Cues and Processes Underlying Metamemory

Malte F. Zimdahl

Inaugural Dissertation
Submitted in partial fulfillment of the requirements for the degree Doctor of Social Sciences in the Graduate School of Economic and Social Sciences at the University of Mannheim

Supervisors:
PD Dr. Monika Undorf
Prof. Dr. Edgar Erdfelder

Dean of the School of Social Sciences:
Prof. Dr. Michael Diehl

Thesis Reviewers:
Prof. Dr. Edgar Erdfelder
Prof. Dr. Beatrice G. Kuhlmann

Defense Committee:
PD Dr. Monika Undorf
Prof. Dr. Edgar Erdfelder
Prof. Dr. Beatrice G. Kuhlmann

Thesis Defense:
May 17, 2022

For my family

Contents

Summary

What does *The Thinker* think about? If he is reflecting on his own thoughts and cognitions, then he is engaging in what is probably one of the most remarkable abilities of human kind: *metacognition*. For over half a century, psychology has examined this ability by especially addressing what people know about their own memory, that is, their *metamemory*. This research showed that people—in general—have an accurate metamemory but sometimes make memory predictions that are systematically different from their actual memory performance, leading to an inaccurate metamemory.

The target of this thesis is to investigate *when* and *why* metamemory is accurate or inaccurate. It does so by building on previous findings showing that we do not have direct access to our own memory but infer its state using different cues and heuristics. In line with this reasoning, I uncover *visual coherence* and *outcome knowledge* as two new cues that people use as basis for their predictions of future memory performance (*judgments of learning*, JOLs). I also show that visual coherence is a valid cue as it aligns JOLs with actual memory performance, whereas outcome knowledge systematically biases the recollection of JOLs towards the outcome of the memory test, leading to hindsight bias on JOLs.

Uncovering new cues that affect JOLs, memory performance, or both is not only of interest in its own right, but also has the potential to provide important insights about the theoretical processes underlying metacognitive judgments. Previously, the *dual-basis view of metacognition* has proven to be useful as a framework to integrate findings of cue effects on JOLs. The current thesis adopts and expands this dual-process perspective by testing the contributions of experience-based, intuitive, automatic processes and theory-based, deliberate, controlled processes to the visual coherence effect and hindsight bias on JOLs.

With that, this dissertation aims at advancing our understanding of people's metamemory. Furthermore, it underlines the reciprocal benefits of bridging together different research fields by showing that the same phenomena occur with JOLs and judgments about the external world. Finally, I hope that the practical implications of this thesis can help learners to improve their study process and prevent them from falling for metacognitive illusions.

Manuscripts

This thesis is the result of research conducted at the *Center for Doctoral Studies in Social and Behavioral Sciences* (CDSS) of the *Graduate School of Economic and Social Sciences* (GESS) at the University of Mannheim. It is based on three manuscripts, one of which has been published and two of which have been submitted for publication.

The three manuscripts aim at expanding our knowledge about the basis of people's metamemory. To this end, I uncover two new cues that are used to infer metacognitive judgments and investigate the underlying theoretical processes from a dual-basis view. I demonstrate that visual coherence impacts JOLs and that experience-based and theory-based processes contribute to this effect (Manuscript I). Furthermore, I show that outcome knowledge produces hindsight bias in metamemory (Manuscript II) and reveal that this is due to automatic rather than controlled processes (Manuscript III).

In the following chapters, I present the overarching theoretical framework of this thesis and illustrate how the three manuscripts are related to each other. By summarizing and discussing the results of the research, I refrain from elaborating on detailed aspects of the manuscripts which are appended to this thesis in the same order as listed below.

Manuscript I

Zimdahl, M. F., & Undorf, M. (2022a). *Visual coherence impacts judgments of learning: Evidence for the dual-basis view of metacognition* [Manuscript submitted for publication]. Department of Psychology, University of Mannheim.

Manuscript II

Zimdahl, M. F., & Undorf, M. (2021). Hindsight bias in metamemory: Outcome knowledge influences the recollection of judgments of learning. *Memory, 29*(5), 559-572. https://doi.org/10.1080/09658211.2021.1919144

Manuscript III

Zimdahl, M. F., & Undorf, M. (2022b). *Warnings and incentives do not reduce hindsight bias on judgments of learning* [Manuscript submitted for publication]. Department of Psychology, University of Mannheim.

1 Introduction

"Cogito ergo sum."[1]

René Descartes (1637/2006)

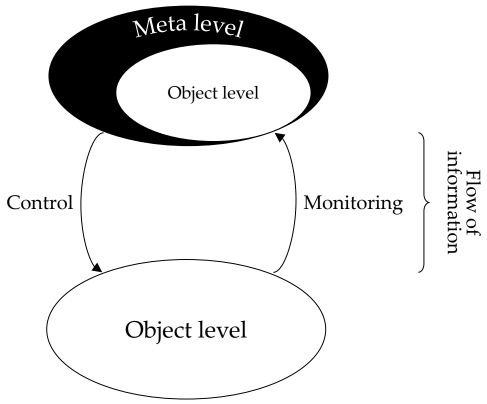
Thinking about one's own thinking has fascinated humans for centuries. The astonishing ability to assess our own memory and thought processes has been a topic not only in psychology but also in philosophy. Even long before René Descartes put down his famous statement into words, it were the scholars of the ancient Greeks who first wrote about their understanding of how memory works and how this is connected to the self (Robinson, 1989). Since then the question *What Do We Know About What We Know* has stimulated a large amount of theoretical considerations and scientific discourses, until it has ultimately led to the psychological discipline that—connecting to its roots—has been named *metamemory*.

As apparent from its Greek prefix, metamemory contains information about our own memory, thoughts, and cognitions (Flavell, 1971). In one of the first attempts to experimentally investigate people's metamemory, it was Joseph Hart (1965) who asked participants to state whether they knew an answer to a question although they were not able to recall it. He then correlated these *feeling-of-knowing judgments* (FOKs) with the memory performance in a subsequent four-alternative recognition test and showed that people have an above-chance accuracy of predicting what they will recognize although they are not able to recall it. This demonstrated that people have information about their own memory and that this information can be congruent with actual memory performance.

However, it took another 25 years until Nelson and Narens (1990) introduced their influential framework of how metamemory and memory are related to each other (see Figure 1). Central to their framework is the differentiation between an object level and a meta level, which are connected via a flow of information. The object level represents our memory and the meta level represents our metamemory. Information is flowing from the object level to the meta level via monitoring processes, whereas the meta level is sending information to the object level via control processes.

[1] I think, therefore I am.

Figure 1
Framework of Metamemory by Nelson and Narens (1990)



Note. The framework differentiates between an object level (which corresponds to memory) and a meta level (which corresponds to metamemory). The meta level includes a dynamic model of the object level. Information is constantly flowing between the levels via monitoring and control processes.

To illustrate this, we can think of a student who is preparing herself for an upcoming exam. While she is studying, the information within the object level (memory) is changing due to her learning progress. At the same time, she has to assess (monitor) what she has learned so far and what is yet to be learned. The result of these monitoring processes is an image of the object level within the meta level (metamemory). Based on this image, she can regulate (control) her learning. This can result in the fact that she ends learning because she thinks that the material is learned well enough or that she continues learning because she thinks that the material is not learned well enough. Alternatively, she might also change her learning strategy because she has experienced that her previous strategy was not successful.

What one can deduct from this example is that optimal learning depends on two factors. First, people's monitoring processes need to be accurate so that their metamemory may be congruent with their memory. Second, people's control processes need to be adequate so that they—based on their metamemory—may choose the right actions for their learning progress. This also implies that accurate monitoring processes are a prerequisite for adequate control processes, otherwise learning will be inefficient (cf. Bjork et al., 2013).

This is best illustrated when we look at defective monitoring or control processes. If monitoring processes are inaccurate, then metamemory will be incongruent with memory. Even if control processes are adequate in this case, the learner will probably spend time on material that she has already learned well enough because she erroneously thinks that she has not. Vice versa, there is a risk that she ends learning on material that she thinks is well enough learned although it is not. If, in turn, control processes are inadequate, learning will still be ineffective because despite of an accurate metamemory, people will draw the wrong conclusions from it. This is the case if the learner attributes her learning progress to factors that occurred during learning but were not responsible for her progress. The result is that she erroneously believes that a certain learning strategy was successful and that she sticks to this strategy. Vice versa, factors that overshadow a successful learning strategy might lead her to think that this strategy is not successful. She therefore might change her strategy to a potentially less successful one.

Because accurate monitoring processes are a necessary precondition for effective self-regulated learning, it is crucial to understand what factors lead to an accurate metamemory. In the past, a variety of studies have investigated the accuracy of people's metamemory by using metacognitive judgments such as the above mentioned FOKs. Apart from that, a large amount of research has concentrated on so-called *judgments of learning* (JOLs, Arbuckle & Cuddy, 1969). JOLs are subjective memory predictions about the likelihood that recently studied information can be recalled in a later test. These memory predictions can then be correlated with actual memory performance to determine the accuracy of people's metamemory.

Over the years, several studies have demonstrated that people are—in general—quite good at predicting their own memory performance with JOLs (e.g., Dunlosky & Metcalfe, 2009; Koriat, 1997). This means that usually there is a substantial correlation between JOLs and actual memory performance, showing that people are able to monitor their cognitive processes and forecast above chance what they will know at test. However, research has also shown that—under some circumstances—JOLs and actual memory performance systematically deviate from each other, thereby producing *metacognitive illusions* (see, e.g., Undorf, 2020; Undorf et al., 2022).

Metacognitive illusions have, as I will elaborate more in the next chapters, contributed much to our current understanding of how people's metamemory works. Specifically, there is substantial evidence that people infer their metamemory by using different cues that vary in their ability to correctly predict actual memory performance (e.g., Koriat, 1997; Undorf et al., 2018). Consequently, metamemory is accurate when people use cues that have predictive validity for their actual memory performance. Conversely, relying on cues that are not valid in predicting actual memory performance or failing to take valid cues into account can lead to an inaccurate metamemory, thereby fostering metacognitive illusions.

To be able to understand why metamemory is (in-)accurate, it is therefore necessary to investigate what cues people use to infer their metamemory and whether these cues are valid in predicting actual memory performance or not. However, at the same time, it is also important to understand what processes underly metacognitive judgments such as JOLs.

In recent years, there has been an active debate about the contributions of two different processes to JOLs: *fluency* and *beliefs*. The former proposes that JOLs are based on people's feelings about how easy it is to process specific material. In this case, certain cues influence people's experience of mastering information directly and intuitively, and these feelings of fluency are then used as a basis for JOLs in an unaware manner. The latter suggests that people have specific convictions about how certain cues affect metamemory. In that case, people detect differences in the material and deliberately apply specific beliefs when making JOLs.

Findings that some cues influence JOLs through feelings of fluency whereas other cues affect JOLs through beliefs are consistent with the *dual-basis view of metacognition* (e.g., Kelley & Jacoby, 1996; Koriat, 1997), assuming that experience-based processes and theory-based processes both contribute to JOLs. However, in recent years, this view has been challenged by an opposing account assuming that JOLs are based solely on beliefs (e.g., Mueller & Dunlosky, 2017; Mueller et al., 2016; Mueller et al., 2013). Investigating whether cue effects on JOLs can be better explained by one single process or two different processes can therefore help to advance the debate about the basis of people's metamemory.

The assumption that judgments are based on two qualitatively different modes of information processing is by far not exclusive to JOLs. In fact, this is an assumption shared by many *dual-process theories* which have been put forward as explanations of phenomena in the domains of (social) cognition, self-regulation, and reasoning (cf. Chaiken & Trope, 1999; Kahneman, 2003; Stanovich & West, 2000). Common to all these theories is the distinction between two kinds of processes that contribute to judgments under uncertainty which can be described as experience-based vs. theory-based, aware vs. unaware, or automatic vs. controlled. Over the years, the idea of dual-process theories has proven to be a helpful guideline as a framework to integrate different, sometimes contradicting, findings. However, with regard to JOLs, there is still much research to do, as Dunlosky and Tauber (2016) noted: "The circumstances under which JOLs are driven by experience, belief, or some combination of those influences remains poorly understood" (p. 75).

In my thesis, I use the dual-process framework to shed light on cues and processes underlying metamemory. More specifically, I uncover two new cues that have not been investigated as basis for metacognitive judgments but are well suited to gain insight into the theoretical basis of JOLs. With this, the current thesis aims at understanding not only what cues people use to form their JOLs, but also what processes contribute to what we know about what we know.

In Chapter 2, I will first review the theoretical foundations of how people form JOLs and what cues they use to monitor their memory. Within this chapter, I will also summarize the idea of dual-process theories and show how they have contributed to a more refined understanding of how metamemory works.

In Chapter 3, I uncover *visual coherence* and *outcome knowledge* as two new cues that people use to form their JOLs and investigate the underlying theoretical processes of these cues. In Manuscript I, I demonstrate that people consistently give higher JOLs to material that is visually coherent than incoherent (Zimdahl & Undorf, 2022a). Furthermore, I show that this influence of visual coherence on JOLs is based on people's feelings of fluency as well as their metacognitive beliefs, thus supporting the dual-basis view of metacognition. In Manuscript II, I reveal that people's recollections of prior JOLs are influenced by the knowledge about the outcome in a memory test, thereby producing hindsight bias in metamemory (Zimdahl & Undorf, 2021). In line with dual-process theories, I examine the contributions of automatic and controlled processes to hindsight bias on JOLs in Manuscript III (Zimdahl & Undorf, 2022b).

In Chapter 4, I discuss and integrate the findings of all three manuscripts within the dual-process framework and illustrate theoretical and practical implications. I conclude this thesis by addressing open questions and highlighting avenues for future research.

2 Theoretical Foundations

> "What, then, is memory development the development of? ... a kind of 'metamemory', perhaps. ... Let's all go out and study it!"
>
> John H. Flavell (1971)

As described in the introduction, the pure existence of a metamemory implies that we can assess our own memory, thoughts, and cognitions. This assessment is especially important in the process of learning, where it enables us to monitor what we have already mastered and what is yet to be learned. The result of these monitoring processes is an image of our memory in our metamemory, as has been illustrated in the framework by Nelson and Narens (1990, see Figure 1).

However, the image of our memory is not necessarily identical with metamemory. If this were the case, then metamemory would be a perfect description of what is going on in our memory. In contrast, metacognitive illusions show that metamemory and memory can systematically deviate from each other. This raises the question: how accurate is our metamemory? How is this measured? And how do we assess our metamemory?

2.1 Metacognitive Judgments

When assessing their metamemory, people have to rely on their own introspections. Research on metamemory has focused on capturing these introspections by asking people to make *metacognitive judgments*. These metacognitive judgments are assumed to reflect what people know and what they believe about their memory and are used to describe the image of the object level within the meta level.

In the past, different metacognitive judgments have been proposed to capture different aspects of metamemory. This is best illustrated by returning to the example of a student who is preparing herself for an upcoming exam. During her learning process, she has to acquire the material, retain it in her memory, and retrieve it whenever needed. Along every step of the way, she can monitor her learning progress by making different metacognitive judgments. *Ease-of-learning judgments* (EOLs), for example, are made in advance of learning and assess the subjective probability that specific material will be learned (e.g., Leonesio & Nelson, 1990). During acquisition, retention, and retrieval, the already mentioned FOKs are used to capture people's probability of having specific material in memory although it is not accessible at that moment (e.g., Hart, 1965). When retrieving material from memory, people can make metacognitive judgments in form of their confidence that their retrieved answer is correct (e.g., Hines et al., 2009).

Among the most used metacognitive judgments, however, are JOLs (Arbuckle & Cuddy, 1969). These metacognitive judgments are made during acquisition and assess the subjective probability that recently studied information can later be recalled at test (e.g., Rhodes, 2016). Usually, this is done by directly asking the participants to state their likelihood of remembering the information on a scale from 0-100%. Furthermore, participants are informed that a JOL of 0% indicates that they will not recall the information, whereas a JOL of 100% indicates that they will definitely recall the information in the memory test. JOLs can either be assessed immediately after an item is studied (e.g., a word pair) or after a certain delay (e.g., after all items have been studied). Apart from that, JOLs can be made for each item or for a complete list of items. The current thesis concentrates on immediate JOLs for each item. This means that participants made subjective memory predictions for every item right after the specific item was presented.

