included this instruction deliberately to avoid any doubt that the use of repetition might be seen as justified; for example, because participants may believe that the prior presentation is meant by the experimenters to indicate truth. This way, we believe we provided the most compelling evidence that the truth effect persists when judgments of truth have consequences; here in the form of monetary gains and losses.

Conclusions

We found that repetition influenced judgments of truth even when the judgments had consequences for participants, when valid advice about factual truth was given, and when repetition was clearly labeled as non-indicative of truth. Further, repetition decreased the likelihood to ask for advice, even when the judgment had consequences. We believe that such seemingly irrational behavior is best explained within the framework of a probabilistic Brunswikian lens model, in which repetition represents an experiential cue that across people and across judgments leads to a seemingly irrational behavior to judge repeated false information as true despite knowing better and despite monetary consequences for the self.

Footnotes

1. Please note that the pre-registration for this experiment states that we use 120 statements. However, given the high monetary costs of the experiment, we decided to reduce the number of statements in the judgment phase to 80, which saves up to 4 Euro per participant in the high incentive condition.

2. There was an apparent outlier in the data analysis for these decision latencies, probably due to a person looking up answers on a mobile device. However, we had no other basis for excluding this person and inclusion/exclusion did not affect any significance decisions, but only the size of the standard deviations.

3. The analysis is based on 186 participants, because four further participants had missing values on one of the DVs.

Appendix

TRUE-RELEVANT

1. People from Europe lose more hair per day than people from Africa.

2. Genetically, chimpanzees are most closely related to humans.

3. Between 1980 and 1990, the birth rate in East Germany was higher than in West Germany.

4. The heart of an average woman beats faster than that of an average man.

5. Babies have more bones than adults.

6. Jaw muscles are the strongest muscles in the human body.

7. The first experimental heart transplant took place in Vienna.

8.When you perceive different smells in both nostrils, you perceive only one of the two smells.

9. The Bundestag 2018 has relatively fewer female members than the Bundestag 2014.

10. The Federal Intelligence Service used to be called the "Federal Property Administration" for camouflage purposes.

11. Sinti and Roma originally come from India and Pakistan.

12. The Federal Republic's coat of arms dates back to the Roman Empire.

13. The Federal Cross of Merit can be revoked after it has been awarded.

14. The melody of the German national anthem was composed by Joseph Haydn.

15. The largest lithium deposits in the world are in bolivia.

16. Apart from humans, gorillas have only leopards as natural enemies.

17. In germany, the daily limit for fine dust is 50 micrograms per cubic meter of air.

18. The price of gas is not contractually linked to the price of oil.

19. Germany has the highest CO_2 emissions per capita in the EU.

20. Elvis and Priscilla Presley had a biological child.

21. Elon Musk has been married three times.

22. Melania Trump was born in the Slovenian town of Novo Mesto.

23. Actor Stan Laurel died of a heart attack.

24. Brad Pitt is banned from entering China.

25. Bond actor Pierce Brosnan worked as a fire-eater for three years.

26. Rolls-Royce uses only bull hides for its leather seats.

27. Nelson Mandela boxed as a hobby.

28. Freddy Mercury was born in Zanzibar.

29. The European parliament has two meeting places.

30. cats can be allergic to humans.

TRUE-IRRELVANT

1) The painting "Au Moulin de la Galette" is by Renoir.

2. Alberto Fujimori was president of Japan from 1990 to 2000.

3. Early scholasticism was a current of the Middle Ages.

4. The play "A Doll's Home" was written by Henrik Ibsen.

5. The Lorentz force acts on electric charges in magnetic fields.

6. The Ising model of theoretical physics describes ferromagnetism.

7. Fbrinogen is a protein produced in the liver.

8. "The Sacrifice of Spring" is a work by Igor Stravinsky.

9. Urania is the epithet of the goddess Aphrodite.

10. A wobbler is a lure used for fishing.

11. The second Punic war was decided in the battle of Zama.

12. The sesame plant has thimble-like leaves.

13. Inorganic pigments are also called mineral colors.

14. First Olympic champion in modern times was James Connolly.

15. The operating depth of a submarine is 100 meters.

16. Europe's largest glacier is Vatnajökull in Iceland.

17. In harmonic oscillations, the frequency is independent of the amplitude.

18. The mass of a body is not equal to its weight.

19. Lucius Superbus was the last Roman king.

20. Semiotics is the general study of sign processes.

21. An oxymoron is a phrase with an internal contradiction.

22. The battle of trafalgar was part of the third coalition war.

23. A panarello is a plastic attachment for the steam nozzles on espresso machines.

24. Syntagma is the antonym of paradigm.

25. Herbert Blumer was the founder of the school of symbolic interactionism.

26. Cloven-hoofed game is the term used in hunting law for cloven-hoofed animals.

27. The motto of the Indian state is "truth alone triumphs".

28. Purse seines are nets used in deep-sea fishing.

29. The largest cemetery in the world is located in Bahrain.

30. Lubaantun is a Mayan ruin in Belize.

FALSE-RELEVANT

1. Aspirin rarely (1 case in 1,000 to 10,000) causes microbleeds in the gastrointestinal tract.

2. People with blood group AB+ can be blood donors for all groups A and B.

3. In heterosexual people, the number of new HIV infections has been stable since 2010.

4. The left lung is bigger than the right lung.

5. The European Union's slogan "United in diversity" is based on an internet survey of students.

6. Women were allowed to vote for the first time in France in 1919.

7. In the event of defense, the Minister of Defense commands the German Armed Forces.

8. The form of government in the United Arab Emirates is a constitutional monarchy.

9. With the Good Friday Agreement, Ireland reaffirmed its demand for reunification with Northern Ireland.

10. The Federal Ministry of the Interior, for Construction and Home Affairs is also responsible for tourism in Germany.

11. In the 2017 federal election, Hamburg had the smallest number of constituencies of all the federal states.

12. More men than women live in Germany.

13. UN missions abroad by the Bundeswehr are not covered by the Basic Law.

14. The European state with the most inhabitants/km² is the Vatican.

15. Russia produces the most aluminum worldwide.

16. Venezuela has less oil reserves than Norway.

17. Rapeseed oil damages groundwater.

18. Solar power is more expensive than coal-fired power in Italy, Spain and Portugal.

19. The USA uses more plastic per capita than any other country.

20. The first world climate conference was held in Geneva in 1990.

21. All EU countries together produce more primary energy than the USA.

22. Michael J. Fox has multiple sclerosis.

23. The Spanish King Felipe VI is descended from the Merovingian dynasty.

24. According to his own statement, Andy Warhol was bisexual.

25. Rachel Markle became Duchess of Essex by marrying Prince Harry.

26. Hugh Grant was convicted twice for having sex in public.

27. Actor Bruce Willis was born in Ems, a town in Rhineland-Palatinate.

28. Will Smith's full name is William Carroll Smith Jr.

29. People need a permanent residence to be able to vote in federal elections.

30. Tsunami means beach wave.

FALSE-IRRELVANT

1. The Greek creation gods were Gaia, Erebos and Creon.

2. A chiasmus is the juxtaposition of contradictory statements.

3. Pentameter is the classical meter of epic poetry.

4. Cacti can reproduce by parthenogenesis.

5. The speed of sound is independent of temperature.

6. The capital of Madagascar is Toamasina.

7. Kinetic pressure is measured in millimeters of water column.

8. The unit of measurement for magnetic flux is a "Henry".

9. Tubular fungi belong to the protist class.

10. Swedish Dag Hammarskjöld was the first UN secretary general.

11. All states of Oceania have only water borders.

12. Viscosity describes the molecular concentration of a substance.

13. Observation of the solar photosphere is possible only with coronographs.

14. Ashi waza refers to a class of kicks in karate.

15. Qín Shi Huángdì was the last Chinese emperor.

16. Coconut islands belong to Indonesia.

17. Lama temple is located in shanghai.

18. The Barringer crater is on the northern moon hemisphere.

19. An emphasis by comparison with something impossible is called aberation.

20. Liberia consists of the territories Tripolitania, Cyrenaika and Fezzan.

21. "At the Rendezvous of Friends" is a work by August Macke.

22. Ghasel is a song form that originated in the North African region.

23. The Castle-Taubig thesis makes statements about the capabilities of a calculating machine.

24. Valparaiso is one of the largest cities in Angola.

25. The conflict over the Falkland Islands took place between Argentina and France.

26. Every odd number greater than 2 can be written as the sum of two prime numbers.

27. The four-color theorem cannot be proven mathematically.

28. "Cum hoc ergo propter hoc" is a figure of syllogistic reasoning.

29. "Parema lapsis" are a class of parasitic microorganisms.

30. Henbane was a popular spice in the Middle Ages.

5 The Truth Effect With Personally Highly Relevant COVID-19 Related Statements

Taken together, the previous three chapters suggest that incentives are not an effective tool to reduce the Moses illusion or the truth effect (see discussion in Chapter 6). However, monetary incentives are not the only way to add consequences to participants' decisions to increase their motivation. As discussed previously, many truth effect experiments use trivia statements that are obscure. Furthermore, these statements often have no relevance for participants' real lives and thus their decisions remain inconsequential. For example, when presented with the statement "The Saturn moon Mimas has more spin than the Saturn moon Pallene" (Unkelbach & Rom, 2017), believing this statement to be true or false is unlikely to have any impact on one's daily life. On the other hand, the statement "The world's most poisonous snake is the Australian Inland Taipan" (Unkelbach & Rom, 2017) could be important for people living in Australia or planning to travel there. In this case, believing or rejecting a statement could have real-life consequences. For example, if one believes the statement, it would be sensible to obtain further information about the Inland Taipan, what it looks like, and so on. Unlike monetary incentives, these consequences are indirect and long-term, but also potentially more dire: The most poisonous snake, for instance, is potentially life-threatening.