To determine the accuracy of these memory predictions, JOLs are compared with actual memory performance. Accuracy can be assessed in two different ways: *absolute accuracy* (or *calibration*) and *relative accuracy* (or *resolution*). Absolute accuracy refers to the congruence between JOLs and memory performance on an overall level. It is measured by comparing the average JOL per participant with the percentage of correct answers given by the same participant in the memory test. For example, if the participant's average JOL is 80% and she remembers 80% of the items in the memory test, then her JOLs are perfectly calibrated. Poorer calibration, on the other hand, emerges when the average JOL is not perfectly aligned with memory performance. When the average JOL is larger than the memory performance, then the person exhibits *overconfidence* in her memory predictions. Likewise, when the average JOL is lower than the memory performance, then the person exhibits *underconfidence* in her memory predictions.

Relative accuracy refers to the congruence between JOLs and memory performance on an item level. It is measured by correlating JOLs per participant with memory performance of the same participant in the memory test. Because memory performance per item is a binary outcome (item is remembered or not), research has used within-person Goodman-Kruskal gamma correlations (Nelson, 1984) to determine relative accuracy. This correlation can range from -1 to 1 and is higher, the better participants can discriminate between items which they remember and items which they do not remember. For example, if the participant gives high JOLs to all items which she later remembers in the memory test and low JOLs to all items which she does not remember, then her

JOLs have perfect resolution and her gamma correlation is 1. A gamma correlation of -1 would mean that she has given low JOLs to all items which she later remembers and high JOLs to all items which she does not remember.

Over the years, numerous studies have found that—in general—JOLs have a moderate accuracy (e.g., Dunlosky & Metcalfe, 2009). Regarding absolute accuracy, JOLs are typically well calibrated or slightly overconfident in a first study-test cycle (e.g., Ariel & Dunlosky, 2011; Koriat, 1997). This, however, significantly changes to a profound underconfidence when the same items are presented in a second study-test cycle (e.g., Koriat et al., 2002; Serra & Dunlosky, 2005). This *underconfidence-with-practice effect* has been attributed to people's failure to fully acknowledge how much they have learned from first to second study-test cycle (e.g., Koriat, 1997; Koriat et al., 2002). Apart from that, JOLs tend to be underconfident when they are made on an aggregate level than when they are made for each item (e.g., Mazzoni & Nelson, 1995).

Regarding relative accuracy, JOLs are typically improving in resolution over multiple study-test cycles (e.g., Finn & Metcalfe, 2008; Koriat et al., 2006). This means that from a second study-test cycle on, people are better in discriminating between items which they remember and which they do not remember when making JOLs (increased resolution), but those JOLs tend to be too low on average when compared with actual memory performance (decreased calibration). Besides, it has been found that delayed JOLs have a higher resolution than immediate JOLs (e.g., Rhodes & Tauber, 2011). Explanations for this *delayed-JOL effect* propose that when making JOLs after a certain amount of time, participants have access to information only from long-term memory, which improves predictive validity compared to JOLs made immediately after studying, which also rely on information from short-term memory that is less predictive of actual memory performance (e.g., Nelson & Dunlosky, 1991; Rhodes & Tauber, 2011).

Above and beyond, there have also been *systematic dissociations* between JOLs and memory performance. For example, people assign unduly high JOLs to words that are written in a larger font compared to words written in a smaller font, although font size affects memory performance to a much smaller extent (e.g., Chang & Brainerd, 2022; Halamish, 2018; Rhodes & Castel, 2008; Undorf & Zimdahl, 2019). In a similar vein, people also assign higher JOLs to words that are presented in a higher volume compared to words presented in a lower volume, but memory performance is not influenced by volume (e.g., Frank & Kuhlmann, 2017; Rhodes & Castel, 2009). Another example of a systematic dissociation between JOLs and memory performance is the *stimulus-size illusion* (Undorf et al., 2017). Stimuli that in the beginning are too small to be recognizable but gradually increase in size rather quickly receive higher JOLs compared to small stimuli that increase rather slowly in size. However, as in the previous examples, memory performance is unaffected by this manipulation.

Systematic dissociations between JOLs and memory performance shed light on the theoretical processes that underly metacognitive judgments. Most importantly, systematic dissociations are incompatible with the idea that people have direct access to their memory traces and use the different strengths of those traces to form their JOLs (e.g., Arbuckle & Cuddy, 1969). If this were true, then people's JOLs should not be influenced by, for example, font size, because the memory traces for words written in a larger font are comparable to those for words written in a smaller font, as indicated by similar actual memory performance for both kinds of fonts (e.g., Rhodes & Castel, 2008). More generally, direct-access accounts propose that JOLs and memory performance are well aligned throughout because they are both based on the same variable (i.e., the memory traces). However, the mere existence of systematic dissociations between JOLs and memory performance makes these accounts implausible. Therefore, it is not surprising that direct-access accounts have been dismissed as theoretical processes that underly JOLs (e.g., King et al., 1980; Rhodes, 2016).

Systematic dissociations between JOLs and memory performance are, however, compatible with the idea that people use cues and heuristics to infer their JOLs. These inferential accounts do not assume that people have direct access to their memory traces, but that they use information that is tied to the material or the circumstances of learning to make inferences about the memorability of the material (e.g., Koriat, 1997). Following this logic, systematic dissociations emerge because people either use invalid cues that are not predictive of memory performance or do not use valid cues that would actually be predictive of memory performance. Inferential accounts are able to explain, for example, the font-size illusion, by showing that people use the invalid cue 'font size' to base their JOLs on. Over the years, researchers have accumulated a great amount of evidence in favor of inferential accounts (e.g., Dunlosky & Tauber, 2016; Koriat, 2007).

2.2 The Cue-Utilization Approach to JOLs

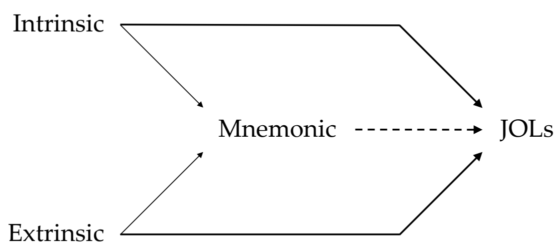
As one of the most prominent inferential accounts, Koriat (1997) introduced the *cue-utilization approach* to JOLs (see Figure 2). According to this approach, people use three different classes of cues to infer their JOLs. *Intrinsic cues* pertain to the material itself and include variables like semantic association (e.g., Mueller et al., 2013; Undorf & Erdfelder, 2015), concreteness (e.g., Begg et al., 1989; Witherby & Tauber, 2017), and emotionality (e.g., Witherby & Tauber, 2018; Zimmerman & Kelley, 2010). *Extrinsic cues* incorporate factors that pertain to the conditions of learning, like how often material has been presented during study (e.g., Undorf et al., 2018), how long the presentation time was (e.g., Jang & Nelson, 2005), and whether the material has been presented massed or distributed (e.g., Dunlosky & Nelson, 1994). Finally, *mnemonic cues* refer to internal factors that signal whether material has been learned or not. Examples for this class of cues are how easily specific material is processed during learning (e.g., Hertzog et al., 2003), how easily it is retrieved from memory (e.g., Koriat & Ma'ayan, 2005), or how familiar it seems (e.g., Metcalfe & Finn, 2008).

Above and beyond differentiating between different classes of cues, the cue-utilization approach (Koriat, 1997) also makes assumptions about the theoretical processes underlying JOLs (see Figure 2). It proposes that JOLs can either be inferred through *theory-based, analytic processes* or *experience-based, non-analytic processes* (see also Koriat & Levy-Sadot, 1999). Inferring JOLs through theory-based, analytic processes means that people have specific theories about how different cues affect their memory and use this knowledge to analytically form their JOLs. For example, people might have a certain idea about how the intrinsic cue 'font size' affects their memory and apply this idea deliberately when making JOLs. The result is that they assign higher JOLs to words that are written in a larger font compared to words that are written in a smaller font, because they have the belief that font size affects memory this way (e.g., Mueller et al., 2014).

Inferring JOLs through experience-based, non-analytic processes means that people's JOLs are based on their experiences or feelings during study or retrieval and that these are used to infer JOLs (e.g., Koriat, 2007). For example, people might feel that words written in a larger font are easier to read than words written in a smaller font and therefore assign higher JOLs to the former than the latter. In this case, the intrinsic cue 'font size' is used in a non-analytic way, because it is mediated through the mnemonic experience of fluency (e.g., Rhodes & Castel, 2008).

Previously, the contributions of fluency and beliefs have been investigated for a number of metacognitive illusions (for reviews, see Undorf, 2020; Yang et al., 2021). For some

Figure 2
Cue-Utilization Approach to JOLs by Koriat (1997)



Note. The cue-utilization approach differentiates between intrinsic, extrinsic, and mnemonic cues that are used to infer JOLs. Solid lines represent theory-based, analytic inferences, dashed lines represent experience-based, non-analytic inferences.

of these illusions, there is strong evidence that they are due to fluency. As an example, experiments on the stimulus-size illusion showed that clarification speed (i.e., whether a stimulus increases quickly or slowly in size) does not influence JOLs directly, but that this influence is mediated by the time that participants need to identify the stimulus (Undorf et al., 2017). Furthermore, post-experimental questionnaires revealed that participants did not detect any differences in clarification speed and showed that they did not have any metacognitive beliefs about the relation between clarification speed and memory performance. It is therefore more likely that the stimulus-size illusion is due to fluency rather than beliefs.

Other metacognitive illusions have been mainly attributed to the contributions of beliefs. The font-size illusion, for example, was originally thought of being due to fluency, because words written in a larger font should be easier to process than words written in a smaller font (Rhodes & Castel, 2008). However, subsequent studies showed that there is no difference in fluency as measured by response time in a lexical decision task for words printed in larger and smaller fonts (Mueller et al., 2014). This was further supported by Undorf and Zimdahl (2019) who demonstrated that the font sizes used in the aforementioned studies for larger (48 pt) and smaller words (18 pt) are both within the *fluent range of print size* (Legge & Bigelow, 2011). This range covers—given a certain distance from the reader to the screen—a certain amount of font sizes for which reading speed is constantly the highest, that is, fluency is maximized. For font sizes that lie below or above this fluent range of print size, reading speed is reduced. Undorf and Zimdahl (2019) varied font sizes between 6 pt and 500 pt and were able to show that fluency (again measured by response time in a lexical decision task) was lower for font sizes below and above the fluent range of print size compared to font sizes within the fluent range of print size. In contrast, JOLs increased monotonically with font size, indicating that the font-size illusion is not due to fluency. Most participants stated that they have a metacognitive belief that font size influences memory performance in a monotonically increasing way, making it likely that the font-size illusion is due to participants' beliefs about how font size affects actual memory performance.

There are also metacognitive illusions for which both contributions of fluency and beliefs have been found. Susser and Mulligan (2015), for example, discovered that people assign higher JOLs to words that they write with their dominant hand compared to words that they write with their non-dominant hand (*motoric fluency illusion*). The researchers showed that the time it took participants to write the words was shorter for words written with the dominant compared to the non-dominant hand, indicating that fluency underlies this illusion. However, in a later study, they also found higher pre-study JOLs for words written with the dominant than with the non-dominant hand (Susser et al., 2017). Pre-study JOLs are assessed before the experiment starts and therefore exclude any form of experience-based processes like fluency (e.g., Castel, 2008). This result was interpreted as evidence that beliefs also play a role in the motoric fluency illusion. In sum, research on metacognitive illusions has shown that JOLs rely on theory-based, analytic processes, experience-based, non-analytic processes, or a combination of both (e.g., Undorf et al., 2022; Yang et al., 2021).

That judgments are assumed to be based on two qualitatively different modes of information processing is by far not limited to JOLs. In fact, it is a characteristic of many dual-process theories that have been proposed for other kinds of judgments (for an overview, see Chaiken & Trope, 1999). To shed light on the theoretical processes underlying JOLs, it is therefore highly advisable to use dual-process theories as a framework to integrate findings from metacognitive research on the formation of JOLs. Before I do this, I will first provide a short review of dual-process theories in general and highlight which parts of the dual-process framework are useful for research on metamemory.

2.3 Dual-Process Theories

Dual-process theories have been proposed as a framework to understand human cognition long before Daniel Kahneman published his widely popular book *Thinking, Fast and Slow* (Kahneman, 2011). Yet, his review underlined the amount of scientific research and psychological phenomena for which dual-process theories have been used as explanations. Although some of the findings that he reviewed have to be read with caution as they have proven to be hard to replicate (see, e.g., Open Science Collaboration, 2015), the general idea of dual-process theories remains valid and helpful.

Dual-process theories assume that human cognition is based on two qualitatively different kinds of information processing. These two different processes have been described in various theories using different labels for the two processes (for overviews, see Chaiken & Trope, 1999; Stanovich & West, 2000). However, because the two processes share common features across theories, they have been summarized under the terms *System 1* and *System 2* (e.g., Kahneman, 2003; Stanovich & West, 2000).

According to Kahneman (2003) the core characteristics of the two systems can be defined as follows:

> The operations of System 1 are typically fast, automatic, effortless, associative, implicit (not available to introspection), and often emotionally charged; they are also governed by habit and are therefore difficult to control or modify. The operations of System 2 are slower, serial, effortful, more likely to be consciously monitored and deliberately controlled; they are also relatively flexible and potentially rule governed. (p. 698)

It is important to state that dual-process theories vary largely in their specificity and their ability to explain and predict certain phenomena, which is also the main target of criticism towards them. According to Gawronski and Creighton (2013), there are three classes of dual-process theories: *phenomenon-specific*, *generalized*, and *formalized* ones. Phenomenon-specific dual-process theories focus on particular phenomena such as prejudice and stereotyping (e.g., Devine, 1989) or impression formation (e.g., Fiske & Neuberg, 1990), but have been criticized for having only limited explanatory value (e.g., Kruglanski et al., 2007). In contrast, generalized dual-process theories such as the distinction between System 1 and System 2 (e.g., Kahneman, 2003; Stanovich & West, 2000) run the risk of lacking conceptual precision, which can make them too unspecific to produce testable hypotheses (e.g., Keren & Schul, 2009). Lastly, formalized dual-process theories such as the *process-dissociation procedure* (Jacoby, 1991) have the ability to mathematically determine the contributions of both kinds of processes to certain phenomena, but their parameters and the interpretation thereof are often subject of controversial debate (e.g., Buchner et al., 1997; Buchner et al., 1995). Regardless of this criticism, there is less dispute about the conclusion that dual-process theories can be useful as a *framework* to understand human cognition.

Regarding JOLs, dual-process theories have been used as a framework to understand metacognition mainly in two different ways. First, as described above, the dual-basis view of metacognition (Kelley & Jacoby, 1996; Koriat, 1997) has stimulated a bulk of research suggesting that metacognitive judgments rely on experience-based, non-analytic processes (System 1) and theory-based, analytic processes (System 2). Second, it has been tested whether JOLs can predict automatic influences and recollection processes of memory (*dual-process theory of memory*, cf. Mandler, 2008; Yonelinas, 2002). For example, Undorf et al. (2016) used the process-dissociation procedure (Jacoby, 1991) to show that JOLs reflect both processes when predicting memory performance. Kuhlmann and Undorf (2018) replicated and extended this finding by demonstrating that JOLs for older adults mirror memory performance that showed an age-related decline in recollection but not in automatic influences. These results underline that applying a dual-process perspective on JOLs can provide meaningful insights into the basis of people's metamemory. The current thesis therefore uses dual-process theories as a framework to investigate the contributions of System 1 and System 2 to cue effects on JOLs.

3 Cues and Processes Underlying Metamemory

> "When people feel that they know something,
> it is very likely that they do know it,
> and when they feel that they do not,
> it is likely that they do not"
>
> Joseph T. Hart (1967)

As has become evident from Chapter 2, people do not have direct access to their metamemory, but use cues to infer their metacognitive judgments like JOLs (e.g., Koriat, 1997). People's metamemory is accurate when they base their metacognitive judgments on cues that are valid with regard to correctly predicting actual memory performance. In contrast, people's metamemory is inaccurate when they base their metacognitive judgments on cues that are not predictive of actual memory performance or fail to incorporate valid cues (see Undorf et al., 2022).

In this chapter, I uncover two new cues that people use to infer their JOLs that have not been investigated before: visual coherence and outcome knowledge. Uncovering new cues for JOLs is an important endeavor for understanding how people assess their metamemory. Detecting valid and invalid cues that have not been considered before can help learners to improve their study process and prevent them from falling for metacognitive illusions.

However, to understand how people assess their metamemory, it is not sufficient to examine what cues people use to infer their metacognitive judgments. Therefore, I also show *how* visual coherence and outcome knowledge impact JOLs. In line with dual-process theories, I investigate the contributions of experience-based, non-analytic and theory-based, analytic processes to the influence of visual coherence and to the influence of outcome knowledge on JOLs. Examining the underlying processes facilitates a better understanding of how people use different cues and can help to integrate findings about the basis of metacognitive judgments like JOLs.

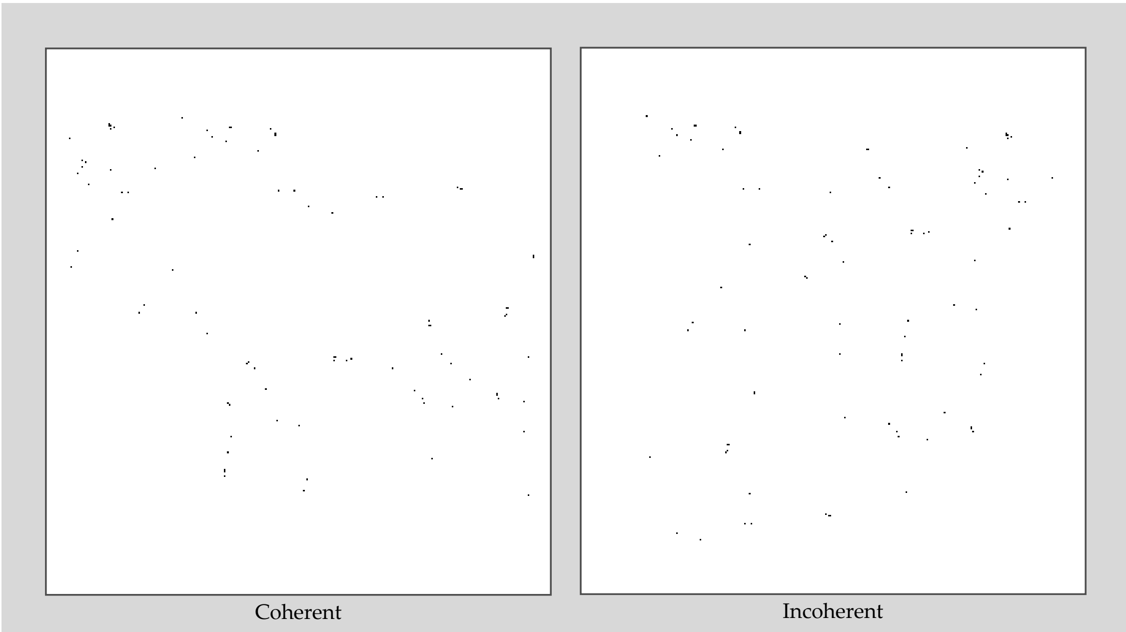
3.1 Visual Coherence & Experience-Based and Theory-Based Processes

Zimdahl, M. F., & Undorf, M. (2022a). *Visual coherence impacts judgments of learning: Evidence for the dual-basis view of metacognition* [Manuscript submitted for publication]. Department of Psychology, University of Mannheim.

Visual coherence refers to the fact whether a dot pattern depicts a certain gestalt or not. To illustrate this, consider the two different stimuli in Figure 3. Both pictures contain the same visual information with regard to the number of dots they are displaying. However, the left picture is a degraded version of a real object, in this case a moose. It is therefore visually coherent as the allegedly random dot pattern forms a gestalt. The right picture does not form a gestalt and is therefore visually incoherent.

Previous research has shown that people have the ability to discriminate between coherent and incoherent pictures even when they are unable to identify what the coherent picture represents (e.g., Bolte & Goschke, 2008). It has been argued that these intuitive judgments of coherence are based on people's feelings of fluency during per-

Figure 3
Examples of Stimuli Used in Zimdahl and Undorf (2022a)



Note. The left picture is a degraded version of a real object (moose) and therefore coherent. The right picture does not form a gestalt and is therefore incoherent. Importantly, both pictures contain the same amount of visual information with regard to the number of dots they are displaying.

ception (e.g., Wippich et al., 1994). This was supported by Topolinski and Strack (2009) who were able to show that increasing feelings of fluency by repeated exposure of the pictures increased the likelihood that the participants judged the pictures to be visually coherent.

The six experiments described in Manuscript I (Zimdahl & Undorf, 2022a) were designed to investigate whether visual coherence affects JOLs. If visual coherence increases fluency and fluency increases JOLs, then there is good reason to expect that JOLs should be higher for coherent than for incoherent pictures. We tested this assumption not only by manipulating the fluency of the pictures (Experiment 2), but also by assessing fluency with three different objective measures (Experiments 3-5) as well as one subjective measure (Experiment 6). Furthermore, we conducted mediational analyses to investigate whether possible effects of visual coherence on JOLs were in fact due to differences in fluency.

Additionally, we asked participants about whether they had noticed any differences in the pictures. With this, we were able to assess metacognitive beliefs and compare JOL patterns between those participants who were aware and those participants who were unaware of differences in visual coherence. This enabled us to investigate the contributions of experience-based, non-analytic and theory-based, analytic processes to an effect of visual coherence on JOLs, thus testing the dual-basis view of metacognition (Kelley & Jacoby, 1996; Koriat, 1997).

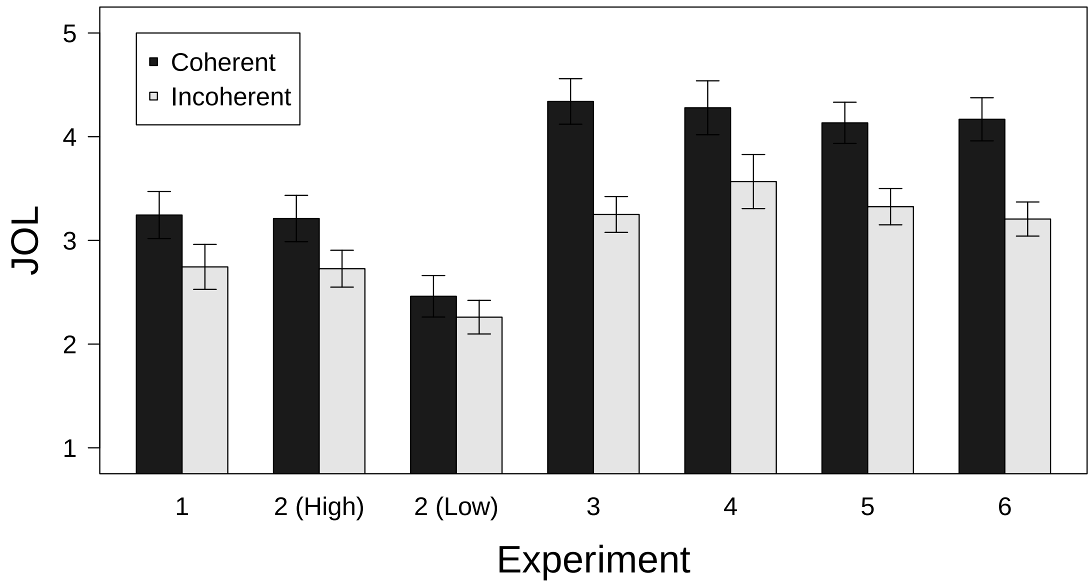
All six experiments contained an initial phase, in which participants were shown 15 coherent and 15 incoherent pictures and made JOLs (on a scale from 1-10) for each picture. Afterwards, we assessed metacognitive beliefs by asking participants whether they had noticed any form of gestalt in the pictures and if so, on how many pictures and how they looked like. Then, after a short filler task, participants administered a recognition test in which the 30 already shown pictures were intermixed with 30 new pictures (15 coherent/15 incoherent).

Experiment 1 (N = 40) showed that JOLs were higher for coherent than for incoherent pictures (see Figure 4). The same was true for actual memory performance (as measured by corrected hit rates $P_r = hits - false\ alarms$, Snodgrass & Corwin, 1988). This demonstrated that participants use visual coherence as a cue to form their JOLs and that visual coherence is valid in predicting actual memory performance.

Experiment 2 (N = 44) was designed to test the contribution of fluency to the visual coherence effect on JOLs. Therefore, we varied the figure-ground contrast of the pictures (high/low) to induce differences in perceptual fluency (cf. Topolinski & Strack, 2009). As expected, we found main effects of contrast and fluency, showing that both factors influence JOLs (see Figure 4). More importantly, a Contrast $\times$ Fluency interaction revealed that the visual coherence effect was absent in low-contrast pictures, indicating

Figure 4

Visual Coherence Effect on JOLs (Zimdahl & Undorf, 2022a)



Note. The figure shows judgments of learning (JOLs) for coherent and incoherent pictures in Experiments 1-6 (Zimdahl & Undorf, 2022a). In Experiment 2, coherent and incoherent pictures were shown with a high or a low figure-ground contrast.

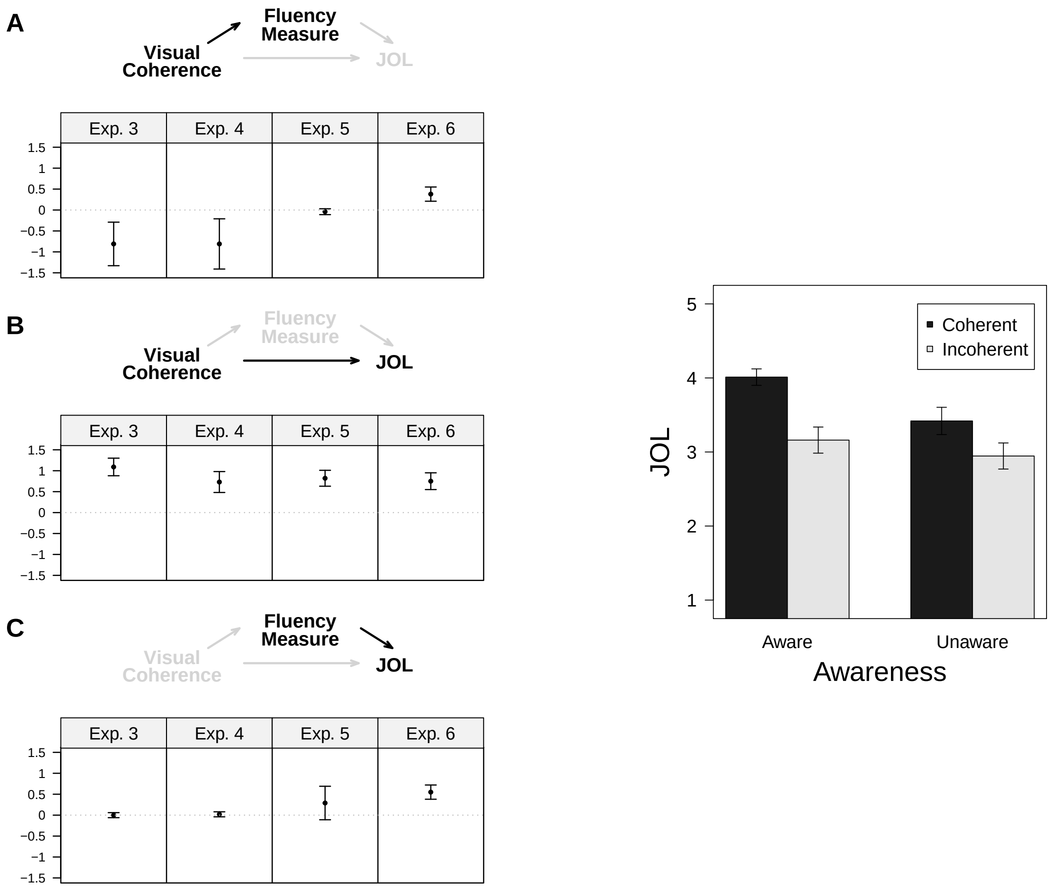
that some fluency is necessary for visual coherence to influence JOLs. However, because this was only indirect evidence for fluency contributions to the visual coherence effect on JOLs, we conducted another four experiments where we used independent measures of fluency. Furthermore, we determined the contribution of fluency to the visual coherence effect on JOLs with mediational analyses.

In Experiment 3 (N = 49), we used self-paced study time as an independent measure of fluency (cf. Koriat & Bjork, 2006; Undorf & Erdfelder, 2011, 2015). This was different from the first two experiments insofar as the time that pictures were shown in the study phase was not fixed to 1 s but was determined by the participants. Results showed that self-paced study time was shorter for coherent than for incoherent pictures, indicating higher fluency for coherent than for incoherent pictures. JOLs (see Figure 4) and memory performance were also higher for coherent than for incoherent pictures.

To investigate whether the visual coherence effect on JOLs is mediated by fluency, we first ran two multilevel regression models (level 1: items, level 2: participants; Kenny et al., 2003; Krull & MacKinnon, 2001) to determine the direct effects of visual coherence on fluency and JOLs as well as the direct effect of fluency on JOLs. Figure 5 (left panel) shows the results of the regression analyses. As can be seen in Panel A, visual coherence was inversely related to fluency for Experiment 3, confirming that visual coherence reduced self-paced study time. Panel B shows that the direct effect of visual coherence on JOLs was positive, indicating that visual coherence increased JOLs when self-paced study time was controlled for. However, there was no direct effect of self-paced study time on JOLs when visual coherence was controlled for (Panel C). Consequently, a mediational analysis revealed that the indirect effect of visual coherence on JOLs through self-paced study time was not significant, confirming that the visual coherence effect on JOLs was not mediated by fluency.

Figure 5

Contributions of Fluency and Beliefs to the Visual Coherence Effect on JOLs



Note. The left panel of the figure shows unstandardized regression coefficients of direct effects for Experiments 3-6 in Zimdahl and Undorf (2022a). Mediational analyses revealed that the indirect effect of visual coherence on JOLs was only significant for Experiment 6, where a subjective fluency measure was used. The right panel of the figure presents the findings from the joint analysis of all experiments reported in Zimdahl and Undorf (2022a). It shows that the visual coherence effect on JOLs is more pronounced in people who were aware of differences across pictures, thus revealing that beliefs also contribute to the visual coherence effect on JOLs.

Because self-paced study had been criticized as a measure of fluency (e.g., Witherby & Tauber, 2017; Yang et al., 2018; Yang et al., 2021), Experiment 4 (N = 43) used a continuous identification task as an independent measure of fluency (e.g., Sanborn et al., 2004; Yang et al., 2018). In this task, a stimulus and a mask are shown in an alternating way until the participants stops the process. In the beginning, the stimulus is shown very briefly (17 ms) and is immediately followed by the mask (983 ms). Subsequently, the duration of the stimulus increases in every sequence (34 ms, 51 ms, 67 ms, ...), whereas the duration of the mask decreases (966 ms, 949 ms, 932 ms, ...). Participants are asked to stop the process as soon as they think they have seen the stimulus long enough to be able to recognize it in the later memory test.