Although Chapter 4.1 introduced statements with higher relevance than usual (Experiment 2), they showed a higher truth effect than statements with low relevance. This could be explained by the referential theory (Unkelbach & Rom, 2017): Relevant statements have probably been heard more often in real life, meaning they contain more reference nodes with already instigated links, thus increasing their perceived truth. Furthermore, the magnitude of personal relevance of the statements was still rather low. Thus, in the following paper, we investigated whether using statements related to the COVID-19 pandemic diminished the truth effect and manipulated experimentally if the statements were well-known or lesser-known, allowing for disentangling how relevant statements are and how well-known they are. These statements were highly relevant as a large part of public attention was focused on COVID-19 (e.g., Ruhrmann & Daube, 2021), and they also had potentially dire long-term consequences. For example, believing or rejecting that SARS-COV-2 can survive on cardboard could change one's behavior and potentially carry serious health risks in case of an infection. Chapter 5.1 is based on the following article:

Unkelbach, C., & Speckmann, F. (2021). Mere repetition increases belief in factually true COVID-19-related information. *Journal of Applied Research in Memory and Cognition*, 10(2), 241–247. https://doi.org/10.1016/j.jarmac.2021.02.001

Please note that some changes in citation style and formatting were undertaken to keep the layout of this dissertation consistent. No changes were made to the content of the article.

5.1 Mere Repetition Increases Belief in Factually True COVID-19-Related Information

Abstract

Repetition increases people's belief that the repeated information is true. Previous research has investigated this increase with largely unknown trivia information (both factually true and false), and more recently, with a focus on factually false information (i.e., "fake news"). We investigate whether this increase in belief also holds for relevant and true information, concretely, information related to the CoViD-19 pandemic. In two experiments (total N = 398), we manipulated repetition for well-known and less-known information about the ongoing pandemic. Overall, repetition increased participants' belief in Corona-related information. However, the increase was significantly larger for less-known compared to well-known statements. This implies an asymmetry for repeating true and false information: Because false information is more likely unknown compared to true information, repetition may benefit false information substantially more compared to true information. Nevertheless, mere repetition increases belief for actually relevant and factually true information, providing a basis for subsequent actions.

Keywords: repetition-induced truth effect, illusory truth effect, fluency, Corona pandemic, CoViD-19, SARS-CoV-2

General Audience Summary

The „truth effect“ is the phenomenon that mere repetition increases belief in the repeated information, both relative to its initial presentation and relative to other, non-repeated information. This truth effect is currently recruited to explain how people come to believe apparently false information, which is a prominent topic in an age of "alternative facts" and "fake news". We investigated if mere repetition may also increase beliefs in factually true and actually relevant information, namely information related to the ongoing CoViD-19 pandemic. Such an increase would suggest a positive implication of the truth effect. Two experiments showed that overall, mere repetition indeed increased participants' belief in true information related to the ongoing pandemic. However, this increase was significantly larger for less-known information; for well-known information, the increases were small. This suggests an asymmetry for the truth effect in the real world: as false information is more likely to be unknown or less known to people (e.g., novel conspiracy theories or novel false claims) compared to true information, the truth effect benefits false information more than true information. Nevertheless, overall, merely repeating information increased participants' belief in the repeated information. Repetition thus may serve as a tool to increase belief in relevant true information related to the ongoing pandemic.

Mere repetition increases belief in factually true CoViD-19-related information

Repetition increases people's belief that the repeated information is true, both relative to its initial presentation and relative to other, non-repeated information. Since the initial demonstration by Hasher et al. (1977), this truth effect or repetition-induced truth effect has become a robust classic (cf. the quantitative meta-analysis by Dechêne et al. 2010). Over the last years, research on this truth effect has increased substantially (see Brashier & Marsh, 2020; Unkelbach et al., 2019). This increase partially follows from the effect's potential to explain how people come to believe apparently false information, which is a prominent topic in an age of "alternative facts" and "fake news" (e.g., Corneille et al., 2020; Pennycook et al., 2018). In addition, the effect might be substantially stronger in reality compared to its laboratory variants (Jalbert et al., 2020). Despite its potential for misinformation and creating false beliefs, we address whether the truth effect may also have positive implications; that is, whether it is possible to increase beliefs in factually true information related to the ongoing CoViD-19 pandemic. As belief is usually a predictor for actions (e.g., Ajzen, 1991), mere repetition may also be a tool for positive influences on behavior in the ongoing pandemic.

In the following, we provide a brief overview of research on the repetition-induced truth effect and delineate why a test of this hypothesis is interesting from a practical and theoretical point of view. Then, we report two experiments that test the truth effect for factually true statements related to the ongoing CoViD-19 pandemic, and based on the data, discuss the implications for theories and applications of the repetition-induced truth effect.

Explanations for the repetition-induced truth effect

The research following the seminal work by Hasher et al. (1977) focused on the underlying mechanisms of the effect. Hasher and colleagues argued that frequency of occurrence directly confers validity to the repeated statements; however, this explanation was refuted by Bacon (1979), who showed that not the objective frequency of occurrence, but the subjective recognition experience predicts increases in rated truth. Building on the potential experiential nature of the effect, Arkes et al. (1991) argued that repetition increases belief in repeated information because it *feels* more familiar compared to novel information. Begg et al. (1992) then showed the different contributions of feelings of familiarity and explicit recollection, thereby providing further evidence for an experiential basis of the effect. This experiential basis was directly tested by Reber and Schwarz (1999), who hypothesized processing fluency as an underlying construct, that is, the experienced ease of ongoing conceptual or perceptual processes (Unkelbach & Greifeneder, 2013). They manipulated processing fluency without repetition via the color contrast of the presented information, which contributed to increases in rated truth without repetition. Combining these explanations, Unkelbach and Rom (2017) presented a referential theory of the effect and argued that the initial presentation of the information links previously unlinked references in memory, thereby creating coherent networks in memory, which should increase recognition, familiarity, and processing fluency (see also Brown & Nix, 1996).

For studies focusing on the underlying mechanism, the content of the repeated information was, for the most part, of little interest; most studies included both true and false information, and this variable rarely had an influence on the results (see Unkelbach & Stahl, 2009). Only more recently came the content into focus, with observations that knowledge about the repeated information does not change the increases in rated truth

(Fazio et al., 2015) and that repetition also increases participants' beliefs in clearly fake news that were even labelled to be disputed by fact checkers (Pennycook et al., 2018), and even for information that was clearly labelled as false, both at encoding and at judgment (e.g., Unkelbach & Greifeneder, 2018).

Building on this shift to the content of the repeated information, we addressed the question if mere repetition could also have positive effects; that is, does mere repetition also increase subjective truth of factually true information related to the ongoing CoViD-19 pandemic? From an applied perspective, this could be a positive side effect of the truth effect, as increasing beliefs in factually true information is most likely a beneficial outcome, potentially leading to more compliance with and acceptance of measures to fight the pandemic (e.g., Ajzen, 1991).

Repeating actually relevant and factually true information

The potential negative implications of increased truth of false information due to mere repetition and potential remedies are by now well documented (e.g., Bago et al., 2020; Brashier et al., 2020). However, we are not aware of studies that investigated potential increases in subjective truth for actually relevant and factually true information. This is potentially due to three reasons. First, from a measurement perspective, rating effects are best investigated with items that have an *a priori* rating in the middle of the scale. Thus, the scale is open to both sides for increases and decreases in ratings. Relevant and true information most likely deviates from this ideal starting point. Second, materials that are relevant and true might introduce confounds that threaten conclusions about underlying processes. And third, with the exception of the original frequency-validity explanation, the presented explanations would predict reduced repetition effects for true and relevant information, which would prevent theoretical distinctions. While the two former points are immediately clear, the latter one might need some explanations.

In a typical truth effect experiment, participants do not know the factual state of the world that is described by the statements (e.g., “The Hawaiian alphabet features fewer letters than the German alphabet.”). Unkelbach and Rom (2017) argued that processing the initially unknown statement links previously unlinked references in memory. References are memory representations that provide meaning for a statement’s elements. Thus, if statements are meaningful for participants (see Hasher et al., 1977; Arkes et al., 1989), the references form a coherent network, which will foster fluent processing, feelings of familiarity, and recognition at later points in time (see also Brown & Nix, 1996). However, when the information describes established states of the world (e.g., “The world is round.”), the initial presentation cannot add new references, which leaves pure perceptual processes to influence subjective truth (Feustel et al., 1983; Johnston et al., 1985).

The same prediction follows from a fluency perspective (Reber & Schwarz, 1999). As analyzed by Reber and Unkelbach (2010), people should be more likely to encounter true information repeatedly compared to false information; fluency due to repetition is thereby a valid cue. This follows from a cooperative communication maxim (Grice, 1975) and from the fact that there are many ways for information to be false (e.g., “The earth is flat”, “The earth is donut-shaped”), while true information can only be truthful in one way (i.e., “The earth is round”; Unkelbach, Koch, & Alves, 2019).[1] If people are more frequently exposed to the same true information compared to the many possible variants of false information, one needs an additional but highly plausible assumption. Repetition increases processing fluency not in a linear, but in a negatively accelerated fashion. Thus, the first repetition has more influence than the second and so forth. This pattern is already evident in the original data by Hasher et al. (1977). The same argument also applies if one substitutes fluency with familiarity or recognition as the underlying causes.

Thus, from virtually all theoretical perspectives, factually true and relevant, that is, currently frequently repeated information, may not benefit from repetition the same way that novel information does.

The present research

As delineated, investigating the influence of repetition on truth for factually true information introduces some uncertainty for the interpretation of underlying processes. Thus, we confined ourselves here to a functional perspective without strong implications for underlying cognitive processes (see De Houwer, 2011). We will test if repetition increases truth ratings of factually true CoViD-19 related information; if one may increase the subjective truth of such information by mere repetition, this would indicate a positive application of the repetition-induced truth effect. Given the present research focus on misinformation (e.g., Pennycook et al., 2018) and observed spread advantages of false information in online communications (e.g., Vosoughi et al., 2018), it is an open question if the same mechanisms that foster belief in false information might also be used beneficially.