Replicating Experiment 3, the time that participants studied the pictures was shorter for coherent than for incoherent pictures, again indicating that fluency was higher for coherent than for incoherent pictures. JOLs (see Figure 4) and memory performance were also higher for coherent than for incoherent pictures. The results of the regression and mediation analyses replicated those of Experiment 3 (see Figure 5). As expected, visual coherence was inversely related to the fluency measure (identification time, Panel A). Furthermore, visual coherence increased JOLs when identification time was controlled for (Panel B). However, again, there was no direct effect of fluency on JOLs when visual coherence was controlled for (Panel C). An non-significant indirect effect of visual coherence on JOLs through identification time confirmed that fluency did not mediate the visual coherence effect on JOLs.

One possible explanation for the result that fluency did not mediate the visual coherence effect on JOLs was that participants in Experiments 3 and 4 did not make intuitive judgments anymore based on quick feelings about how fluently they perceived the pictures. In fact, the times that participants took to study the pictures in Experiments 3 and 4 were much longer (self-paced study time: M = 7.16 s; identification time: M = 11.43 s) than the fixed presentation times in Experiments 1 and 2 (1 s). Therefore, we restricted self-paced study time to 5 s in Experiment 5 (N = 39).

However, although we were able to replicate the visual coherence effect on JOLs (see Figure 4) and memory performance, we did not find differences in restricted self-paced study time between coherent and incoherent pictures. This was confirmed by regression analyses (see Figure 5), showing that visual coherence was not related to restricted self-paced study time (Panel A). The other effects replicated the previous experiments. Visual coherence increased JOLs when restricted self-paced study time was controlled for (Panel B), fluency was not related to JOLs when visual coherence was controlled for (Panel C), and the indirect effect of visual coherence on JOLs through restricted self-paced study time was not significant. This suggested that objective measures might have difficulties in assessing fluency contributions in quick and intuitive judgments.

We therefore conducted a final experiment (N = 43), in which we used a subjective measure to assess the perceived fluency of the pictures. This was done in a separate phase before the study phase by showing the participants each picture for 1 s and asking them directly how easy it was for them to perceive the picture. Previous research (Graf et al., 2018) had shown that this is sufficient to capture subjective fluency (see also Reber et al., 2004). Presentation times of the pictures in the study phase were again set to 1 s. Besides again replicating the visual coherence effect on JOLs (see Figure 4) and memory performance, subjective fluency ratings were also higher for coherent than for incoherent pictures, showing that participants expressed that they had perceived coherent pictures more easily than incoherent pictures.

Regression analyses (see Figure 5) in this experiment revealed that, as expected, visual coherence was positively related to subjective fluency (Panel A). Again, visual coherence increased JOLs when subjective fluency was controlled for (Panel B). More importantly—and differing from all previous experiments—fluency had a positive direct effect on JOLs when visual coherence was controlled for (Panel C). Furthermore, a mediational analysis revealed that the indirect effect of visual coherence on JOLs through subjective fluency was significant, indicating that subjective fluency partially mediated the visual coherence effect on JOLs.

Experiment 6 therefore deviated from the previous experiments in showing that subjective, but not objective fluency contribute to the visual coherence effect on JOLs. This was probably due to the different presentation times of the pictures in the study phase of the experiments. To be able to capture quick and intuitive feelings of fluency, the self-paced study time and identification time in Experiments 3 and 4 were most likely too long compared to the relatively short presentation time in Experiment 6. Consequently, objective measures of fluency turned out to have difficulties in assessing quick and intuitive feelings of fluency. However, with a subjective measure of fluency we were able to find contributions of experience-based, non-analytic processes to the visual coherence effect on JOLs.

Regarding the contributions of theory-based, analytic processes to the visual coherence effect on JOLs, we assessed metacognitive beliefs in each of the six experiments by asking participants whether they had noticed any form of gestalt in the pictures and if so, on how many and how they looked like. With this, we could compare JOL patterns between people who were aware of differences across pictures and people who were unaware. In Experiments 1-5, the visual coherence effect on JOLs did not significantly differ between those participants who were aware and those participants who were unaware of difference in visual coherence. In contrast, Experiment 6 showed a more pronounced visual coherence effect in people who were aware of differences in visual coherence. However, the number of participants who were unaware of differences in visual coherence was rather small in every experiment, allowing only tentative conclusions about the contribution of beliefs to the visual coherence effect on JOLs.

Therefore, we conducted a joint analysis of all six experiments which provided us with more power to test the contribution of beliefs to the visual coherence effect on JOLs. When we compared all participants who were aware of differences across pictures (n = 185) with all participants who were unaware (n = 73), we found a significant interaction between coherence and awareness (see Figure 5, right panel). This indicated that the visual coherence effect on JOLs was more pronounced in people who were aware of differences across pictures, thus revealing a contribution of beliefs to the visual coherence effect on JOLs.

Taken together, the six experiments reported in MANUSCRIPT I (Zimdahl & Undorf, 2022a) uncovered visual coherence as a new cue that people use to form their JOLs. All experiments exhibited a robust visual coherence effect with higher JOLs for coherent than for incoherent pictures. Furthermore, visual coherence was found to be a valid cue, as memory performance was also higher for coherent than for incoherent pictures in five of six experiments (lack of differences in Experiment 2 was possibly due to very low levels of memory performance for this experiment). Using mediational analyses, we showed that fluency contributes to the visual coherence effect on JOLs which was demonstrated with a subjective fluency measure but not with objective fluency measures. Moreover, a joint analysis of all six experiments revealed that metacognitive beliefs also contribute to the visual coherence effect on JOLs. The results therefore support the dual-basis view of metacognition (Kelley & Jacoby, 1996; Koriat, 1997).

3.2 Outcome Knowledge

Zimdahl, M. F., & Undorf, M. (2021). Hindsight bias in metamemory: Outcome knowledge influences the recollection of judgments of learning. *Memory, 29*(5), 559-572. https://doi.org/10.1080/09658211.2021.1919144

In his doctoral dissertation, Baruch Fischhoff (1975) discovered that people remembered being better in predicting certain event outcomes after they knew what the outcome was. This laid the foundations for a rich and productive research program on the influence of outcome knowledge on the recollections of prior judgments, which is better known as *hindsight bias*. Today, hindsight bias is one of the most studied biases in psychology (e.g., Roese & Vohs, 2012) and has proven to be extremely robust and widespread (for reviews, see Christensen-Szalanski & Willham, 1991; Guilbault et al., 2004; Pohl & Erdfelder, 2022).

At this point, it is important to state that hindsight bias is by no means restricted to event outcomes. It has been shown to affect people when they had to judge the truth of different statements (e.g., Campbell & Tesser, 1983; Hoffrage et al., 2000), provide numerical estimates for unknown quantities (e.g., Erdfelder & Buchner, 1998; Hell et al., 1988), or state their confidence in given answers (e.g., Ackerman et al., 2020). What is common to all those judgments is that they are initially made in a state of uncertainty. However, knowledge about the correct outcome removes this uncertainty and makes it hard for people to remember their previous judgments prior to knowing the correct outcome (e.g., Erdfelder et al., 2007; Pohl & Erdfelder, 2022; Pohl & Hell, 1996).

Despite the obvious similarities between judgments that have been used to investigate hindsight bias and JOLs, no study has ever examined whether JOLs are also affected by the influence of outcome knowledge. Yet, this is a highly relevant question, especially for the process of learning: If someone remembers in hindsight giving more accurate judgments in foresight, she might be unable to understand what has led to her faulty judgments in the first place and how her recollected judgments deviate from her original ones. Or, to borrow Fischhoff's (1975) words: "The very outcome knowledge which gives us the feeling that we understand what the past was all about may prevent us from learning anything from it" (pp. 298-299).

Therefore, Manuscript II (Zimdahl & Undorf, 2021) was designed to investigate whether people use outcome knowledge as a cue to form their JOLs, hence, if there is hindsight bias in metamemory. To be more precise, the three experiments described below aimed at examining the influence of outcome knowledge on the *recollection* of JOLs. This distinction is important as hindsight bias is concerned with how biased people remember their naïve judgments that they made prior to knowing the correct outcome. The current experiments therefore extended previous research on hindsight bias by investigating whether outcome knowledge influences the recollection of JOLs.

Outcome knowledge has been demonstrated to influence new JOLs in studies using multiple study-test cycles on the same material (e.g., Finn & Metcalfe, 2008; Serra & Ariel, 2014). A common finding in these studies is that people remember their performance from a past test and use this information to form new JOLs (*memory-for-past-test heuristic*, Ariel & Dunlosky, 2011; Finn & Metcalfe, 2007). However, it was an open question whether outcome knowledge also influences the recollection of prior made JOLs.

To test this, participants in Experiment 1a (N = 54) of Manuscript II (Zimdahl & Undorf, 2021) were asked to study 60 weakly related word pairs and made JOLs for each of them. After a short filler task, participants engaged in a cued recall test where they were presented with the first word of each word pair and were asked to enter the second word. They received feedback about whether they were correct and were instructed to remember their JOLs from the first phase of the experiment. Results showed that recollected JOLs (RJOLs) were closer to the correct outcome than original JOLs (OJOLs), thus revealing hindsight bias on JOLs.

We were able to demonstrate hindsight bias on JOLs using three different measures. First, a significant interaction between JOL type (OJOL, RJOL) and recall success (yes, no) showed that the difference in JOLs between recalled and not-recalled word pairs was larger for RJOLs than for OJOLs. Second, resolution (as measured by within-subjects Goodman-Kruskal gamma correlations between JOLs and recall performance, Goodman & Kruskal, 1954) was higher for RJOLs than for OJOLs. Third, we obtained a significantly positive ΔHB as a standard measure of hindsight bias magnitude (Pohl, 1992). ΔHB compares how much on average the OJOLs and the RJOLs are away from the correct outcome and is significantly positive when the mean difference between RJOLs and the correct outcome is smaller than the mean difference between OJOLs and the correct outcome. With this, we combined measures from metamemory and hindsight bias literature to show converging evidence for hindsight bias on JOLs.

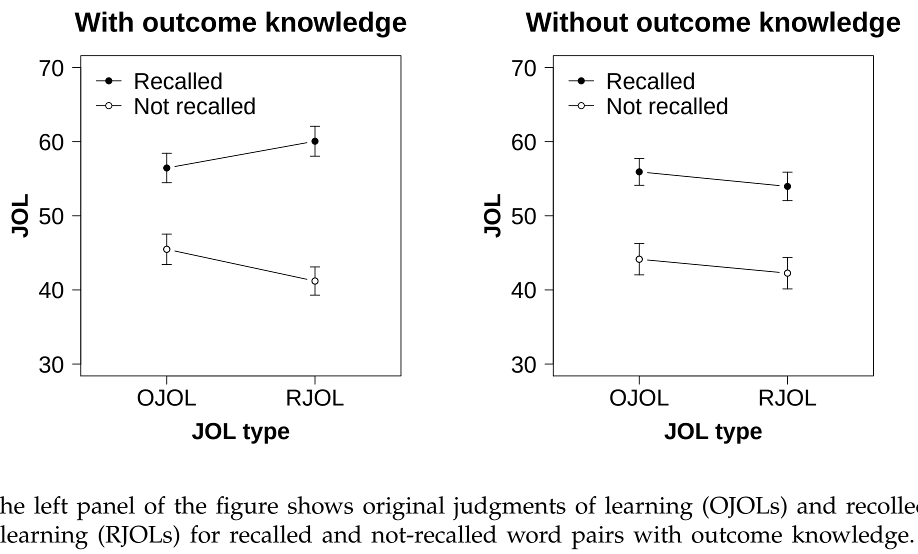
Experiment 1b (N = 49) replicated hindsight bias on JOLs with a non-student sample using the same measures as in Experiment 1a. This time, the sample consisted of Amazon Mechanical Turk workers who—divergent from the previous experiment—were paid for their participation and performed the experiment online. Despite these differences between experiments, hindsight bias on JOLs turned out to be robust. This experiment therefore did not only fulfill the important purpose of replicating the newly found hindsight bias on JOLs (see Asendorpf et al., 2013, for an excellent review on the significance of replication for good science), but also showed the generalizability of hindsight bias on JOLs across different samples and methods of conducting an experiment. Furthermore, it demonstrated that recruiting participants via online platforms such as Amazon Mechanical Turk is worth considering for obtaining valid experimental results (see also Crump et al., 2013), especially when conducting research in the laboratory is not an option (e.g., in the case of a global pandemic).

Although Experiments 1a and 1b both provided initial evidence for hindsight bias on JOLs, they also left room for alternative explanations why RJOLs were biased towards the correct outcome. Because participants recollected their JOLs after receiving feedback about whether they were right or wrong for every word pair, it was unclear whether JOL recollection would also be biased when people do not have outcome knowledge. Therefore, in Experiment 2 (N = 96), which we additionally preregistered at the Open Science Framework (https://osf.io/nehf2), we asked participants to recollect their JOLs in a separate phase after trying to recall half of the word pairs. With this, participants had outcome knowledge for some, but not all items when recollecting their JOLs.

The results of Experiment 2 underlined the crucial role of outcome knowledge in producing hindsight bias on JOLs. As can be seen in Figure 6, RJOLs were only biased towards the correct outcome for word pairs with outcome knowledge (left panel), but not for word pairs without outcome knowledge (right panel). All three measures mentioned above confirmed that hindsight bias on JOLs was limited to word pairs that had been tested before the JOL recollection (i.e., with outcome knowledge). This demonstrated that RJOLs were not biased towards the correct outcome when people do not have outcome knowledge. Furthermore, it showed that when people have outcome knowledge about whether they were right or wrong at recalling the word pairs, they use it as a cue to recollect their JOLs. It is therefore the first study showing that outcome knowledge influences the recollection of JOLs.

This finding is important insofar as it has implications for the process of learning (for a detailed description of this, see Manuscript II and Section 4.1). When people recollect their JOLs biased towards the correct outcome after they have been tested, then they think they had been better in predicting their memory performance than they actually were. This prevents them from recognizing that their OJOLs were less accurate than they think they were. Consequently, they might not sufficiently engage in identifying causes for their faulty judgments that allows them to improve their JOLs in the future. In this case, hindsight bias hinders learning, as it has been described for other judgments as well (see Fischhoff, 1975).

Figure 6
Hindsight Bias on JOLs (Zimdahl & Undorf, 2021, Experiment 2)



Note. The left panel of the figure shows original judgments of learning (OJOLs) and recollected judgments of learning (RJOLs) for recalled and not-recalled word pairs with outcome knowledge. The right panel of the figure represents OJOLs and RJOLs for recalled and not-recalled word pairs without outcome knowledge. RJOLs for word pairs with outcome knowledge are closer to the correct outcome, thus demonstrating hindsight bias on JOLs. RJOLs for word pairs without outcome knowledge are not biased, thus showing that hindsight bias on JOLs is limited to word pairs with outcome knowledge.

However, one could also argue that hindsight bias on JOLs fosters learning. As described earlier, metacognitive judgments are the basis on which people decide how to control their learning process. RJOLs that are more accurate than OJOLs have the potential to attribute cognitive resources more appropriately, thus making learning more efficient. Instead of focusing on prior faulty judgments, integrating the outcome knowledge into their JOLs allows learners to spend more time on material that is not learned well enough and stop concentrating on material that has been accomplished in the test. In this case, hindsight bias is viewed as a by-product of learning, leading to an improvement in predicting actual memory performance (see also Ackerman et al., 2020).

To test the robustness of hindsight bias on JOLs, we conducted another pre-registered experiment (https://osf.io/mrkzq), in which we manipulated the levels of OJOLs. In Experiment 3 (N = 212), we told participants prior to the study phase either that people find the memory test to be easy and recall 80% of the word pairs (high-anchor group), or that they find the task to be difficult and recall 20% of the word pairs (low-anchor group). With this, we were able to investigate whether hindsight bias on JOLs is limited to a certain range of JOLs. This was important because OJOLs located around the middle of the scale (as in the previous experiments) allowed RJOLs for recalled and not-recalled word pairs to be maximally biased towards the correct outcome.

The results of Experiment 3 again showed clear hindsight bias on JOLs as indicated by a significant interaction between JOL type and recall success, higher resolution of RJOLs than of OJOLs, and a significantly positive ΔHB. More importantly, hindsight bias appeared in both anchoring groups, showing that it is not limited to JOLs located around the middle of the scale. As in the previous experiments, hindsight bias on JOLs was only evident in word pairs with outcome knowledge, demonstrating that this is necessary to produce a biased recollection of JOLs.

Taken together, Manuscript II (Zimdahl & Undorf, 2021) uncovered outcome knowledge as a cue that people use to recollect their JOLs. It therefore extended research on the influence of outcome knowledge on JOLs (e.g., Ariel & Dunlosky, 2011; Finn & Metcalfe, 2007; Serra & Ariel, 2014) by showing that it also influences people's memory for prior made JOLs. Furthermore, it extended research on hindsight bias (e.g., Pohl & Erdfelder, 2022; Roese & Vohs, 2012), demonstrating that outcome knowledge affects metamemory judgments in a similar way than judgments about the external world.

However, at this point, the theoretical basis of hindsight bias on JOLs was unclear. Because hindsight bias on JOLs can have consequences for learning, it is crucial to understand what processes contribute to this effect or whether it can be prevented. Manuscript III (Zimdahl & Undorf, 2022b) was therefore designed to investigate whether hindsight bias on JOLs can be reduced by using two different debiasing methods: *warnings* and *incentives*.

3.3 Automatic and Controlled Processes

Zimdahl, M. F., & Undorf, M. (2022b). *Warnings and incentives do not reduce hindsight bias on judgments of learning* [Manuscript submitted for publication]. Department of Psychology, University of Mannheim.

The experiments reported in Manuscript II (Zimdahl & Undorf, 2021) revealed that outcome knowledge influences the recollection of prior made metacognitive judgments, thereby leading to hindsight bias on JOLs. While hindsight bias could be demonstrated for a variety of judgments, this study was the first to show that it also affects judgments of memorability.

Manuscript III (Zimdahl & Undorf, 2022b) examined whether hindsight bias on JOLs can be reduced by warnings and incentives. This has the potential to shed light on the theoretical processes that underly this bias. As described earlier, the dual-basis view of metacognition (Kelley & Jacoby, 1996; Koriat, 1997) assumes that both experience-based, non-analytic processes and theory-based, analytic processes contribute to JOLs. This is in line with other dual-process theories that propose two qualitatively different modes of information processing (see Section 2.3).

Within this framework, one core distinction between System 1 and System 2 is whether processes are assumed to occur in an automatic or a controlled fashion (Bargh, 1994; Kahneman, 2003). Consequently, if warnings and incentives do not reduce hindsight bias on JOLs, this would provide evidence for the conclusion that hindsight bias on JOLs is robust and hard to avoid. Hence, it would support automatic processes as explanation for hindsight bias on JOLs. If, on the other hand, warnings and incentives reduce hindsight bias on JOLs, this would provide evidence for the conclusion that hindsight bias on JOLs can be prevented and might be due to lack of people's awareness or motivation to avoid it. Hence, this would support controlled processes as explanation for hindsight bias on JOLs.

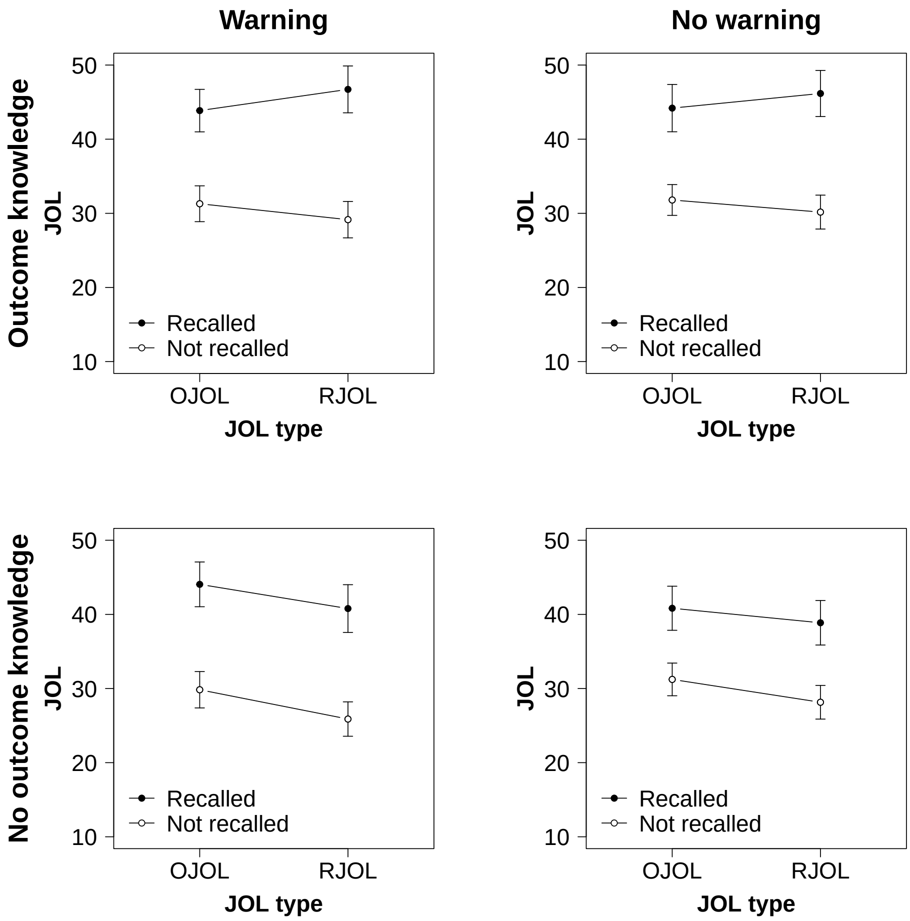
Previously, warnings and incentives have been used to reduce hindsight biases on judgments other than JOLs (e.g., Fischhoff, 1977; Hell et al., 1988; Pohl & Hell, 1996). Most of these attempts were not successful in reducing hindsight bias, indicating that hindsight bias on judgments about the external world is robust. In contrast, biases on JOLs other than hindsight bias could be reduced using different debiasing methods (see, e.g., Undorf et al., 2022). Especially warnings were found to be capable of reducing biases on JOLs, such as the font-size illusion (Rhodes & Castel, 2008). It was therefore an open question whether warnings and incentives reduce hindsight bias on JOLs.

All three experiments reported in Manuscript III (Zimdahl & Undorf, 2022b) used the same procedure as Experiments 2 and 3 of Manuscript II (Zimdahl & Undorf, 2021). Participants learned 60 word pairs and made JOLs for each of them. Afterwards, participants administered the first half of the cued recall test and were then asked to recollect their JOLs. The second half of the memory test concluded the experiments.

In Experiment 1 (N = 82), one group of participants (n = 41) received a detailed description of hindsight bias on JOLs prior to the JOL recollection phase. Additionally, this group was instructed to avoid hindsight bias when recollecting the JOLs. Another

Figure 7

Hindsight Bias on JOLs for the Warning and No-Warning Group (Zimdahl & Undorf, 2022b, Experiment 1)



Note. The two left panels show original judgments of learning (OJOLs) and recollected judgments of learning (RJOLs) for recalled and not-recalled word pairs in the warning group. The two right panels represent OJOLs and RJOLs for recalled and not-recalled word pairs in the no-warning group. Word pairs with outcome knowledge (top panels) show hindsight bias on JOLs that is similar across groups. Word pairs without outcome knowledge (bottom panels) do not show hindsight bias.

group of participants (n = 41) did not receive any information about hindsight bias on JOLs during the experiment. As can be seen in Figure 7, participants in both groups exhibited hindsight bias on JOLs for word pairs with outcome knowledge only, replicating the findings from Zimdahl and Undorf (2021). Importantly, hindsight bias on JOLs was not reduced in the warning group compared to the no-warning group.

To provide evidence for the conclusion that hindsight bias on JOLs was similar for the warning and the no-warning group, we used three different analyses. First, we compared the hindsight bias measures (interaction between JOL type and recall success, resolution, and ΔHB) across groups and found no differences in the respective ANOVA and t-tests. Second, we corroborated our findings with Bayesian analyses, revealing that there was more evidence for the null hypothesis assuming similar hindsight biases across groups. Third, because reduced hindsight bias on JOLs could also result in more correct recollections of JOLs, we also compared those across groups. Again, frequentist as well as Bayesian analyses showed that hindsight bias on JOLs was not reduced for the warning group when looking at correct recollections of JOLs.

Experiment 1 therefore supported automatic processes as explanation of hindsight bias on JOLs. Although participants were aware of hindsight bias on JOLs and instructed to avoid it, outcome knowledge still influenced recollections of JOLs in the same way as for participants who had no information about hindsight bias on JOLs. However, there was still the possibility that participants were simply not motivated to avoid hindsight bias on JOLs. We therefore conducted another two experiments to investigate whether an extrinsic motivation manipulation reduce hindsight bias on JOLs.

In Experiments 2 (N = 52) and 3 (N = 52), participants were incentivized for correct JOL recollections with monetary bonuses. We assigned half of the word pairs in each experiment a low value of 1 point, and the other half of the word pairs a high value of 10 points (Experiment 2) or 30 points (Experiment 3). At the end of the experiments, point values were converted into monetary bonuses and the five participants with the most accurate JOL recollections (i.e., highest point values) received the bonus additional to their usual compensation. Participants were informed about this procedure immediately before recollecting their JOLs. A post-experimental questionnaire in Experiment 3 confirmed that participants' motivation to correctly recollect JOLs was higher for high- than for low-value items.

Both experiments revealed similar hindsight biases on JOLs for high- and low-value items, showing that biased JOL recollections were not due to differences in participants' motivation. Again, frequentist and Bayesian analyses provided evidence for the conclusion that incentives did not influence hindsight bias on JOLs. Together with Experiment 1, these results demonstrate that hindsight bias on JOLs is a robust phenomenon that is resistant against the use of warnings and incentives as debiasing methods.

In sum, all experiments reported in Manuscript III (Zimdahl & Undorf, 2022b) support automatic rather than controlled processes as explanation for hindsight bias on JOLs. This bias was not reduced when people were aware of it or extrinsically motivated to avoid it, indicating that hindsight bias on JOLs does not arise because people are unwilling to correctly recollect their JOLs but because they are unable to do so.

4 Discussion

> "Metacognition is simultaneously a topic of interest in its own right and a bridge between areas, e.g., between decision making and memory, between learning and motivation, and between learning and cognitive development."
>
> Thomas O. Nelson and Louis Narens (1994)

In my thesis, I aimed at further understanding what people know about what they know. Over the years, a number of studies showed that we have the astonishing ability to think about our own thinking and—more precisely—judge our own memory. This metamemory is especially relevant in the process of learning, where accurate judgments about our own memory enable us to efficiently allocate resources towards an optimal study success. It is therefore crucial to understand how accurate our metamemory is and what factors contribute to the (in-)accuracy of our own memory assessments.

As outlined earlier, people do not have direct access to their memory, but infer their metacognitive judgments from cues such as semantic association (e.g., Mueller et al., 2013; Undorf & Erdfelder, 2015), concreteness (e.g., Begg et al., 1989; Witherby & Tauber, 2017), font size (e.g., Rhodes & Castel, 2008; Undorf & Zimdahl, 2019), or volume (e.g., Frank & Kuhlmann, 2017; Rhodes & Castel, 2009). While some of these cues are valid in predicting actual memory performance (semantic association and concreteness), other cues are invalid in doing so (font size and volume), leading to metacognitive illusions (for reviews, see Undorf et al., 2022; Yang et al., 2021).

The current thesis expands this list of cues that people use to infer their metacognitive judgments by showing that visual coherence and outcome knowledge also influence JOLs. In Manuscript I (Zimdahl & Undorf, 2022a), we demonstrate that participants consistently give higher JOLs to coherent pictures than to incoherent pictures. Moreover, recognition memory is also higher for coherent than for incoherent pictures, showing that visual coherence is a valid cue in predicting actual memory performance. In contrast, outcome knowledge systematically biases the recollection of prior made JOLs towards the outcome of the memory test (Manuscript II, Zimdahl & Undorf, 2021). Interestingly enough, however, this bias leads to recollected JOLs that are more accurate than original JOLs.

Apart from the cues that people use to infer their metacognitive judgments, this thesis also sheds light on the theoretical processes underlying JOLs. Supporting the dual-basis view of metacognition (Kelley & Jacoby, 1996; Koriat, 1997), we show that both fluency and beliefs underly the visual coherence effect on JOLs (Manuscript I, Zimdahl & Undorf, 2022a). Above and beyond, Manuscript III (Zimdahl & Undorf, 2022b) demonstrates that the biased recollection of JOLs is more likely due to automatic rather than to controlled processes, which is in line with research on hindsight bias with judgments about the external world (see, e.g., Pohl & Erdfelder, 2022; Pohl & Hell, 1996).

Taken together, the results of the three manuscripts described here expand our knowledge about the basis of people's metacognitive judgments by uncovering two new cues that people use to infer their JOLs and investigating the underlying processes. The findings have both theoretical and practical implications by (1) supporting the dual-basis view of metacognition, (2) revealing that subjective measures can be used to assess fluency contributions to metamemory, (3) demonstrating significant similarities between metacognitive judgments and judgments about the external world, and (4) highlighting important aspects of metacognitive judgments in the process of learning.

4.1 Implications

The dual-basis view of metacognition (Kelley & Jacoby, 1996; Koriat, 1997) assumes that metacognitive judgments rely on experience-based, non-analytic and theory-based, analytic processes. Over the years, several studies have provided evidence for this view by showing that cues affect JOLs through fluency, beliefs, or both (see, e.g., Undorf, 2020; Yang et al., 2021). In contrast, other researchers have made the claim that JOLs rely exclusively on beliefs (e.g., Mueller & Dunlosky, 2017; Mueller et al., 2016; Mueller et al., 2013). Supporting this view, it could be shown that some cues, such as font size, impact JOLs mainly through theory-based, analytic processes (e.g., Mueller et al., 2014; Undorf & Zimdahl, 2019).

The current thesis, however, underlines the role of fluency for metacognitive judgments such as JOLs. The finding of a visual coherence effect on JOLs in people who are unaware of differences in the material shows that this effect is not solely due to people's beliefs about how visual coherence might impact their memory (Zimdahl & Undorf, 2022a). Besides, decreasing fluency through reducing the figure-ground contrast also decreased the visual coherence effect on JOLs. Most importantly, assessing fluency through a subjective measure demonstrated that the visual coherence effect on JOLs is partly mediated by experience-based, non-analytic processes. These results provide evidence for the contribution of fluency to cue effects on JOLs. Together with the finding that the visual coherence effect in JOLs was more pronounced in people who were aware of differences in the material, this research supports the dual-basis view of metacognition (Kelley & Jacoby, 1996; Koriat, 1997).

Manuscript I (Zimdahl & Undorf, 2022a) also stresses the importance of assessing fluency with the appropriate measures. In the recent past, there has been evolving criticism on the use of self-paced study time as "(at least in some situations) a potentially insensitive or invalid measure of processing fluency" (Yang et al., 2021, p. 326). The rationale behind this is that self-paced study time is not a process-pure measure of fluency as it also includes people's motivation or perceived importance of the material (see also Undorf & Ackerman, 2017). It has been argued that self-paced study time should be replaced by other fluency measures, such as identification time in a continuous identification task (e.g., Sanborn et al., 2004; Yang et al., 2018). The current findings, however, show that both measures can have difficulties in assessing quick and intuitive feelings of fluency. In contrast, a subjective measure of fluency (see Graf et al., 2018; Reber et al., 2004) is able to detect differences in the experience of mastering the material which mediate a part of the visual coherence effect on JOLs.

It is important to state that we do not argue to abandon self-paced study time or identification time as measures of fluency. Rather, we suggest that subjective and objective measures of fluency should be used dependent on the experimental conditions in which the contributions of fluency are assessed (cf. Winkielman et al., 2003). With regard to the visual coherence effect on JOLs (Zimdahl & Undorf, 2022a), objective measures turned out to be difficult to assess quick and intuitive feelings of fluency because they were accompanied by much longer presentation times of the pictures. Instead, a subjective measure was capable of showing that the visual coherence effect on JOLs was partly mediated by fluency. We therefore recommend to use subjective measures when assessing quick and intuitive feelings of fluency. In other contexts, it might be more appropriate to use objective measures of fluency (e.g., stimulus-size illusion, Undorf et al., 2017).