We report two experiments that test the truth effect for well-known CoViD-19 information and less-known CoViD-19 information. The sole difference between experiments is that we made the distinction of “well-known” and “less-known” ad hoc in Experiment 1, but empirically validated this distinction in Experiment 2. Both experiments were pre-registered. The pre-registrations, statements, data files, analysis scripts, and the materials can be found online at https://osf.io/hvj8s/?view_only=4ea0be8f4e9e4b2bafa8921c804e50c7. These are the only two experiments we conducted so far in this line of research.

All information was factually true to the best of our knowledge at the time of the experiments. Although it would have been interesting to include false statements about the COVID-19 pandemic, this option seemed unethical given the potential for introducing false beliefs.

Both experiments included as a between-participants manipulation whether the statements used a "precise" labelling or a "colloquial" labelling. The precise condition used the terms "CoViD-19" and "SARS-CoV-2", while the colloquial condition used the terms "Corona disease" and "Corona virus", respectively. We did not have specific hypotheses for this variation (see pre-registrations) and this factor did not influence any of the results (i.e., no main effects or interactions). The factor is included in all analyses, but we omit this manipulation from the report and the results for clarity and brevity.

Experiment 1

Method

Participants and Design. We recruited 198 participants from Amazon's Mechanical Turk platform on April 9th, 2020. They received $ 1.50 for participating. At this date, there were 458.623 confirmed CoViD-19 cases and 16.954 deaths related to CoViD-19 in the US (https://covid.cdc.gov/). We manipulated repetition (repeated vs. novel) and statement set (well-known vs. less-known) within participants. Please note that we use the term "novel" for information that appears for the first time within a given experimental session.

Materials. We collected statements about CoViD-19 disease and the SARS-CoV-2 virus from the homepages of the European Union (https://europa.eu/european-union/coronavirus-response_en), the American Center for Disease Control (https://www.cdc.gov/coronavirus/2019-nCoV/index.html), the World Health Organization (https://www.who.int/emergencies/diseases/novel-coronavirus-2019), and Wikipedia (https://en.wikipedia.org/wiki/COVID-19_pandemic). We collected 16 statements for which we assumed that they are well-known, due to their wide-spread coverage (e.g., "Contact via some contaminated surfaces can lead to infection with the Corona virus."), and 16 statements for which we assumed that they are less-known, due to their less frequent appearance (e.g., "Hypertension is a predictor for intensive care unit admission after a Corona virus infection."). Both statement sets were of comparable length (171 vs. 175 words, respectively). However, we did not explicitly control for other potential differences between the sets such as concreteness or specificity. The statements are available online at the OSF (see above).

Procedure. Participants first read a consent form that complied with European data storage regulations. If they agreed to participate, the experiment instructed them that they would first see a list of statements about CoViD-19 disease and SARS-CoV2 and then rate a list of statements in the second part for believability. Once they clicked a "continue" button, the presentation phase began. The survey program randomly picked eight well-known statements and eight less-known statements. Each statement was shown for two seconds and participants passively observed the statements. Participants had to click a button to continue to the next statement. Next, the evaluation phase began. The specific instructions stated: "For the next task, please rate the following statements on how believable they are. "Then, participants rated the 16 repeated and the 16 novel statements on a 7-point Likert scale with the end points labelled "Not very believable" and "Very believable". Presentation order was randomized. Upon finishing, participants answered demographic questions and indicated their level of concentration during the experiment.

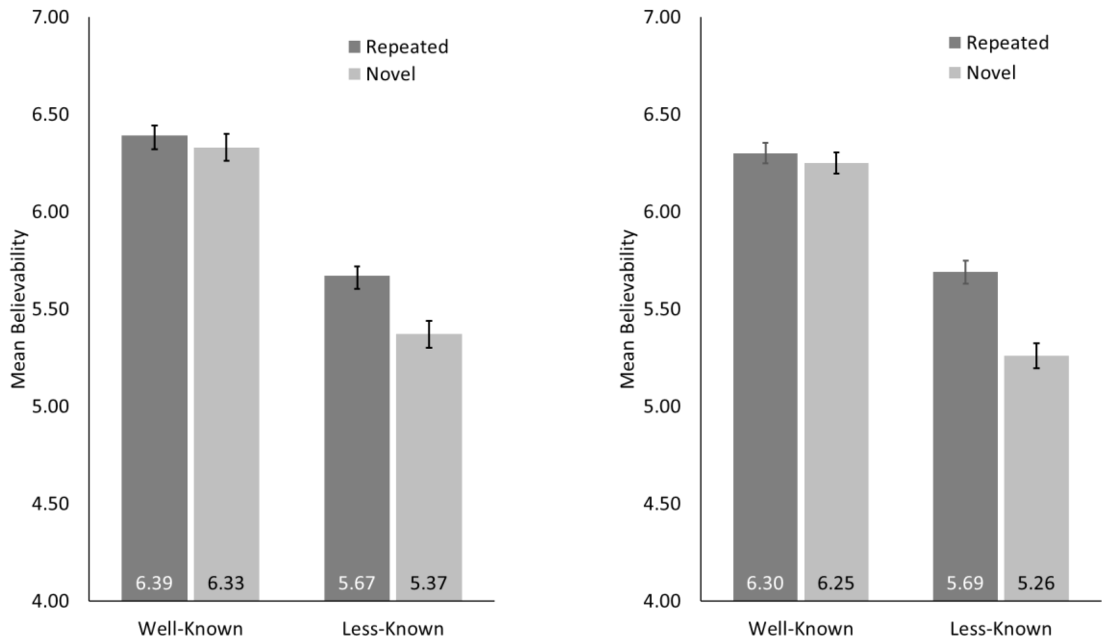
Results

Figure 5.1's left panel shows participants' mean believability ratings of repeated and new statements as a function of well-known and less-known statements. We analyzed these means with a repetition (repeated vs. novel) by statement set (well-known vs. less-known) repeated-measures ANOVA. As Figure 5.1 suggests, we replicated the typical truth effect for information related to the CoViD-19 pandemic. Overall, participants rated repeated statements as more believable (M = 6.03, SD = 0.93) compared to novel statements (M = 5.85, SD = 0.96), $F(1, 196) = 27.00$, $p < .001$, $\eta_p^2 = .121$, 95%CI[.058, .193].

This main effect was qualified by an interaction, $F(1, 196) = 15.79$, $p < .001$, $\eta_p^2 = .075$, 95%CI[.026, .139]. The truth effect was significantly smaller for well-known compared to less-known statements. Within well-known statements, there was no significant difference between repeated (M = 6.39, SD = 0.72) and novel statements (M = 6.33, SD = 0.69), $F(1, 196) = 2.42$, $p = .122$, $\eta_p^2 = .012$, 95%CI[.000, .049]. Within less-known statements, the difference between repeated (M = 5.67, SD = 0.98) and novel statements (M = 5.37, SD = 0.69) was highly significant, $F(1, 196) = 34.79$, $p < .001$, $\eta_p^2 = .151$, 95%CI[.081, .226].

Figure 5.1

Participants mean believability ratings of well-known and less well-known statements related to the CoViD-19 pandemic as a function of repetition (i.e., old vs. new in the experimental setting). The left panel shows the data from Experiment 1 (n = 198) and the right panel shows the data from Experiment 2 (n = 200). Higher values indicate higher believability; the scale range was from 1 to 7. The error bars show the standard error of the means.



In addition, participants rated the statements we had classified as well-known as more believable (*M* = 6.36, *SD* = 0.70) compared to less-known statements (*M* = 5.52, *SD* = 0.98), $F(1, 196) = 231.10$, $p < .001$, $\eta_p^2 = .541$, 95%CI[.465, .601]. As we only included true information, this main effect implies that less knowledge leads to less belief; that is, for the less-known statements, participants knew less often that the statements are factually true, and judged them accordingly.

Discussion

Mere repetition increased the believability of statements about SARS-CoV-2 and the resulting CoViD-19 disease. This effect was qualified by the presumed knowledge about the statements. For statements we a priori classified as well-known, participants showed only a non-significant increase in rated believability.

As discussed in the introduction, this interaction may follow from explanations of the truth effect based on processing fluency perspective as well as referential networks. However, it is also possible that mere ceiling effects prevented an empirical realization of the truth effect. Thus, we cannot make firm claims on the process level. On the functional level, though, the clear result is that for what we labelled well-known statements, there was only a non-significant increase in rated believability, while for less-known statements, there was a significant increase.

While the clear nature of the observed interaction lends credibility to our ad-hoc classification, it seems prudent to substantiate this factor empirically. Thus, Experiment 2 replicated Experiment 1, and participants assessed whether they knew the presented information or not with a binary response format. We expected to replicate Experiment 1’s results and obtain a higher frequency of participants “known” responses for the “well-known” statement set compared to the “less-known” set, thereby validating our ad-hoc classifications.

Experiment 2

Method

Participants, design, materials, and procedure. We recruited 200 participants from Amazon’s Mechanical Turk platform on April 20th, 2020. They received $1.90 for participating. At this date, there were 774.873 confirmed CoViD-19 cases and 41.756 deaths related to CoViD-19 in the US (https://covid.cdc.gov/). The sole difference to Experiment 1 besides time of data collection was that participants answered a binary "I knew this / I did not know this" choice for all presented statements after the believability rating phase. Specifically, after the truth evaluation phase, the program instructed participants: "For the next task, please indicate for each of the following statements whether you knew this before partaking in the study or you did not know this." All statements appeared in a new randomized order. Upon finishing this knowledge judgment phase, participants answered demographic questions and indicated their level of concentration during the experiment.

Results

Manipulation Check. Participants rated 87.38% of the "well-known" statement set as "known", and 56.31% of the "less-known" set as "known", χ^2 (1, N = 6393) = 1350.28, $p <$ .001, supporting our classification.

Truth effect analysis. Figure 5.1's right panel shows participants' truth evaluations in Experiment 2. As the Figure shows, we almost exactly replicated Experiment 1. First, we again found the typical truth effect. Participants rated repeated statements as more believable (M = 5.99, SD = 0.85) compared to novel statements (M = 5.75, SD = 0.98), F(1, 198) = 43.70, $p < .001$, $\eta_p^2 = .181$, 95%CI[.106, .258].

Second, this main effect was again qualified by an interaction, F(1, 198) = 31.71, $p <$ .001, $\eta_p^2 = .138$, 95%CI[.071, .212]. The truth effect was significantly smaller for well-known compared to less-known statements. Within well-known statements, there was again no significant difference between repeated (M = 6.30, SD = 0.75) and novel statements (M = 6.25, SD = 0.76), F(1, 198) = 1.74, $p = .189$, $\eta_p^2 = .009$, 95%CI[.000, .042]. Within less-known statements, the difference between repeated (*M* = 5.69, *SD* = 0.83) and novel statements (*M* = 5.26, *SD* = 0.92) was highly significant, $F(1, 198) = 55.70$, $p < .001$, $\eta_p^2 = .220$, 95%CI[.140, .298].

Third, participants rated the statements we classified as well-known as more believable (*M* = 6.28, *SD* = 0.75) compared to less-known statements (*M* = 5.47, *SD* = 0.90), $F(1, 198) = 280.42$, $p < .001$, $\eta_p^2 = .586$, 95%CI[.515, .641].

Truth and self-reported knowledge. Exploratorily, we also analyzed the believability ratings as a function of self-reported knowledge, repetition, and their interaction. This analysis mirrors the results of the analysis based on a priori classified statement sets. We needed to exclude nine participants from this analysis because they had missing values on one of the dependent variables. For the remaining participants, the typical truth effect on rated believability remained unchanged: Repeated (*M* = 5.84, *SD* = 1.10) and novel statements (*M* = 5.57, *SD* = 1.21) differed significantly, $F(1, 189) = 36.76$, $p < .001$, $\eta_p^2 = .163$, 95%CI[.090, .241].

This difference was again qualified by an interaction with self-reported knowledge, $F(1, 189) = 15.40$, $p < .001$, $\eta_p^2 = .075$, 95%CI[.026, .141]. The truth effect was significantly smaller for self-classified “known” compared to “unknown” statements. However, even within “known” statements, there was a significant difference between repeated (*M* = 6.43, *SD* = 0.75) and novel statements (*M* = 6.33, *SD* = 0.79), $F(1, 189) = 10.70$, $p = .001$, $\eta_p^2 = .053$, 95%CI[.013, .113]. Within self-classified “unknown” statements, the difference between repeated (*M* = 5.20, *SD* = 1.05) and novel statements (*M* = 4.78, *SD* = 1.03) was highly significant, $F(1, 189) = 29.50$, $p < .001$, $\eta_p^2 = .135$, 95%CI[.068, .210].

The difference between self-classified known statements (*M* = 6.42, *SD* = 0.75) compared to self-classified unknown statements (*M* = 4.99, *SD* = 1.06) was highly significant, $F(1, 189) = 391.77$, $p < .001$, $\eta_p^2 = .675$, 95%CI[.614, .719].

Discussion

Experiment 2 by and large replicated Experiment 1. We again observed a truth effect for information related to the CoViD-19 pandemic and an interaction with statement set: The truth effect was significantly smaller for well-known compared to less-known statements. In addition, participants' binary decisions supported our a priori classification of the statements. In addition to Experiment 1, based on participants' classification of the presented information, we also observed a significant increase in believability for well-known information. Thus, different from our ad-hoc classification, we found significant increases in rated believability both for self-rated "known" and "unknown" statements, although the increase was significantly larger for "unknown" statements.

The knowledge results also fit with our explanation for the difference in repetition effects between what we classified as well-known and less-known statements. If one assumes that people acquire knowledge by being repeatedly exposed to information, and that the effect of repetition on belief is negatively accelerated (i.e., the first repetition has a stronger effect compared to the second repetition, etc.), the interaction of repetition and knowledge (cf. Figure 5.1) directly follows. This argument implies an asymmetry for the repetition of true and false information. As true information has a higher chance of being repeated (cf. Reber & Unkelbach, 2010; Unkelbach et al., 2019b), the repetition-induced truth effect should be substantially stronger for false compared to true information.

General Discussion

We investigated whether one might increase the believability of factually true information related to the ongoing CoViD-19 pandemic by mere repetition; that is, is there a repetition-induced truth effect for relevant and factually true information? This would indicate a positive side of the truth effect, as increases in belief should lead to higher compliance with measures to fight the pandemic. For example, if people believe that washing their hands regularly and thoroughly prevents the spread of the virus, they should be more likely to follow such a behavioral recommendation (e.g., Ajzen, 1991). The answer to the question is differentiated. Overall, we found in both experiments a clear repetition-induced truth effect for actually relevant and factually true information. However, this effect was moderated by how well-known the information was. Based on our ad-hoc classification of well-known and less-known information, the truth effect was significantly smaller and no longer significant for well-known statements (Exp. 1 and 2). Based on participants' self-classifications, the truth effect was significantly smaller, but still significant for well-known information (Exp. 2). This reduction based on assumed knowledge structures mirrors results by Brashier et al. (2020), who found that focusing on accuracy at encoding eliminates the truth effect, but only for participants with the relevant knowledge. Similarly, Bago et al. (2020) found that deliberating about "fake news" reduces belief in such news only if participants had relevant knowledge. Finally, beyond the increase in subjective truth, repeating information keeps it also more accessible, which is a necessary pre-condition for information to influence behavior (e.g., Wyer, 2008).

On a functional level, it thus follows that one may harvest the impact of repetition on true information related to the CoViD-19 pandemic; yet, the increases for already well-known information are small and less reliable compared to the increases for less-known or unknown information. This may follow from the discussed underlying processes or because existing beliefs cannot increase further (i.e., ceiling effects). Independent of the underlying cause, the observed interactions emphasize an insidious aspect of the repetition-induced truth effect. Because repetition increases believability more for less-known or unknown compared to well-known information, and just because false information is more likely to be unknown, repeating true information may be less effective to increase beliefs in true information, while repeating false information may be a very effective strategy. Nevertheless, even if the information ecology seems saturated with information related to the CoViD-19 pandemic, there are still overall increases in believability due to mere repetition.

Footnotes

1. We are aware that the earth is not perfectly round in the geometrical sense.

6 General Discussion

Previous research investigating the moderators of cognitive illusions like the Moses illusion and the truth effect has often focused on task-related features, such as different phrasing of the statements/questions or different presentations (see Chapter 1). Person-related features were mainly investigated in the context of individual differences (e.g., Cantor & Marsh, 2017; De keersmaecker et al., 2020; Hannon & Daneman, 2001; Umanath et al., 2014). One relatively understudied potential person-related moderator is motivation. In Chapter 1, I argued that increased motivation through meaningful consequences could lead to higher performance by increasing effort. Consequently, monetary incentives could serve as a suitable manipulation of consequences that increase motivation, and in turn, effort. Chapters 2 and 3 directly tested this previously proposed incentives-effort-performance link (Bonner & Sprinkle, 2002).

Chapter 2.1 used a basic Moses illusion paradigm with three different incentive levels (high, medium, none) but found no substantial effect of incentive level on the number of Moses responses. However, participants did spend more time deliberating their response. This sign of increased effort indicates that the motivation manipulation was successful.

A similar pattern occurred in Chapter 3.1, which substituted the Moses illusion paradigm with a truth paradigm but kept the incentivized design. Again, incentives increased response times, but the influence on the number of “true” judgments was minor. Both chapters demonstrated the incentives-effort link but did not provide strong evidence for the effort-performance link. However, at least for the truth effect, this could be due to participants’ lack of knowledge regarding the statements. If fluency is the only cue available because there is no prior knowledge, then manipulations weakening the reliance on fluency will likely be unsuccessful.

Chapter 4.1 addressed this problem by providing participants with advice in addition to the previously used incentives for correct responses. So, even if participants lacked the required knowledge to judge the statements, they could rely on advice, especially if it is highly valid. However, while participants did understand the advice and relied more on highly valid advice, the truth effect still persisted in the face of high incentives. Concretely, high incentives did increase the probability that participants follow 100% valid advice, but they did not diminish the effect of repetition. Interestingly, the repetition effect was larger for statements supposed to be more relevant than typical truth paradigm statements. Furthermore, participants did not seek advice when presented with this option even with high incentives, suggesting that they prefer to rely on their intuition even when presented with a highly valid alternative for their decision.

Chapters 2-4 introduced direct consequences for participants by incentivizing correct responses with money. And while the potential bonus in the high incentives conditions was much larger than the typical compensation for in our lab, this type of consequence could have been too irrelevant and artificial to meaningfully reduce the illusions.

Since three chapters did not find compelling evidence for an attenuation of the Moses illusion or the truth effect through monetary incentives, Chapter 5 investigated a different manipulation of motivation.

Chapter 5.1 used COVID-19 related statements in a typical truth effect paradigm. Because they were related to an acute global pandemic, these statements were highly relevant and deciding whether to believe them or not could potentially have strong negative consequences for one's own health. Thus, the consequences for decisions in Chapter 5.1 were implicitly embedded into the statements. However, the truth effect did not show for well-known statements (possibly due to ceiling effects). Lesser-known statements yet again were rated as true more often when repeated. For ethical reasons and to avoid the spread of misinformation regarding COVID-19, we chose only factually true statements, but the results should be generalizable to factually false statements as well, in which case participants could suffer health consequences when mistakenly believing them.