Manuscript II (Zimdahl & Undorf, 2021) was the first study to reveal that JOLs are affected by hindsight bias, as are many judgments about the external world (for overviews, see Pohl & Erdfelder, 2022; Roese & Vohs, 2012). In the past, hindsight bias has been demonstrated for plausibility ratings for answers to almanac questions (Campbell & Tesser, 1983; Fischhoff, 1977), likelihoods of possible event outcomes (Fischhoff & Beyth, 1975), and quantitative estimates (Erdfelder & Buchner, 1998; Hell et al., 1988). These judgments refer to the external world and thereby differ from JOLs, which refer to internal, cognitive processes. In studies investigating hindsight bias on judgments about the external world, the correct outcome (the answer to the almanac question, the event outcome, etc.) is identical for all participants and is set before participants make their judgments. Conversely, the correct outcome in studies investigating hindsight bias on JOLs (i.e., whether or not a study item is remembered) results from a learning process that unfolds during the study and is unknown until each individual participant is tested. However, the finding of hindsight bias on JOLs indicates that the cognitive processes underlying metacognitive judgments and judgments about the external world are similar and deserve mutual attention from researchers of metamemory and Judgment and Decision Making (for notable examples, see Bröder & Undorf, 2019; Undorf & Bröder, 2020; Undorf et al., 2018).

This point is stressed by the results of MANUSCRIPT III (Zimdahl & Undorf, 2022b), showing that attempts to reduce hindsight bias on JOLs proved to be as difficult as to reduce hindsight bias on judgments about the external world. While especially warnings were effective in reducing or even eliminating biased JOLs (e.g., Carpenter & Olson, 2012; Koriat & Bjork, 2006; Rhodes & Castel, 2008), they were of no avail in reducing hindsight bias on JOLs. Together with the finding that monetary incentives did not affect hindsight bias on JOLs either, these results provide convergent evidence that hindsight biases on metacognitive judgments and on judgments about the external world share significant similarities in how they arise and how they can be explained.

The finding of hindsight bias on JOLs (Zimdahl & Undorf, 2021) has implications for the process of learning. At this point, however, it is not yet clear whether biased recollection of JOLs is beneficial or detrimental for learning. On the one hand, hindsight bias on JOLs has the potential to foster learning when people use biased JOL recollections that are closer to the actual outcome as a basis for new JOLs in a subsequent study-test cycle *on the same material*. In this case, people would update their knowledge state about the memorability of the items by integrating the outcome of the memory test into the recollections of their JOLs. This resembles research on the memory-for-past-test heuristic (e.g., Ariel & Dunlosky, 2011; Finn & Metcalfe, 2007) showing that people base their JOLs in a second study-test cycle on the performance in the preceding memory test. With this, people would replace their faulty original judgments with better predictions which can be a useful strategy for learning. However, as previous research on the underconfidence-with-practice effect (e.g., Koriat et al., 2002; Serra & Dunlosky, 2005) demonstrated, this could mean that an increase in resolution of people's JOLs is accompanied by a decrease in calibration, leading to an underestimation of actual memory performance with JOLs.

On the other hand, however, the current research suggests that people do not realize that their recollected JOLs are more accurate than their original JOLs, even when they are fully aware of hindsight bias and are instructed to actively avoid it (Zimdahl & Undorf, 2022b). This suggests that by believing that they were already more accurate in their memory predictions than they actually had been, they prevent themselves from learning why their original judgments were faulty. In this case, it could be that they do not improve in their metacognitive accuracy in a second study-test cycle *on different material*. Improved metacognitive accuracy would therefore be limited to a specific set of items to which people have outcome knowledge but would not generalize to other situations in which different items are tested (cf. Koriat & Bjork, 2006). However, more research is needed to examine the implications of hindsight bias on JOLs for learning.

4.2 Future Directions

The research presented here show that visual coherence and outcome knowledge impact JOLs and investigated the theoretical processes underlying these influences. In line with dual-process theories, we could demonstrate that visual coherence affects JOLs through fluency and beliefs, whereas outcome knowledge affects the recollection of JOLs mainly through automatic rather than controlled processes. However, there are open questions that should be addressed by future research.

First, the results of Manuscript I (Zimdahl & Undorf, 2022a) hinted at some of the methodological challenges when investigating the contributions of experience-based, non-analytic and theory-based, analytic processes to cue effects on JOLs (cf. Yang et al., 2021). While those concerning fluency have already been mentioned in the previous section (e.g., assessing quick and intuitive feelings with subjective and objective measures), those concerning beliefs have to be considered as well. The evidence that beliefs contribute to the visual coherence effect on JOLs was based on the finding that this effect was more pronounced in people who were aware of differences in visual coherence than in people who were unaware of any differences in the material. Whether people were aware or unaware of differences in visual coherence was assessed via a mid-experimental questionnaire, in which participants stated whether they had seen any form of gestalt during the study phase or not. Furthermore, they were asked on how many pictures they had perceived a gestalt and to describe those perceptions as precisely as possible. Such questionnaires are an established method to assess people's beliefs about how different cues might affect JOLs (e.g., Mueller et al., 2014; Undorf & Zimdahl, 2019).

However, previous studies could demonstrate that holding a metacognitive belief does not mean that this belief is automatically applied to metacognitive judgments but often times has to be activated to affect JOLs (e.g., Kornell & Hausman, 2017; Kornell et al., 2011; Tauber et al., 2019). Thus, we found a more pronounced visual coherence effect on JOLs in people who could *potentially* base their metacognitive judgments on a specific belief about how visual coherence might affect JOLs (Zimdahl & Undorf, 2022a). Future studies could pursue this line of research by using different methods to assess metacognitive beliefs such as varying the point in time of the questionnaire (Undorf & Zimdahl, 2019), or using pre-study JOLs (Mueller et al., 2014; Witherby & Tauber, 2017), *global differentiated predictions* (GPREDs, Frank & Kuhlmann, 2017; Hertzog et al., 2009), or the learner-observer-task (Undorf & Erdfelder, 2011; Undorf et al., 2017).

Second, Manuscripts II and III (Zimdahl & Undorf, 2021, 2022b) provide evidence for hindsight bias on JOLs that is due to automatic rather than controlled processes. However, more research is needed to understand which specific processes contribute most to hindsight bias on JOLs. As outlined in Manuscript III (Zimdahl & Undorf, 2022b), hindsight bias on JOLs is compatible with theories assuming an *immediate assimilation* of original JOLs and outcome knowledge (Fischhoff, 1975, 1977), the *trace-strength hypothesis* proposing that the outcome knowledge impairs the retrieval of the original JOLs (Hell et al., 1988), and the *anchoring-and-adjustment approach* suggesting that the outcome knowledge is used as an anchor to reconstruct the original JOLs (Tversky & Kahneman, 1974). According to Erdfelder and Buchner (1998), these different explanatory approaches address different stages of the process that participants move through when being asked to remember their original judgments (see also Erdfelder et al., 2007). While the immediate assimilation and the trace-strength hypothesis refer to a recollection stage, the anchoring and adjustment takes place at a reconstruction stage after recollection failed. Future research could examine the influence of recollection and reconstruction processes to hindsight bias by using multinomial processing tree models (Batchelder & Riefer, 1999; Erdfelder & Buchner, 1998).

Third, this thesis calls for an integrative model about the contributions of dual processes to metacognitive judgments such as JOLs. Since the dual-basis view of metacognition (Kelley & Jacoby, 1996; Koriat, 1997) was introduced, 25 years of research have accumulated a bulk of evidence that JOLs rely on experience-based, non-analytic and theory-based, analytic processes. Especially over the last decade, the contributions of fluency and beliefs have been studied for a variety of metacognitive phenomena, leading to an enhanced insight into the basis of JOLs (for overviews, see Undorf, 2020; Undorf et al., 2022; Yang et al., 2021). Yet, little is known about *when* fluency, beliefs, or a combination of both contribute to JOLs and *why* each of those factors underly some metacognitive phenomena but not others. Consequently, the next step would be to integrate those findings into a model that can make predictions about the relative contributions of both kinds of processes to cue effects on JOLs.

4.3 Conclusion

In my thesis, I uncovered two new cues that people use to infer their metacognitive judgments and investigated the underlying theoretical processes of the effects that these two cues have on metamemory. In line with the dual-basis view of metacognition, the current research shows that visual coherence affects JOLs through fluency and beliefs. In contrast, outcome knowledge produces hindsight bias on JOLs through automatic rather than controlled processes. These results provide important insights into people's metamemory and lead to a better understanding of what we know about what we know.

5 Bibliography

Ackerman, R., Bernstein, D. M., & Kumar, R. (2020). Metacognitive hindsight bias. *Memory & Cognition*, *48*(5), 731–744. https://doi.org/10.3758/s13421-020-01012-w

Arbuckle, T. Y., & Cuddy, L. L. (1969). Discrimination of item strength at time of presentation. *Journal of Experimental Psychology*, *81*(1), 126–131. https://doi.org/10.1037/h0027455

Ariel, R., & Dunlosky, J. (2011). The sensitivity of judgment-of-learning resolution to past test performance, new learning, and forgetting. *Memory & Cognition*, *39*(1), 171–184. https://doi.org/10.3758/s13421-010-0002-y

Asendorpf, J. B., Conner, M., de Fruyt, F., de Houwer, J., Denissen, J. J. A., Fiedler, K., Fiedler, S., Funder, D. C., Kliegl, R., Nosek, B. A., Perugini, M., Roberts, B. W., Schmitt, M., van Aken, M. A. G., Weber, H., & Wicherts, J. M. (2013). Recommendations for increasing replicability in psychology. *European Journal of Personality*, *27*(2), 108–119. https://doi.org/10.1002/per.1919

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In R. S. Wyer JR. & T. K. Srull (Eds.), *Handbook of social cognition: Basic processes; Applications* (pp. 1–40). Lawrence Erlbaum Associates, Inc.

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review*, *6*(1), 57–86. https://doi.org/10.3758/BF03210812.

Begg, I., Duft, S., Lalonde, P., Melnick, R., & Sanvito, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory and Language*, *28*(5), 610–632. https://doi.org/10.1016/0749-596X(89)90016-8

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: beliefs, techniques, and illusions. *Annual Review of Psychology*, *64*, 417–444. https://doi.org/10.1146/annurev-psych-113011-143823

Bolte, A., & Goschke, T. (2008). Intuition in the context of object perception: intuitive gestalt judgments rest on the unconscious activation of semantic representations. *Cognition*, *108*(3), 608–616. https://doi.org/10.1016/j.cognition.2008.05.001

Bröder, A., & Undorf, M. (2019). Metamemory viewed through the judgment lens. *Acta Psychologica*, *197*, 153–165. https://doi.org/10.1016/j.actpsy.2019.04.011

Buchner, A., Erdfelder, E., Steffens, M. C., & Martensen, H. (1997). The nature of memory processes underlying recognition judgments in the process dissociation procedure. *Memory & Cognition*, *25*(4), 508–517. https://doi.org/10.3758/bf03201126

Buchner, A., Erdfelder, E., & Vaterrodt-Plünnecke, B. (1995). Toward unbiased measurement of conscious and unconscious memory processes within the process dissociation framework. *Journal of Experimental Psychology: General*, *124*(2), 137–160. https://doi.org/10.1037/0096-3445.124.2.137

Campbell, J. D., & Tesser, A. (1983). Motivational interpretations of hindsight bias: An individual difference analysis. *Journal of Personality*, *51*, 605–620. https://doi.org/10.1111/j.1467-6494.1983.tb00868.x

Carpenter, S. K., & Olson, K. M. (2012). Are pictures good for learning new vocabulary in a foreign language? Only if you think they are not. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(1), 92–101. https://doi.org/10.1037/a0024828

Castel, A. D. (2008). Metacognition and learning about primacy and recency effects in free recall: The utilization of intrinsic and extrinsic cues when making judgments of learning. *Memory & Cognition*, *36*, 429–437. https://doi.org/10.3758/MC.36.2.429

Chaiken, S., & Trope, Y. (Eds.). (1999). *Dual-process theories in social psychology*. The Guilford Press.

Chang, M., & Brainerd, C. J. (2022). Association and dissociation between judgments of learning and memory: A meta-analysis of the font size effect. *Metacognition and Learning*, Advance online publication. https://doi.org/10.1007/s11409-021-09287-3

Christensen-Szalanski, J. J., & Willham, C. F. (1991). The hindsight bias: A meta-analysis. *Organizational Behavior and Human Decision Processes*, *48*, 147–168. https://doi.org/10.1016/0749-5978(91)90010-Q

Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PloS one*, *8*(3), e57410. https://doi.org/10.1371/journal.pone.0057410

Descartes, R. (1637/2006). *A discourse on the method of correctly conducting one's reason and seeking truth in the sciences*. Oxford University Press.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*(1), 5–18. https://doi.org/10.1037/0022-3514.56.1.5

Dunlosky, J., & Nelson, T. O. (1994). Does the sensitivity of judgments of learning (JOLs) to the effects of various study activities depend on when the JOLs occur? *Journal of Memory and Language*, *33*(4), 545–565. https://doi.org/10.1006/jmla.1994.1026

Dunlosky, J., & Metcalfe, J. (2009). *Metacognition*. Sage.

Dunlosky, J., & Tauber, S. U. K. (Eds.). (2016). *The Oxford handbook of metamemory*. Oxford University Press.

Erdfelder, E., Brandt, M., & Bröder, A. (2007). Recollection biases in hindsight judgments. *Social Cognition*, *25*, 114–131. https://doi.org/10.1521/soco.2007.25.1.114

Erdfelder, E., & Buchner, A. (1998). Decomposing the hindsight bias: A multinomial processing tree model for separating recollection and reconstruction in hindsight. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 387–414. https://doi.org/10.1037/0278-7393.24.2.387

Finn, B., & Metcalfe, J. (2007). The role of memory for past test in the underconfidence with practice effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(1), 238–244. https://doi.org/10.1037/0278-7393.33.1.238

Finn, B., & Metcalfe, J. (2008). Judgments of learning are influenced by memory for past test. *Journal of Memory and Language*, *58*(1), 19–34. https://doi.org/10.1016/j.jml.2007.03.006

Fischhoff, B. (1975). Hindsight is not equal to foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, *1*, 288–299. https://doi.org/10.1037/0096-1523.1.3.288

Fischhoff, B. (1977). Perceived informativeness of facts. *Journal of Experimental Psychology: Human Perception and Performance*, *3*, 349–358. https://doi.org/10.1037/0096-1523.3.2.349

Fischhoff, B., & Beyth, R. (1975). I knew it would happen: Remembered probabilities of once-future things. *Organizational Behavior and Human Performance*, *13*, 1–16. https://doi.org/10.1016/0030-5073(75)90002-1

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. *Advances in Experimental Social Psychology*, *23*, 1–74. https://doi.org/10.1016/S0065-2601(08)60317-2

Flavell, J. H. (1971). First discussant's comments: What is memory development the development of? *Human Development*, *14*, 272–278. https://doi.org/10.1159/000271221

Frank, D. J., & Kuhlmann, B. G. (2017). More than just beliefs: Experience and beliefs jointly contribute to volume effects on metacognitive judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(5), 680–693. https://doi.org/10.1037/xlm0000332

Gawronski, B., & Creighton, L. A. (2013). Dual process theories. In D. E. Carlston (Ed.), *The Oxford handbook of social cognition* (pp. 282–312). Oxford University Press. https://doi.org/10.1037/t03782-000.