6.1 The Relation of Effort and Performance

Despite adding meaningful consequences, both through monetary incentives or highly relevant statements, we could not attenuate the Moses illusion or the truth effect, providing further evidence to their general robustness. However, judging by the increased response times in Chapters 2 - 4 (cf. "effort duration", Bonner & Sprinkle, 2002, p.306), this was probably due to a disruption in the effort-performance link rather than the incentives-effort link. In other words, we likely managed to increase motivation but it seems that motivation is not an effective tool to attenuate cognitive illusions (cf. Enke et al., 2021). It may be that both the Moses illusion and the truth effect belong to a group of cognitive illusions that cannot be avoided through motivation alone.

But why would motivation have no effect on the Moses illusion? Consider again the semantic network and partial matching explanation for the Moses illusion from Chapter 1.2.1. When a participant reads a question, activation spreads through the network to all nodes connected to each of the concepts in the question. For example, if the question were "How many animals of each kind did Moses take on the ark?", activation would spread to nodes associated with animals (e.g., dogs, cats, etc.) and to nodes associated with Moses and the Ark (e.g., Old Testament, water, etc.). The difference between both groups of nodes is their exclusivity to either the distorted or undistorted term: The animal-related nodes are distinguishing features for Noah, whereas Old Testament and water are similar features, that is, Moses and Noah share them (cf. Kamas et al., 1996). Higher numbers of similar features make detecting the distortion harder and lower numbers of similar features make it easier, while the opposite holds true for distinguishing features (the more, the easier). Concretely, if the total activation contained in distinguishing features does not rise above a certain threshold, participants will accept the question as a partial match and treat it as undistorted. In Chapter 2.1, participants in the high incentives condition had higher response times, implying they spent more time deliberating the correct response. It is possible that this led to higher overall activation in the semantic network, and thus a more likely detection of a distortion, but since the distorted terms in our experiment were specifically chosen to have many overlapping features, a large part of the activation spread into the nodes representing shared features, making the detection of a distortion unlikely. In other words, making participants read the questions more intently or for longer might change the overall activation level but will not change the activation level difference between shared and unique nodes. As the detection of a distortion is directly related to the difference in activation between distinguishing features compared to similar features (cf. Kamas et al., 1996) this might explain why motivation showed little effects in Chapter 2.1.

For why motivation did not influence the truth effect, consider the fluency explanation from Chapter 1.1.1: People process repeated statements more easily, perceiving them as more fluent, and thus more likely rate them as true compared to novel statements. They do so under the (implicit) assumption that processing fluency is a valid cue for truth (cf. Reber & Unkelbach, 2010). In Chapter 3.1, participants rated mostly unknown statements as either true or false. Assuming they had no other cues to rely on for this forced judgment and given that fluency is a generally valid cue, they used fluency to judge truth. This process is unlikely to change through increased motivation since more intent reading of the statements in the judgment phase will not provide participants with other cues to base their judgment on. As long as participants do not learn that fluency is not a valid cue in experimental situations, they are unlikely to change their strategy. In contrast, Chapter 4.1 provided participants with advice in addition to incentivizing correct responses. Although previous research showed that knowledge and advice do not attenuate the truth effect (Fazio et al., 2015; Unkelbach & Greifeneder, 2018), we hypothesized that advice may serve as an alternative basis for judgment other than fluency, if correct responses are incentivized, because actively ignoring the advice would likely reduce the bonus payment.

Nevertheless, Chapter 4.1 showed that advice did not eliminate the truth effect, even if combined with monetary incentives. To provide a possible explanation for why incentives combined with advice still failed to eliminate the truth effect, consider the previously mentioned referential model (Unkelbach & Rom, 2017). Reading a statement in the presentation phase should instigate a new referential network. If the statement is not obviously false, this new network should have several coherent links between references, making it more likely that a person will judge the statement the references are corresponding to as true (i.e., the truth effect). Now, if a person receives incongruent advice, this poses as an incompatible reference for the rest of the network. From a cognitive dissonance perspective (Festinger, 1957) the person should have two options from this point on. First, modifying the referential network to compensate for the newly received incompatible advice, or second, disregarding the advice by discrediting the source (i.e., the advisor). In other words, people would have to reject an outside source to keep following their intuition. If this theory is true, then the second option (i.e., disregarding the advice) would be the easier option in most cases and should be chosen more often. Introducing monetary consequences should increase the stakes for this decision, though it is possible that people will still follow their intuition. However, the outside source (i.e., advisor) would then have to be discredited even more to reduce the dissonance emerging through ignoring valid advice despite real monetary consequences, which should be investigated in future research. While only speculative, this theory fits the data, and would suggest that the truth effect and its related phenomena, such as fake news, are hard to attenuate.

6.2 Potential Future Attenuation Strategies

Considering the evidence, it seems both the Moses illusion and the truth effect are resistant to many attenuation attempts. However, based on the reasoning in the previous chapters, a combination of motivation and strategy teaching might reduce both illusions.

For the truth effect, Unkelbach (2006, 2007) showed that teaching participants about the validity of fluency as a cue through feedback on their decisions could even reverse the truth effect. Similarly, Nadarevic and Aßfalg (2017) showed that explaining repetition and asking participants to not rely on it as a cue, while not eliminating the truth effect entirely, weakened it. Taken together, attenuation of the truth effect might be possible by setting an accuracy goal in people (through adding consequences to their decisions) and providing them with strategies that can help avoid the truth effect (e.g., how to interpret repetition, identifying the memory source, or assessing believability of statements when first seen, cf. Brashier et al., 2020). Future research will have to investigate which strategies are effective and how strong the motivation manipulation must be to diminish the truth effect.

For the Moses illusion, Kamas et al. (1996) showed that in all their attempts to reduce the illusion, only sensibilizing participants to distinguishing features between the distorted and undistorted terms had an effect. In their fourth experiment, they presented participants with additional, related questions before showing them the actual questions. The related questions could focus either on the shared features between the terms (e.g., “What religions study the story of Moses?”, p. 694) or on the distinguishing features (e.g., “What sea did Moses part?”, p. 694). Distortion detection improved for questions preceded by a related question that focused on distinguishing features. This result is in line with the semantic network theory (cf. Chapter 1.2.1), since more activation spreading into unshared nodes makes it more likely that a distortion is detected. While this experiment used a task-related manipulation (i.e., preceding questions), it could also be possible to provide participants with strategies to come up with distinguishing features by themselves. For example, an experimental instruction could be: “For each question, please carefully consider each word and try to remember as many different facts or stories related to it as possible”. When reading the question and arriving at the name “Moses”, this could help participants remember the other stories Moses appeared in (e.g., the flaming bush, parting the sea, receiving the ten commandments), which serve as distinguishing features from Noah and would make it more likely that participants detect the distortion. Of course, employing this strategy would result in additional effort, which participants could be averse to, but combining this with incentivized decisions (resulting in more effort) could reduce the Moses illusion. Future research will have to investigate the effectiveness of such strategy-teaching instructions combined with incentives.

Although I proposed this attenuation attempt through teaching strategies in addition to incentivizing decisions specifically for the Moses illusion and the truth effect, it is possible that this would reduce other cognitive illusions as well. In line with this notion, Agnoli and Krantz (1989) wanted to reduce the conjunction fallacy (overestimating the probability of compound events; e.g., Tversky & Kahneman, 1983). They recruited participants with "little mathematical background" (p. 515) and provided them with basic algebra training using Venn diagrams, which reduced the strength of the conjunction fallacy and stands in contrast to the original findings by Tversky and Kahneman (1983) who claimed that statistical training had little effect. This suggests that providing participants with the right training and strategies can reduce the use of heuristics and thus the impact of cognitive illusions. Future research should investigate what forms of teaching and which strategies and trainings are effective for which cognitive illusions.

6.2.1 Comparing Explanations

Given that both the Moses illusion and the truth effect are not moderated by motivation, and the semantic network theory and the referential theory are similar, it makes sense to compare the two.

The referential theory posits that memory references corresponding to the elements of a statement are organized in a network with the number of coherent links within the network determining the likelihood that a statement is judged as true (Unkelbach & Rom, 2017). Importantly, the referential theory also explains the extension of existing networks. When novel information is presented in a statement, new links will form between existing memory structures and the references corresponding to the novel information. The semantic network model, on the other hand, posits that reading a question sends activation through the memory network and activation accrues at different concepts based on the number and strength of connections to the concepts contained in the question.

Both models have in common that the number of links between memory nodes determines the strength of the illusion, but the semantic network model also considers secondary links between related nodes that do not directly correspond to concepts in the question. For example, when reading the statement “Moses took two animals of each kind on the ark”, the referential theory would predict that the nodes “animals”, “two”, and “ark” share excitatory links, but the node “Moses” would share an inhibitory link with “ark”, prohibiting parallel constraint satisfaction and resulting in the rejection of the network (cf. Unkelbach & Rom, 2017). A semantic network would also consider the links between implicit concepts (e.g., “Bible”, “Old Testament”, “water”) that are not explicitly mentioned in the question. Because Moses and Noah share many features, there would be many strong links between Moses and the other nodes activated by the question, resulting in missing the distortion.

Future research should investigate how these two models are compatible. For instance, it would be interesting to investigate what would happen if the Moses statements were repeated, and how the predictions of the referential theory would change if it also considered secondary implicit links.

6.3 Implications

The implications of the reported research differ between the two cognitive illusions. The Moses illusion only occurs under certain circumstances: The distorted and undistorted terms have to be semantically similar and cannot be the target of the question. As mentioned before, inserting Nixon instead of Moses will make distortion detection trivial, as will directly asking for the name of the person who took two animals of each kind on the ark (as used as knowledge checks in previous experiments). A typical conversation does not usually contain such trick questions, and thus the risk of falling for the Moses illusion in everyday conversations is negligible. As such, the failed attempts to substantially attenuate the Moses illusion through incentives have little impact in real life. However, the multiple-choice format introduced in Chapter 2 will make future research on this illusion easier. Moreover, our findings provide further evidence for the partial matching explanation. Kamas et al. (1996) concluded, “[i]t seems that people cannot easily become more vigilant at detecting distortions, even when they try” (p. 696), which can also be concluded from our results.

For the truth effect, our results suggest that adding consequences to decisions does not substantially reduce the truth effect, regardless of type of consequences (monetary or potential health consequences), presence or absence of external advice, or relevance of the statements. This further supports previous research demonstrating the robustness of the truth effect and suggests increased motivation alone is unlikely to reduce heuristic decision making. However, as the truth effect is often encountered in everyday situations, our results also have implications for real life. Importantly, the truth effect plays a role in the spread of fake news: Vosoughi et al. (2018) found that false news stories spread faster and more broadly on Twitter, especially if they were related to politics. Pennycook et al. (2018) showed that repetition of fake news increases their believability. Moreover, informing participants that these items were contended by fact checkers did not reduce the effect. The results from Chapter 4.1 are in line with these findings: Fact checkers are similar to our advice in that they provide a highly valid assessment about the truth status of a given statement. However, our results suggest that even incentivizing correct truth judgements is unlikely to make people rely on advice from fact checkers. Similarly, the results from Chapter 3.1 suggest that even giving people money to avoid being influenced by repetition of fake news will likely have no effect. Finally, we showed that even relevant statements that have potential health consequences for oneself are believed more when repeated, suggesting that the role of the truth effect in fake news is not limited to statements about events without personal consequences. This especially raises concerns about the spread of (mis)information during the current COVID-19 pandemic.

6.4 Limitations

Despite the important theoretical and practical consequences, it is important to consider some limitations of this work. The following limitations are in addition to those in the respective chapters (2-5) and apply to the dissertation as a whole rather than to individual experiments. While I discussed several cognitive illusions, I only provided evidence for the Moses illusion and truth effect, and thus the findings cannot be directly generalized to other cognitive illusions. However, based on the arguments mentioned in Chapter 6.2, it would be promising to investigate other cognitive illusions with the findings of this work in mind. For example, future research could investigate the effectivity of incentives in combination with strategy-teaching for several cognitive illusions, which could result in a classification system for cognitive illusions: Those that are more easily overcome and those that are especially robust.

Comparing this work to the research by Enke et al. (2021), one could argue that the incentives in Chapters 2-4 were relatively low. Nevertheless, participants in the high incentives condition of our experiments could earn a bonus of up to eight times of our standard laboratory compensation, but due to the higher payment for participant samples from Germany and the U.S., we were unable to provide incentives as high as those in the experiments by Enke et al. (2021). This could mean that motivation was not manipulated as strongly as possible and some of the effects could have been underestimated. Yet it is unlikely that even very high incentives would have led to substantial effects, as even the very high incentives used by Enke et al. (2012) led to small or no effects. Moreover, even if the illusions had been attenuated by incentivizing participants with thousands of dollars, this would not be a feasible manipulation for most laboratories or in real life.

6.5 Conclusion

In this dissertation, I have shown that the truth effect and the Moses illusion as examples of cognitive illusions are robust even in the face of relevant decision consequences for participants. Across four chapters, neither monetary incentives for correct responses, nor monetary incentives paired with highly valid advice, nor relevant statements with potential health consequences substantially reduced the illusions. This suggests that they are not due to laboratory conditions and their lack of performance-based incentives, as is often typical for psychological research. Some research has claimed that the general lack of performance-based incentives in psychological research leads to “variability in the data” and “ultimately may impede theoretical advances” (Hertwig & Ortmann, 2001, p.398). However, the results in this work are more compatible with a literature review by Camerer and Hogart (1999). They reviewed 74 papers and found that “[i]ncentives improve performance [only] in easy tasks that are effort-responsive[…]” (p. 34). While I presented evidence in this dissertation that effort did increase for participants in the high incentives conditions (as indicated by higher response times), it seems that neither the Moses illusion nor the truth effect are responsive to effort. I argued that increased motivation and effort could attenuate the illusions when paired with strategy-teaching. Providing participants with a (possibly effortful) strategy and then incentivizing them to employ that strategy might diminish both illusions but offering incentives without a means to overcome the illusion will likely be ineffective. Future research will have to address this hypothesis, as well as the generalizability of this attenuation strategy to other cognitive illusions.

References

Agnoli, F., & Krantz, D. H. (1989). Suppressing natural heuristics by formal instruction: The case of the conjunction fallacy. *Cognitive Psychology*, *21*(4), 515–550. https://doi.org/10.1016/0010-0285(89)90017-0

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Alves, H., Koch, A., & Unkelbach, C. (2017). Why good is more alike than bad: Processing implications. *Trends in Cognitive Sciences*, *21*(2), 69–79. https://doi.org/10.1016/j.tics.2016.12.006

Arkes, H. R., Boehm, L. E., & Xu, G. (1991). Determinants of judged validity. *Journal of Experimental Social Psychology*, *27*(6), 576–605. https://doi.org/10.1016/0022-1031(91)90026-3

Arkes, H. R., Hackett, C., & Boehm, L. (1989). The generality of the relation between familiarity and judged validity. *Journal of Behavioral Decision Making*, *2*(2), 81–94. https://doi.org/10.1002/bdm.3960020203

Bacon, F. T. (1979). Credibility of repeated statements: Memory for trivia. *Journal of Experimental Psychology: Human Learning and Memory*, *5*(3), 241–252. https://doi.org/10.1037/0278-7393.5.3.241

Bago, B., Rand, D. G., & Pennycook, G. (2020). Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines. *Journal of Experimental Psychology: General*, *149*(8), 1608–1613. https://doi.org/10.1037/xge0000729

Begg, I. M., Anas, A., & Farinacci, S. (1992). Dissociation of processes in belief: Source recollection, statement familiarity, and the illusion of truth. *Journal of Experimental Psychology: General*, *121*(4), 446–458. https://doi.org/10.1037/0096-3445.121.4.446

Biernat, M. (1989). Motives and values to achieve: Different constructs with different effects. *Journal of Personality*, *57*(1), 69–95. https://doi.org/10.1111/j.1467-6494.1989.tb00761.x

Blendermann, M. F., Little, J. L., & Gray, K. M. (2020). How “none of the above” (NOTA) affects the accessibility of tested and related information in multiple-choice questions. *Memory*, *28*(4), 473–480. https://doi.org/10.1080/09658211.2020.1733614

Bonner, S. E., & Sprinkle, G. B. (2002). The effects of monetary incentives on effort and task performance: Theories, evidence, and a framework for research. *Accounting, Organizations and Society*, *27*(4–5), 303–345. https://doi.org/10.1016/S0361-3682(01)00052-6

Brashier, N. M., Eliseev, E. D., & Marsh, E. J. (2020). An initial accuracy focus prevents illusory truth. *Cognition*, *194*, 104054. https://doi.org/10.1016/j.cognition.2019.104054

Brashier, N. M., & Marsh, E. J. (2020). Judging truth. *Annual Review of Psychology*, *71*(1), 499–515. https://doi.org/10.1146/annurev-psych-010419-050807

Bredart, S., & Modolo, K. (1988). Moses strikes again: Focalization effect on a semantic illusion. *Acta Psychologica*, *67*(2), 135–144. https://doi.org/10.1016/0001-6918(88)90009-1

Brown, A. S., & Nix, L. A. (1996). Turning lies into truths: Referential validation of falsehoods. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(5), 1088–1100. https://doi.org/10.1037/0278-7393.22.5.1088

Brunstein, J. C., & Heckhausen, H. (2018). Achievement motivation. In J. Heckhausen & H. Heckhausen (Eds.), *Motivation and Action* (3rd ed.).

Brunswik, E. (1952). *The conceptual framework of psychology*. University of Chicago Press.

Butler, A. C. (2018). Multiple-choice testing in education: Are the best practices for assessment also good for learning? *Journal of Applied Research in Memory and Cognition*, *7*(3), 323–331. https://doi.org/10.1016/j.jarmac.2018.07.002

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

Camerer, C. F., & Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty*, *19*, 7–42. https://doi.org/10.1023/A:1007850605129

Cantor, A. D., & Marsh, E. J. (2017). Expertise effects in the Moses illusion: detecting contradictions with stored knowledge. *Memory*, *25*(2), 220–230. https://doi.org/10.1080/09658211.2016.1152377

Cerasoli, C. P., Nicklin, J. M., & Ford, M. T. (2014). Intrinsic motivation and extrinsic incentives jointly predict performance: A 40-year meta-analysis. *Psychological Bulletin*, *140*(4), 980–1008. https://doi.org/10.1037/a0035661

Chapman, L. J. (1967). Illusory correlation in observational report. *Journal of Verbal Learning and Verbal Behavior*, *6*(1), 151–155. https://doi.org/10.1016/S0022-5371(67)80066-5

Corneille, O., Mierop, A., & Unkelbach, C. (2020). Repetition increases both the perceived truth and fakeness of information: An ecological account. *Cognition*, *205*, 104470. https://doi.org/10.1016/j.cognition.2020.104470

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition*, *31*(3), 187–276. https://doi.org/10.1016/0010-0277(89)90023-1

De Houwer, J. (2011). Why the cognitive approach in psychology would profit from a functional approach and vice versa. *Perspectives on Psychological Science*, *6*(2), 202–209. https://doi.org/10.1177/1745691611400238

De keersmaecker, J., Dunning, D., Pennycook, G., Rand, D. G., Sanchez, C., Unkelbach, C., & Roets, A. (2020). Investigating the robustness of the illusory truth effect across individual differences in cognitive ability, need for cognitive closure, and cognitive style. *Personality and Social Psychology Bulletin*, *46*(2), 204–215. https://doi.org/10.1177/0146167219853844