Goodman, L. A., & Kruskal, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association*, *49*(268), 732. https://doi.org/10.2307/2281536

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

Guilbault, R. L., Bryant, F. B., Brockway, J. H., & Posavac, E. J. (2004). A meta-analysis of research on hindsight bias. *Basic and Applied Social Psychology*, *26*, 103–117. https://doi.org/10.1207/s15324834basp2602{\&}3{\textunderscore}1

Halamish, V. (2018). Can very small font size enhance memory? *Memory & Cognition*, *46*(6), 979–993. https://doi.org/10.3758/s13421-018-0816-6

Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology*, *56*(4), 208–216. https://doi.org/10.1037/h0022263

Hart, J. T. (1967). Second-try recall, recognition, and the memory-monitoring process. *Journal of Educational Psychology*, *58*, 193–197. https://doi.org/10.1037/h0024908

Hell, W., Gigerenzer, G., Gauggel, S., Mall, M., & Müller, M. (1988). Hindsight bias: An interaction of automatic and motivational factors? *Memory & Cognition*, *16*, 533–538. https://doi.org/10.3758/BF03197054

Hertzog, C., Dunlosky, J., Robinson, A. E., & Kidder, D. P. (2003). Encoding fluency is a cue used for judgments about learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 22–34. https://doi.org/10.1037/0278-7393.29.1.22

Hertzog, C., Price, J., Burpee, A., Frentzel, W. J., Feldstein, S., & Dunlosky, J. (2009). Why do people show minimal knowledge updating with task experience: Inferential deficit or experimental artifact? *Quarterly Journal of Experimental Psychology*, *62*(1), 155–173. https://doi.org/10.1080/17470210701855520

Hines, J. C., Touron, D. R., & Hertzog, C. (2009). Metacognitive influences on study time allocation in an associative recognition task: An analysis of adult age differences. *Psychology and Aging*, *24*(2), 462–475. https://doi.org/10.1037/a0014417

Hoffrage, U., Hertwig, R., & Gigerenzer, G. (2000). Hindsight bias: A by-product of knowledge updating? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 566–581. https://doi.org/10.1037/0278-7393.26.3.566

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*(5), 513–541. https://doi.org/10.1016/0749-596X(91)90025-F

Jang, Y., & Nelson, T. O. (2005). How many dimensions underlie judgments of learning and recall? Evidence from state-trace methodology. *Journal of Experimental Psychology: General*, *134*(3), 308–326. https://doi.org/10.1037/0096-3445.134.3.308

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, *58*, 697–720. https://doi.org/10.1037/0003-066X.58.9.697

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

Kelley, C. M., & Jacoby, L. L. (1996). Adult egocentrism: Subjective experience versus analytic bases for judgment. *Journal of Memory and Language*, *35*(2), 157–175. https://doi.org/10.1006/jmla.1996.0009

Kenny, D. A., Korchmaros, J. D., & Bolger, N. (2003). Lower level mediation in multilevel models. *Psychological Methods*, *8*(2), 115–128. https://doi.org/10.1037/1082-989X.8.2.115

Keren, G., & Schul, Y. (2009). Two is not always better than one: A critical evaluation of two-system theories. *Perspectives on Psychological Science*, *4*(6), 533–550. https://doi.org/10.1111/j.1745-6924.2009.01164.x

King, J. F., Zechmeister, E. B., & Shaughnessy, J. J. (1980). Judgments of knowing: The influence of retrieval practice. *The American Journal of Psychology*, *93*(2), 329–343. https://doi.org/10.2307/1422236

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*(4), 349–370. https://doi.org/10.1037/0096-3445.126.4.349

Koriat, A. (2007). Metacognition and consciousness. In P. D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *The Cambridge handbook of consciousness* (pp. 289–325). Cambridge University Press. https://doi.org/10.1017/CBO9780511816789.012

Koriat, A., & Bjork, R. A. (2006). Mending metacognitive illusions: A comparison of mnemonic-based and theory-based procedures. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(5), 1133–1145. https://doi.org/10.1037/0278-7393.32.5.1133

Koriat, A., & Levy-Sadot, R. (1999). Processes underlying metacognitive judgments: Information-based and experience-based monitoring of one's own knowledge. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 483–502). The Guilford Press.

Koriat, A., & Ma'ayan, H. (2005). The effects of encoding fluency and retrieval fluency on judgments of learning. *Journal of Memory and Language*, *52*, 478–492. https://doi.org/10.1016/j.jml.2005.01.001

Koriat, A., Ma'ayan, H., & Nussinson, R. (2006). The intricate relationships between monitoring and control in metacognition: Lessons for the cause-and-effect relation between subjective experience and behavior. *Journal of Experimental Psychology: General*, *135*(1), 36–69. https://doi.org/10.1037/0096-3445.135.1.36

Koriat, A., Sheffer, L., & Ma'ayan, H. (2002). Comparing objective and subjective learning curves: Judgments of learning exhibit increased underconfidence with practice. *Journal of Experimental Psychology: General*, *131*(2), 147–162. https://doi.org/10.1037/0096-3445.131.2.147

Kornell, N., & Hausman, H. (2017). Performance bias: Why judgments of learning are not affected by learning. *Memory & Cognition*, *45*(8), 1270–1280. https://doi.org/10.3758/s13421-017-0740-1

Kornell, N., Rhodes, M. G., Castel, A. D., & Tauber, S. K. (2011). The ease-of-processing heuristic and the stability bias: Dissociating memory, memory beliefs, and memory judgments. *Psychological Science*, *22*(6), 787–794. https://doi.org/10.1177/0956797611407929

Kruglanski, A. W., Pierro, A., Mannetti, L., Erb, H.-.-P., & Young Chun, W. (2007). On the parameters of human judgment. *Advances in Experimental Social Psychology*, *39*, 255–303. https://doi.org/10.1016/S0065-2601(06)39005-3

Krull, J. L., & MacKinnon, D. P. (2001). Multilevel modeling of individual and group level mediated effects. *Multivariate Behavioral Research*, *36*(2), 249–277. https://doi.org/10.1207/S15327906MBR3602{\textunderscore}06

Kuhlmann, B. G., & Undorf, M. (2018). Is all metamemory monitoring spared from aging? A dual-process examination. *Psychology and Aging*, *33*(8), 1152–1167. https://doi.org/10.1037/pag0000318

Legge, G. E., & Bigelow, C. A. (2011). Does print size matter for reading? A review of findings from vision science and typography. *Journal of Vision*, *11*(5). https://doi.org/10.1167/11.5.8

Leonesio, R. J., & Nelson, T. O. (1990). Do different metamemory judgments tap the same underlying aspects of memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(3), 464–467. https://doi.org/10.1037//0278-7393.16.3.464.

Mandler, G. (2008). Familiarity breeds attempts: A critical review of dual-process theories of recognition. *Perspectives on Psychological Science*, *3*(5), 390–399. https://doi.org/10.1111/j.1745-6924.2008.00087.x

Mazzoni, G., & Nelson, T. O. (1995). Judgments of learning are affected by the kind of encoding in ways that cannot be attributed to the level of recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(5), 1263–1274. https://doi.org/10.1037/0278-7393.21.5.1263

Metcalfe, J., & Finn, B. (2008). Familiarity and retrieval processes in delayed judgments of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(5), 1084–1097. https://doi.org/10.1037/a0012580

Mueller, M. L., & Dunlosky, J. (2017). How beliefs can impact judgments of learning: Evaluating analytic processing theory with beliefs about fluency. *Journal of Memory and Language*, *93*, 245–258. https://doi.org/10.1016/j.jml.2016.10.008

Mueller, M. L., Dunlosky, J., & Tauber, S. K. (2016). The effect of identical word pairs on people's metamemory judgments: What are the contributions of processing fluency and beliefs about memory? *Quarterly Journal of Experimental Psychology*, *69*(4), 781–799. https://doi.org/10.1080/17470218.2015.1058404

Mueller, M. L., Dunlosky, J., Tauber, S. K., & Rhodes, M. G. (2014). The font-size effect on judgments of learning: Does it exemplify fluency effects or reflect people's beliefs about memory? *Journal of Memory and Language*, *70*, 1–12. https://doi.org/10.1016/j.jml.2013.09.007

Mueller, M. L., Tauber, S. K., & Dunlosky, J. (2013). Contributions of beliefs and processing fluency to the effect of relatedness on judgments of learning. *Psychonomic Bulletin & Review*, *20*, 378–384. https://doi.org/10.3758/s13423-012-0343-6

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, *95*(1), 109–133. https://doi.org/10.1037/0033-2909.95.1.109

Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The 'delayed-JOL effect.' *Psychological Science*, *2*, 267–270. https://doi.org/10.1111/j.1467-9280.1991.tb00147.x

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (pp. 125–173). Academic Press. https://doi.org/10.1016/S0079-7421(08)60053-5

Nelson, T. O., & Narens, L. (1994). Why investigate metacognition? In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing*. The MIT Press. https://doi.org/10.7551/mitpress/4561.003.0003

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

Pohl, R. F. (1992). Der Rückschau-Fehler: Systematische Verfälschung der Erinnerung bei Experten und Novizen. *Kognitionswissenschaft*, *3*, 38–44.

Pohl, R. F., & Erdfelder, E. (2022). Hindsight bias. In R. F. Pohl (Ed.), *Cognitive illusions: Intriguing phenomena in thinking, judgment and memory (Third Edition)* (pp. 436–454). Routledge. https://doi.org/10.4324/9781003154730-31

Pohl, R. F., & Hell, W. (1996). No reduction in hindsight bias after complete information and repeated testing. *Organizational Behavior and Human Decision Processes*, *67*, 49–58. https://doi.org/10.1006/obhd.1996.0064

Reber, R., Wurtz, P., & Zimmermann, T. D. (2004). Exploring “fringe” consciousness: The subjective experience of perceptual fluency and its objective bases. *Consciousness and Cognition*, *13*(1), 47–60. https://doi.org/10.1016/S1053-8100(03)00049-7

Rhodes, M. G. (2016). Judgments of learning: Methods, data, and theory. In J. Dunlosky & S. U. K. Tauber (Eds.), *The Oxford handbook of metamemory* (pp. 65–80). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199336746.013.4

Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology: General*, *137*(4), 615–625. https://doi.org/10.1037/a0013684

Rhodes, M. G., & Castel, A. D. (2009). Metacognitive illusions for auditory information: Effects on monitoring and control. *Psychonomic Bulletin & Review*, *16*(3), 550–554. https://doi.org/10.3758/PBR.16.3.550

Rhodes, M. G., & Tauber, S. K. (2011). The influence of delaying judgments of learning on metacognitive accuracy: a meta-analytic review. *Psychological Bulletin*, *137*(1), 131–148. https://doi.org/10.1037/a0021705

Robinson, D. N. (1989). *Aristotle's psychology*. Columbia University Press.

Roese, N. J., & Vohs, K. D. (2012). Hindsight bias. *Perspectives on Psychological Science*, *7*(5), 411–426. https://doi.org/10.1177/1745691612454303

Sanborn, A. N., Malmberg, K. J., & Shiffrin, R. M. (2004). High-level effects of masking on perceptual identification. *Vision Research*, *44*(12), 1427–1436. https://doi.org/10.1016/j.visres.2004.01.004

Serra, M. J., & Ariel, R. (2014). People use the memory for past-test heuristic as an explicit cue for judgments of learning. *Memory & Cognition*, *42*, 1260–1272. https://doi.org/10.3758/s13421-014-0431-0

Serra, M. J., & Dunlosky, J. (2005). Does retrieval fluency contribute to the underconfidence-with-practice effect? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(6), 1258–1266. https://doi.org/10.1037/0278-7393.31.6.1258

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, *117*(1), 34–50. https://doi.org/10.1037/0096-3445.117.1.34

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, *23*(5), 645–665. https://doi.org/10.1017/S0140525X00003435

Susser, J. A., & Mulligan, N. W. (2015). The effect of motoric fluency on metamemory. *Psychonomic Bulletin & Review*, *22*(4), 1014–1019. https://doi.org/10.3758/s13423-014-0768-1

Susser, J. A., Panitz, J., Buchin, Z., & Mulligan, N. W. (2017). The motoric fluency effect on metamemory. *Journal of Memory and Language*, *95*, 116–123. https://doi.org/10.1016/j.jml.2017.03.002

Tauber, S. K., Witherby, A. E., & Dunlosky, J. (2019). Beliefs about memory decline in aging do not impact judgments of learning (JOLs): A challenge for belief-based explanations of JOLs. *Memory & Cognition*, *47*(6), 1102–1119. https://doi.org/10.3758/s13421-019-00919-3

Topolinski, S., & Strack, F. (2009). The architecture of intuition: Fluency and affect determine intuitive judgments of semantic and visual coherence and judgments of grammaticality in artificial grammar learning. *Journal of Experimental Psychology: General*, *138*(1), 39–63. https://doi.org/10.1037/a0014678

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

Undorf, M. (2020). Fluency illusions in metamemory. In A. M. Cleary & B. L. Schwartz (Eds.), *Memory quirks* (pp. 150–174). Routledge. https://doi.org/10.4324/9780429264498-12

Undorf, M., & Ackerman, R. (2017). The puzzle of study time allocation for the most challenging items. *Psychonomic Bulletin & Review*, *24*(6), 2003–2011. https://doi.org/10.3758/s13423-017-1261-4

Undorf, M., Böhm, S., & Cüpper, L. (2016). Do judgments of learning predict automatic influences of memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(6), 882–896. https://doi.org/10.1037/xlm0000207

Undorf, M., & Bröder, A. (2020). Cue integration in metamemory judgements is strategic. *Quarterly Journal of Experimental Psychology*, *73*(4), 629–642. https://doi.org/10.1177/1747021819882308

Undorf, M., & Erdfelder, E. (2011). Judgments of learning reflect encoding fluency: conclusive evidence for the ease-of-processing hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(5), 1264–1269. https://doi.org/10.1037/a0023719

Undorf, M., & Erdfelder, E. (2015). The relatedness effect on judgments of learning: A closer look at the contribution of processing fluency. *Memory & Cognition*, *43*(4), 647–658. https://doi.org/10.3758/s13421-014-0479-x

Undorf, M., Navarro-Báez, S., & Zimdahl, M. F. (2022). Metacognitive illusions. In R. F. Pohl (Ed.), *Cognitive illusions: Intriguing phenomena in thinking, judgment and memory (Third Edition)* (pp. 307–323). Routledge. https://doi.org/10.4324/9781003154730-22

Undorf, M., Söllner, A., & Bröder, A. (2018). Simultaneous utilization of multiple cues in judgments of learning. *Memory & Cognition*, *46*, 507–519. https://doi.org/10.3758/s13421-017-0780-6

Undorf, M., & Zimdahl, M. F. (2019). Metamemory and memory for a wide range of font sizes: What is the contribution of perceptual fluency? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(1), 97–109. https://doi.org/10.1037/xlm0000571

Undorf, M., Zimdahl, M. F., & Bernstein, D. M. (2017). Perceptual fluency contributes to effects of stimulus size on judgments of learning. *Journal of Memory and Language*, *92*, 293–304. https://doi.org/10.1016/j.jml.2016.07.003

Winkielman, P., Schwarz, N., Fazendeiro, T. A., & Reber, R. (2003). The hedonic marking of processing fluency: Implications for evaluative judgment. In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 189–217). Lawrence Erlbaum Associates Publishers.

Wippich, W., Mecklenbräuker, S., & Krisch, S. (1994). Priming effects and intuitive judgments. *Swiss Journal of Psychology*, *53*(2), 63–77.