Dechêne, A., Stahl, C., Hansen, J., & Wänke, M. (2010). The truth about the truth: A meta-analytic review of the truth effect. *Personality and Social Psychology Review*, *14*(2), 238–257. https://doi.org/10.1177/1088868309352251

Eisenhardt, K. M. (1989). Agency theory: An assessment and review. *Academy of Management Review*, *14*(1), 57–74. https://doi.org/10.5465/amr.1989.4279003

Enke, B., Gneezy, U., Hall, B., Martin, D., Nelidov, V., Offerman, T., & van de Ven, J. (2021). Cognitive biases: mistakes or missing stakes? In *NBER working paper* (Issue 9). https://doi.org/10.3386/w28650

Erickson, T. D., & Mattson, M. E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior*, *20*(5), 540–551. https://doi.org/10.1016/S0022-5371(81)90165-1

Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, *8*(3), 223–241. https://doi.org/10.1177/1745691612460685

Fazio, L. K., Brashier, N. M., Payne, B. K., & Marsh, E. J. (2015). Knowledge does not protect against illusory truth. *Journal of Experimental Psychology: General*, *144*(5), 993–1002. https://doi.org/10.1037/xge0000098

Fazio, L. K., Rand, D. G., & Pennycook, G. (2019). Repetition increases perceived truth equally for plausible and implausible statements. *Psychonomic Bulletin and Review*, *26*(5), 1705–1710. https://doi.org/10.3758/s13423-019-01651-4

Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior. *Advances in Experimental Social Psychology*, *23*, 75–109.

Festinger, L. (1957). *A theory of cognitive dissonance* (Vol. 2). Stanford University press.

Feustel, T. C., Shiffrin, R. M., & Salasoo, A. (1983). Episodic and lexical contributions to the repetition effect in word identification. *Journal of Experimental Psychology: General*, *112*(3), 309–346. https://doi.org/10.1037/0096-3445.112.3.309

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

Garcia-Marques, T., Silva, R. R., Reber, R., & Unkelbach, C. (2015). Hearing a statement now and believing the opposite later. *Journal of Experimental Social Psychology*, *56*, 126–129. https://doi.org/10.1016/j.jesp.2014.09.015

Gehringer, W. L., & Engel, E. (1986). Effect of ecological viewing conditions on the Ames' distorted room illusion. *Journal of Experimental Psychology: Human Perception and*

Performance, *12*(2), 181–185. https://doi.org/10.1037/0096-1523.12.2.181

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, *102*(4), 684–704. https://doi.org/10.1037/0033-295X.102.4.684

Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). *Heuristics and biases: The psychology of intuitive judgment.* Cambridge university press.

Grice, H. P. (1975). Logic of conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Speech acts* (p. Vol. 3; pp. 41–58). Seminar Press.

Hamilton, D. L., & Gifford, R. K. (1976). Illusory correlation in interpersonal perception: A cognitive basis of stereotypic judgments. *Journal of Experimental Social Psychology*, *12*(4), 392–407. https://doi.org/10.1016/S0022-1031(76)80006-6

Hammond, K. R. (1955). Probabilistic functioning and the clinical method. *Psychological Review*, *62*(4), 255–262. https://doi.org/10.1037/h0046845

Hannon, B., & Daneman, M. (2001). Susceptibility to semantic illusions: An individual-differences perspective. *Memory and Cognition*, *29*(3), 449–461. https://doi.org/10.3758/BF03196396

Hansen, J., Dechêne, A., & Wänke, M. (2008). Discrepant fluency increases subjective truth. *Journal of Experimental Social Psychology*, *44*(3), 687–691. https://doi.org/10.1016/j.jesp.2007.04.005

Hasher, L., Goldstein, D., & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning and Verbal Behavior*, *16*(1), 107–112. https://doi.org/10.1016/S0022-5371(77)80012-1

Hertwig, R., & Ortmann, A. (2001). Experimental practices in economics: A methodological challenge for psychologists? *Behavioral and Brain Sciences*, *24*(3), 383–403. https://doi.org/10.1017/S0140525X01004149

Heyman, J., & Ariely, D. (2004). Effort for payment: A tale of two markets. *Psychological Science*, *15*(11), 787–793. https://doi.org/10.1111/j.0956-7976.2004.00757.x

Jacoby, L. L., Kelley, C. M., Brown, J., & Jasechko, J. (1989). Becoming famous overnight: Limits on the ability to avoid unconscious influences of the past. *Journal of Personality and Social Psychology*, *56*(3), 326–338.

Jalbert, M., Newman, E., & Schwarz, N. (2019). Trivia claim norming: Methods report and data. In *ResearchGate*. https://doi.org/10.6084/m9.figshare.9975602

Jalbert, M., Newman, E., & Schwarz, N. (2020). Only half of what I'll tell you is true: Expecting to encounter falsehoods reduces illusory truth. *Journal of Applied Research in Memory and Cognition*, *9*(4), 602–613.

Johar, G. V., & Roggeveen, A. L. (2007). Changing false beliefs from repeated advertising: The role of claim-refutation alignment. *Journal of Consumer Psychology*, *17*(2), 118–127. https://doi.org/10.1016/S1057-7408(07)70018-9

Johnston, W. A., Dark, V. J., & Jacoby, L. L. (1985). Perceptual fluency and recognition judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(1), 3–11. https://doi.org/10.1037/0278-7393.11.1.3

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*(1), 54–69.

https://doi.org/10.1037/a0028347

Kamas, E. N., & Reder, L. M. (1995). The role of familiarity in cognitive processing. In *Sources of coherence in reading.* (pp. 177–202). Lawrence Erlbaum Associates, Inc.

Kamas, E. N., Reder, L. M., & Ayers, M. S. (1996). Partial matching in the Moses illusion: Response bias not sensitivity. *Memory & Cognition*, *24*(6), 687–699. https://doi.org/10.3758/BF03201094

Karelaia, N., & Hogarth, R. M. (2008). Determinants of linear judgment: A meta-analysis of lens model studies. *Psychological Bulletin*, *134*(3), 404–426. https://doi.org/10.1037/0033-2909.134.3.404

Kruglanski, A. W., & Webster, D. M. (1996). Motivated closing of the mind: “Seizing” and “freezing.” *Psychological Review*, *103*(2), 263–283. https://doi.org/10.1037/0033-295X.103.2.263

Kurzban, R., Duckworth, A., Kable, J. W., & Myers, J. (2013). Cost-benefit models as the next, best option for understanding subjective effort. *Behavioral and Brain Sciences*, *36*(6), 707–726. https://doi.org/10.1017/S0140525X13001532

Lacassagne, D., Béna, J., & Corenille, O. (2021). *Is the earth a perfect square? Repetition increases the perceived truth of highly implausable statements*. https://doi.org/10.31234/osf.io/fce8z

Lakens, D. (2014). Performing high-powered studies efficiently with sequential analyses. *European Journal of Social Psychology*, *44*(7), 701–710. https://doi.org/10.1002/ejsp.2023

Lakens, D. (2017). Equivalence tests: A practical primer for t tests, correlations, and metaanalyses. *Social Psychological and Personality Science*, *8*(4), 355–362. https://doi.org/10.1177/1948550617697177

Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science*, *1*(2), 259–269. https://doi.org/10.1177/2515245918770963

Langer, E. J. (1975). The illusion of control. *Journal of Personality and Social Psychology*, *32*(2), 311–328. https://doi.org/10.1037/0022-3514.32.2.311

Law, S., Hawkins, S. A., & Craik, F. I. M. (1998). Repetition-induced belief in the elderly: Rehabilitating age-related memory deficits. *Journal of Consumer Research*, *25*(2), 91–107. https://doi.org/10.1086/209529

Le Bon, G. (1996). The crowd: A study of the popular mind. Retrieved from Project Gutenberg: http://www.gutenberg.org/cache/epub/445/pg445.html. (Original work published 1895).

Ledgerwood, A. (2015). *Practical and painless: Five easy strategies to transition your lab*. Talk presented in a symposium on best practices at the annual conference of the Society for Personality and Social Psychology, Long Beach, CA.

Lee, D. S., Kim, E., & Schwarz, N. (2015). Something smells fishy: Olfactory suspicion cues improve performance on the Moses illusion and Wason rule discovery task. *Journal of Experimental Social Psychology*, *59*, 47–50. https://doi.org/10.1016/j.jesp.2015.03.006

Lev-Ari, S., & Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology*, *46*(6), 1093–1096. https://doi.org/10.1016/j.jesp.2010.05.025

Levordashka, A., & Lakens, D. (2015). Replication of Study 2 by Song & Schwarz (2008, Psychological Science). *OSF*. https://osf.io/u5rfy/

Locke, E. A. (1968). Toward a theory of task motivation and incentives. *Organizational Behavior and Human Performance*, *3*(2), 157–189. https://doi.org/10.1016/0030-5073(68)90004-4

Locke, E. A., Shaw, K. N., Saari, L. M., & Latham, G. P. (1981). Goal setting and task performance: 1969-1980. *Psychological Bulletin*, *90*(1), 125–152. https://doi.org/10.1037/0033-2909.90.1.125

Lowell, E. L. (1952). The effect of need for achievement on learning and speed of performance. *The Journal of Psychology*, *33*(1), 31–40. https://doi.org/10.1080/00223980.1952.9712815

Lunney, G. H. (1970). Using analysis of variance with a dichotomous dependent variable: An empirical study. *Journal of Educational Measurement*, *7*(4), 263–269. https://doi.org/10.1111/j.1745-3984.1970.tb00727.x

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, *87*(3), 252–271. https://doi.org/10.1037/0033-295X.87.3.252

McCaul, K. D., Hockemeyer, J. R., Johnson, R. J., Zetocha, K., Quinlan, K., & Glasgow, R. E. (2006). Motivation to quit using cigarettes: A review. *Addictive Behaviors*, *31*(1), 42–56. https://doi.org/10.1016/j.addbeh.2005.04.004

McGlone, M. S., & Tofighbakhsh, J. (2000). Birds of a feather flock conjointly (?): Rhyme as reason in aphorisms. *Psychological Science*, *11*(5), 424–428. https://doi.org/10.1111/1467-9280.00282

Nadarevic, L., & Aßfalg, A. (2017). Unveiling the truth: warnings reduce the repetition-based truth effect. In *Psychological Research* (Vol. 81, Issue 4, pp. 814–826). https://doi.org/10.1007/s00426-016-0777-y

Newman, E. J., Jalbert, M. C., Schwarz, N., & Ly, D. P. (2020). Truthiness, the illusory truth effect, and the role of need for cognition. *Consciousness and Cognition*, *78*(December 2019), 102866. https://doi.org/10.1016/j.concog.2019.102866

Olds, J. M., & Westerman, D. L. (2012). Can fluency be interpreted as novelty? Retraining the interpretation of fluency in recognition memory. *Journal of Experimental Psychology: Learning Memory and Cognition*, *38*(3), 653–664. https://doi.org/10.1037/a0026784

Oswald, M. E., & Grosjean, S. (2004). Confirmation bias. In R. F. Pohl (Ed.), *Cognitive illusions: A handbook on fallacies and biases in thinking, judgement and memory* (pp. 79–96). Psychology Press; Taylor & Francis.

Park, H., & Reder, L. M. (2004). Moses Illusion: Implication for human cognition. In R. F. Pohl (Ed.), *Cognitive illusions: a handbook on fallacies and biases in thinking, judgement, and memory* (pp. 275–291). Psychology Press; Taylor & Francis.

Pennycook, G., Cannon, T. D., & Rand, D. G. (2018). Prior exposure increases perceived accuracy of fake news. *Journal of Experimental Psychology: General*, *147*(12), 1865–1880. https://doi.org/10.1037/xge0000465

Pohl, R. F. (Ed.). (2004). *Cognitive illusions: a handbook on fallacies and biases in thinking, judgement, and memory*. Psychology Press; Taylor & Francis.

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

Reber, R., & Unkelbach, C. (2010). The epistemic status of processing fluency as source for judgments of truth. *Review of Philosophy and Psychology*, *1*(4), 563–581. https://doi.org/10.1007/s13164-010-0039-7

Reder, L. M., & Cleeremans, A. (1990). The role of partial matches in comprehension: The moses illusion revisited. In *Psychology of Learning and Motivation - Advances in Research and Theory* (Vol. 25, Issue C, pp. 233–258). https://doi.org/10.1016/S0079-7421(08)60258-3

Reder, L. M., & Kusbit, G. W. (1991). Locus of the Moses Illusion: Imperfect encoding, retrieval, or match? *Journal of Memory and Language*, *30*(4), 385–406. https://doi.org/10.1016/0749-596X(91)90013-A

Ruhrmann, G., & Daube, D. (2021). Die Rolle der Medien in der COVID-19-Pandemie. In A. W. Lohse (Ed.), *Infektionen und Gesellschaft* (pp. 119–134). Springer Berlin Heidelberg.

Scholl, S. G., Greifeneder, R., & Bless, H. (2014). When fluency signals truth: Prior successful reliance on fluency moderates the impact of fluency on truth judgments. *Journal of Behavioral Decision Making*, *27*(3), 268–280. https://doi.org/10.1002/bdm.1805

Schwartz, M. (1982). Repetition and Rated Truth Value of Statements. *The American Journal of Psychology*, *95*(3), 393. https://doi.org/10.2307/1422132

Shafto, M., & MacKay, D. G. (2000). The Moses, mega-Moses, and Armstrong illusions: Integrating language comprehension and semantic memory. *Psychological Science*, *11*(5), 372–378. https://doi.org/10.1111/1467-9280.00273

Silva, R. R., Garcia-Marques, T., & Reber, R. (2017). The informative value of type of repetition: Perceptual and conceptual fluency influences on judgments of truth.

Consciousness and Cognition, *51*, 53–67. https://doi.org/10.1016/j.concog.2017.02.016

Skurnik, I., Yoon, C., Park, D. C., & Schwarz, N. (2005). How warnings about false claims become recommendations. *Journal of Consumer Research*, *31*(4), 713–724. https://doi.org/10.1086/426605

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. https://doi.org/10.1521/soco.2008.26.6.791

Speckmann, F., & Unkelbach, C. (2021a). *Monetary incentives do not reduce the repetition-induced truth effect*. Manuscript in preparation.

Speckmann, F., & Unkelbach, C. (2021b). The Moses illusion. In R. F. Pohl (Ed.), *Cognitive illusions*. In press.

Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest*, *1*(1), 1–26. https://doi.org/10.1111/1529-1006.001

Terborg, J. R. (1976). The motivational components of goal setting. *Journal of Applied Psychology*, *61*(5), 613–621. https://doi.org/10.1037/0021-9010.61.5.613

Terborg, J. R., & Miller, H. E. (1978). Motivation, behavior, and performance: A closer examination of goal setting and monetary incentives. *Journal of Applied Psychology*, *63*(1), 29–39. https://doi.org/10.1037/0021-9010.63.1.29

Thompson, S. C., & Spacapan, S. (1991). Perceptions of control in vulnerable populations. *Journal of Social Issues*, *47*(4), 1–21. https://doi.org/10.1111/j.1540-4560.1991.tb01831.x

Topolinski, S., & Strack, F. (2009). The analysis of intuition: Processing fluency and affect in judgements of semantic coherence. *Cognition & Emotion*, *23*(8), 1465–1503. https://doi.org/10.1080/02699930802420745

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, *90*(4), 293–315. https://doi.org/10.1037/0033-295X.90.4.293

Umanath, S., Dolan, P. O., & Marsh, E. J. (2014). Ageing and the Moses Illusion: Older adults fall for Moses but if asked directly, stick with Noah. *Memory*, *22*(5), 481–492. https://doi.org/10.1080/09658211.2013.799701

Unkelbach, C. (2006). The learned interpretation of cognitive fluency. *Psychological Science*, *17*(4), 339–345. https://doi.org/10.1111/j.1467-9280.2006.01708.x

Unkelbach, C. (2007). Reversing the truth effect: Learning the interpretation of processing fluency in judgments of truth. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(1), 219–230. https://doi.org/10.1037/0278-7393.33.1.219

Unkelbach, C., Alves, H., & Koch, A. (2020). Negativity bias, positivity bias, and valence asymmetries: Explaining the differential processing of positive and negative information. In *Advances in Experimental Social Psychology* (1st ed., Vol. 62). Elsevier Inc. https://doi.org/10.1016/bs.aesp.2020.04.005

Unkelbach, C., & Greifeneder, R. (2013). The Experience of Thinking. In C. Unkelbach & R. Greifeneder (Eds.), *The Experience of Thinking: How the Fluency of Mental Processes*

Influences Cognition and Behavior. Psychology Press. https://doi.org/10.4324/9780203078938

Unkelbach, C., & Greifeneder, R. (2018). Experiential fluency and declarative advice jointly inform judgments of truth. *Journal of Experimental Social Psychology*, *79*(December 2017), 78–86. https://doi.org/10.1016/j.jesp.2018.06.010

Unkelbach, C., Koch, A., & Alves, H. (2019). The evaluative information ecology: On the frequency and diversity of “good” and “bad.” *European Review of Social Psychology*, *30*(1), 216–270. https://doi.org/10.1080/10463283.2019.1688474

Unkelbach, C., Koch, A., Silva, R. R., & Garcia-Marques, T. (2019). Truth by repetition: Explanations and implications. *Current Directions in Psychological Science*, *28*(3), 247–253. https://doi.org/10.1177/0963721419827854

Unkelbach, C., & Rom, S. C. (2017). A referential theory of the repetition-induced truth effect. *Cognition*, *160*, 110–126. https://doi.org/10.1016/j.cognition.2016.12.016

Unkelbach, C., & Speckmann, F. (2021). Mere repetition increases belief in factually true COVID-19-related information. *Journal of Applied Research in Memory and Cognition*, *10*(2), 241–247. https://doi.org/10.1016/j.jarmac.2021.02.001

Unkelbach, C., & Stahl, C. (2009). A multinomial modeling approach to dissociate different components of the truth effect. *Consciousness and Cognition*, *18*(1), 22–38. https://doi.org/10.1016/j.concog.2008.09.006

Van Oostendorp, H., & De Mul, S. (1990). Moses beats Adam: A semantic relatedness effect on a semantic illusion. *Acta Psychologica*, *74*(1), 35–46. https://doi.org/10.1016/0001-6918(90)90033-C van Oostendorp, H., & Kok, I. (1990). Failing to notice errors in sentences. *Language and Cognitive Processes*, *5*(2), 105–113. https://doi.org/10.1080/01690969008402100

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559

Vroom, V. (1964). *Work and motivation*. John Wiley.

Wendling, M. (2018). *The (almost) complete history of "fake news."* BBC News. https://www.bbc.com/news/blogs-trending-42724320

Whittlesea, B. W. A., Jacoby, L. L., & Girard, K. (1990). Illusions of immediate memory: Evidence of an attributional basis for feelings of familiarity and perceptual quality. *Journal of Memory and Language*, *29*(6), 716–732. https://doi.org/10.1016/0749-596X(90)90045-2

Wyer, R. S. (2008). The role of knowledge accessibility in cognition and behavior: Implications for consumer information processing. In C. P. Haugtvedt, P. M. Herr, & F. R. Kardes (Eds.), *Handbook of consumer psychology* (pp. 31–76). Taylor & Francis Group/Lawrence Erlbaum Associates.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, *9*(2, Pt.2), 1–27. https://doi.org/10.1037/h0025848

Zipf, G. K. (1949). *Human behavior and the principle of least effort.* Addison-Wesley.