Witherby, A. E., & Tauber, S. K. (2017). The concreteness effect on judgments of learning: Evaluating the contributions of fluency and beliefs. *Memory & Cognition*, *45*(4), 639–650. https://doi.org/10.3758/s13421-016-0681-0

Witherby, A. E., & Tauber, S. K. (2018). Monitoring of learning for emotional faces: how do fine-grained categories of emotion influence participants' judgments of learning and beliefs about memory? *Cognition & Emotion*, *32*(4), 860–866. https://doi.org/10.1080/02699931.2017.1360252

Yang, C., Huang, T. S.-T., & Shanks, D. R. (2018). Perceptual fluency affects judgments of learning: The font size effect. *Journal of Memory and Language*, *99*, 99–110. https://doi.org/10.1016/j.jml.2017.11.005

Yang, C., Yu, R., Hu, X., Luo, L., Huang, T. S.-T., & Shanks, D. R. (2021). How to assess the contributions of processing fluency and beliefs to the formation of judgments of learning: methods and pitfalls. *Metacognition and Learning*, *16*(2), 319–343. https://doi.org/10.1007/s11409-020-09254-4

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*(3), 441–517. https://doi.org/10.1006/jmla.2002.2864

Zimdahl, M. F., & Undorf, M. (2021). Hindsight bias in metamemory: Outcome knowledge influences the recollection of judgments of learning. *Memory*, *29*(5), 559–572. https://doi.org/10.1080/09658211.2021.1919144

Zimdahl, M. F., & Undorf, M. (2022a). *Visual coherence impacts judgments of learning: Evidence for the dual-basis view of metacognition* [Manuscript submitted for publication]. Department of Psychology, University of Mannheim.

Zimdahl, M. F., & Undorf, M. (2022b). *Warnings and incentives do not reduce hindsight bias on judgments of learning* [Manuscript submitted for publication]. Department of Psychology, University of Mannheim.

Zimmerman, C. A., & Kelley, C. M. (2010). "I'll remember this!": Effects of emotionality on memory predictions versus memory performance. *Journal of Memory and Language*, *62*, 240–253. https://doi.org/10.1016/j.jml.2009.11.004

A Acknowledgments

"Which is more important," asked Big Panda,
"the journey or the destination?"

"The company." said Tiny Dragon.

from *Big Panda & Tiny Dragon*
by James Norbury

As a close friend once said: a PhD can be a lonesome journey, even more so during a global pandemic. This is the reason why I am forever grateful for the many tiny dragons that accompanied me along my way. And although I hope that each and every one of them know how much they mean to me, I would like to take the opportunity to thank them all for being the greatest company one could ever wish for.

Thank you ...

... Monika, for offering me the chance to make this journey. Our paths crossed a long time ago and I am more than happy that you introduced me to the wonderful world of metacognition. Your passion about science, your striving for new ideas, and your determination inspired me to start this PhD. Thank you for always taking your time, believing in me, and most importantly: being honest with me at every step of the way.

... Edgar, for supporting me and having an open door to your office at all times. I remember that it was your lecture on my first day as a student that showed me that this path was the right way to go. Thank you for making my workplace feel like more than a place to work: various evenings in the Dölle-Kino, summer gatherings in your garden, and Christmas events with the legendary Psycho choir made this time so unique.

... Bea, not only for providing me with highly appreciated feedback throughout my dissertation, but also for your amazing support outside the workplace. Thank you for your encouragement in the scientific world as well as when you attended my concerts.

... Rüdiger, for our conversations about life, music, and hindsight bias. This dissertation would not have been possible without your contributions to science, and I feel honored to have pursued this trail for a while to build it a bit further.

... to my colleagues Lili, Letti, Sofia, Philipp, David, Ruben, and Pedro who have become so much more than colleagues. Your presence has made this journey a little less lonesome and I am incredibly grateful for many shared memories with you.

... to my invaluable friends Elli, Sophie, Feli, Patty, Kathi, and Marisa. You entered my life long before this journey started, and I am extremely blessed to know that you will be there long after it ends. Your friendship, your warmth, and your love helped me through times that were rough and made happy moments even brighter.

... Franzi, for various coffee breaks, after work beers, and virtual sport sessions. Thank you for reminding me of my life beyond the PhD, enjoying it with me, and motivating me to pursue my goals inside and outside academia.

... Pia, for your moral support, your loyalty, and your appreciative way with which you surround the people who are important to you. I feel lucky that our friendship defies the distance and that you always make sure that the ones around you are fine.

... Tabea, for sharing my passions and so many phone calls during the past years. Although we do not see each other often, I am fortunate that you are a part of my life and I am a part of yours.

... Steffen, for always laughing at my jokes and giving me the impression that I am at least remotely funny. I love talking to you about everything and I am grateful that you are there whenever I need a friendly piece of advice.

... Martini, for listening to my stories and telling me yours. Our lives became parallel in an unexpected moment, and I am deeply moved to have a friend like you who cares and gives in a way only few other people do.

... Martin, for being you. You have been an anchor in my life since the day I attended your first friends gathering uninvited. Throughout the years, we have become best friends who share the same interests, the same humor, and the same memories. I would not be at this point in my life without you and I will never forget that you were there for me when I needed you the most.

... Lisa, for going most of the way with me. You and your family will always have a special place in my heart.

... to my grandparents, both those who are no longer with us, and those who can celebrate this moment with me. To my grandpa, who I would have loved to meet, because I know that it would have been a joy to make music with him. To my grandma, who—despite the fact that she is now longer in another place than I had the chance to spend time with her—still inspires me up to this day. And to my other grandparents, for always promoting me and putting faith in me.

... to my twin-sister Noëmi, who brings sunshine into rainy days and music into quiet places. We are sharing a life since the moment we were born, and I cannot express what it means to me that this connection lasts up to this day and far beyond. I admire you for all you do and who you are: a creative mind with brilliant ideas, a dedicated maker with admirable enthusiasm, and—luckily for me—the best musical companion I will ever have.

... to my Mom and Dad, for making me the person that I am. You raised Noëmi and me in a way that allows us to be where we want to be. Your values, your education, and your love are the reason for what we have achieved so far, and your unconditional support assures us that we can accomplish everything we strive for. For this, and many other reasons, I will love you forever from the bottom of my heart.

Love, Malte
March 2022

D Copies of Manuscripts

Visual Coherence Impacts Judgments of Learning: Evidence for the Dual-Basis View of Metacognition

Malte F. Zimdahl and Monika Undorf

Department of Psychology, School of Social Sciences, University of Mannheim

Abstract

People can discriminate between visually coherent pictures depicting fragmented objects and visually incoherent pictures depicting random visual information even when they are unable to identify what is depicted. The current study examined the effect of visual coherence on metamemory judgments (judgments of learning; JOLs) and the contributions of experience-based, non-analytic and theory-based, analytic processes to this effect. In six experiments, participants saw coherent and incoherent pictures, made a JOL for each picture, and took a recognition memory test. All six experiments revealed higher JOLs for coherent than for incoherent pictures. Recognition memory was higher for coherent than for incoherent pictures in five of six experiments. Consistent with the notion that experience-based, non-analytic processes contribute to the visual coherence effect on JOLs, reducing fluency through a low figure-ground contrast decreased the visual coherence effect (Experiment 2) and the independent fluency measures self-paced study time (Experiment 3) and identification time from a continuous identification task (Experiment 4) indicated that coherent pictures were processed more fluently than incoherent pictures. However, only an independent measure of subjective fluency mediated the visual coherence effect on JOLs (Experiment 6). Consistent with the notion that theory-based, analytic processes contribute to the visual coherence effect on JOLs, a joint analysis of all experiments revealed that this effect was more pronounced in people who could base their JOLs on explicit beliefs about visual coherence. Overall, the current study shows that visual coherence impacts JOLs through experience-based, non-analytic processes and theory-based, analytic processes, thus supporting the dual-basis view of metacognition.

Keywords: metamemory, judgments of learning, visual coherence, fluency, beliefs

---

This research was supported by the University of Mannheim's Graduate School of Economic and Social Sciences funded by the German Research Foundation (DFG) and by a grant from the DFG to Monika Undorf (UN 345/1-3).

The data and program code for all experiments are available at https://osf.io/c8xfn.

Correspondence concerning this article should be addressed to Malte F. Zimdahl, Department of Psychology, School of Social Sciences, University of Mannheim, A5, 6, 68159 Mannheim, Germany. Email: malte.zimdahl@psychologie.uni-mannheim.de

People's predictions about their future memory performance (judgments of learning; JOLs) are generally quite accurate (e.g., Koriat, 1997). This has been found in numerous studies showing reliable correlations between JOLs and actual memory performance (see, e.g., Dunlosky & Metcalfe, 2009; Koriat, 2007). However, there has also been evidence for systematic dissociations between JOLs and actual memory performance that lead to inaccurate memory predictions (see, e.g., Undorf, 2020). Overall, JOLs predict subsequent memory performance well above chance, but are not perfectly aligned with actual memory performance.

Moderate accuracy of JOLs as well as systematic dissociations between JOLs and actual memory performance support the idea that people base their memory predictions on cues of different validities (e.g., Koriat, 1997; Undorf et al., 2018). For example, because word pair relatedness influences memory performance, basing JOLs on relatedness fosters JOL accuracy (e.g., Mueller et al., 2013; Undorf & Erdfelder, 2015). Conversely, font size is a misleading cue for JOLs: While people usually assign higher JOLs to words that are written in larger compared to smaller font sizes, memory performance often does not increase with increasing font size (Rhodes & Castel, 2008; Undorf & Zimdahl, 2019; Yang et al., 2018; for a meta-analysis, see Luna et al., 2018). Over the years, a variety of cues have been identified as impacting JOLs, some of which are valid such as concreteness (e.g., Begg et al., 1989; Witherby & Tauber, 2017) or animacy (Li et al., 2016), and some of which are misleading such as volume (Frank & Kuhlmann, 2017; Rhodes & Castel, 2009) or stimulus size (Undorf et al., 2017).

The Basis of JOLs

It has been argued that cues impact JOLs through two different types of processes: experience-based, non-analytic processes, and theory-based, analytic processes (dual-basis view of metacognition; Kelley & Jacoby, 1996; Koriat, 1997, 2007). As an example for experience-based, non-analytic contributions to memory predictions, there is evidence that people use subjective experiences such as fluency—the ease of processing—to form JOLs (Besken, 2016; Undorf & Erdfelder, 2011; Undorf et al., 2017; Yang et al., 2018). Consistent with this view, it was found that people assign higher JOLs to materials that they read fast rather than slow (e.g., Ball et al., 2014; Undorf & Zander, 2017) and to items that they write with their dominant hand rather than with their non-dominant hand (Susser & Mulligan, 2015; Susser et al., 2017). Thus, processing materials fluently can result in high JOLs.

As an example for theory-based, analytic processes, it has been found that JOLs are based on peoples' explicit beliefs about how memory works (Koriat et al., 2004; Mueller et al., 2014). Support for this idea comes from studies in which JOLs that are assessed before the respective item is studied and therefore cannot be influenced by experience-based processes (pre-study JOLs) are higher for large-font words than for small-font words (e.g., Mueller et al., 2014). Also, there is evidence that JOLs are aligned with the beliefs people report in questionnaires (e.g., Mueller et al., 2016; Undorf & Zimdahl, 2019). Thus, explicit beliefs that specific materials are memorably can result in high JOLs.

There is an ongoing debate about the contributions of fluency and beliefs to cue effects on JOLs. As an example, the font-size effect on JOLs was originally thought to be driven by fluency (Rhodes & Castel, 2008). However, later research demonstrated that the font-size effect is largely due to participants' beliefs about how font size affects their memory (Mueller et al., 2014; Undorf & Zimdahl, 2019; but see Yang et al., 2018). Both fluency and beliefs have been found to contribute to a number of other cue effects on JOLs, including relatedness (Mueller et al., 2013; Undorf & Erdfelder, 2015), word frequency (Jia et al., 2016; Mendes et al., 2020, 2021), and, most relevant for present purposes, semantic coherence. Undorf and Zander (2017) found that people give higher JOLs to groups of three words that share a common associate (*coherent triads*; e.g., *silk-cream-even*, solution: *smooth*) than to groups of three words that have no common associate (*incoherent triads*; e.g., *deck-stool-pocket*). When participants could read the triads at their own pace, reading times for coherent word triads were shorter than for incoherent word triads, indicating that fluency was higher for coherent than for incoherent word triads. Also important, this semantic coherence effect on JOLs was not only evident in people who were aware of relations between triad words but also in people who were unaware of relations between triad words. This indicated that beliefs alone cannot explain the semantic coherence effect on JOLs. However, a more pronounced semantic coherence effect on JOLs in aware people suggested that theory-based, analytic processes also contribute to this effect. Thus, consistent with the dual-basis view of metacognition (Kelley & Jacoby, 1996; Koriat, 1997, 2007), both fluency and beliefs contribute to the semantic coherence effect on JOLs.

Visual Coherence

The present study aimed at testing contributions of fluency and beliefs to JOLs by investigating *visual* coherence. Visually coherent material are pictures of meaningful objects that are visually degraded to such an extent that they are hardly identifiable (Bowers et al., 1990). Nonetheless, these pictures are coherent in the sense of depicting a gestalt. Incoherent pictures, in contrast, depict random visual information that does not form a gestalt. Bowers et al. (1990) demonstrated that people can discriminate coherent from incoherent pictures even when they were unable to identify what is depicted (see also Bolte & Goschke, 2008). Moreover, Topolinski and Strack (2009) showed that increasing perceptual fluency through repeated presentations increased the likelihood that pictures were judged as coherent regardless of whether they were actually coherent. The current study extends this line of research by investigating whether visual coherence also impacts JOLs for pictures and to what extent fluency and beliefs contribute to effects of visual coherence on JOLs.

By examining visual coherence, we extend the literature on coherence effects in metamemory that has so far focused on semantic coherence (cf. Undorf & Zander, 2017). This is important because metacognitive processes might differ between pictorial and verbal materials (see, e.g., Besken, 2016; Undorf & Bröder, 2021). A crucial difference is that pictorial materials have higher imagery value than verbal materials (cf. Besken, 2016). Because people use imagery value as a cue for JOLs (e.g., Koriat, 1997), the information available to participants when predicting their future memory differs between pictorial and verbal materials. For instance, Carpenter and Olson (2012) found that participants assigned higher JOLs to foreign words that were paired with a picture during learning than to foreign words that were paired with a translation, despite similar actual memory performance. Pairing foreign words with a picture thus increased overconfidence in JOLs. Taken together, due to differences in the basis and accuracy of metacognitive judgments for pictorial and verbal materials, it was an open question whether coherence would also affect JOLs for pictorial materials.

The Current Study

In the current study, six experiments investigated the effect of visual coherence on JOLs. In each experiment, participants saw coherent and incoherent pictures and made a JOL for each picture. We hypothesized that visual coherence would increase subjective memorability, resulting in higher JOLs for coherent than for incoherent pictures. Because Experiment 1 revealed a robust effect of visual coherence on JOLs, the subsequent experiments investigated the contributions of fluency and beliefs to this effect. In Experiment 2, we manipulated fluency independent of visual coherence by changing the figure-ground contrast of coherent and incoherent pictures. We expected higher JOLs for coherent than for incoherent pictures as well as higher JOLs for high-contrast than for low-contrast pictures. To obtain more direct evidence for contributions of fluency, Experiments 3-6 included independent measures of processing fluency. Fluency was operationally defined as self-paced study time in Experiments 3 and 5, as identification time from a continuous identification task in Experiment 4, and as subjective fluency ratings in Experiment 6. If fluency contributes to the visual coherence effect on JOLs, independent fluency measures should indicate that coherent pictures are processed more fluently than incoherent pictures (i.e., shorter self-paced study time, shorter identification time, and higher subjective fluency ratings). Furthermore, the visual coherence effect on JOLs should be mediated by fluency.

To test whether beliefs contribute to the visual coherence effect on JOLs, we asked participants about potential gestalt perceptions after the JOL phase. The rationale was that only participants who report that they have perceived gestalts during the JOL phase could base their JOLs on relevant beliefs. If beliefs contribute to the visual coherence effect on JOLs, this effect should be more pronounced in people who are aware of differences in visual coherence than in people who are unaware of differences in visual coherence.

Experiment 1

In Experiment 1, participants saw coherent and incoherent pictures for 1 s each. Immediately after the presentation of each picture, participants made a JOL. After reporting potential gestalt perceptions, participants took a recognition test.

Method

Participants

The sample size for this experiment was determined by a power analysis using G*Power 3 (Faul et al., 2007). It showed that a sample size of 36 participants provided us with high power (1 - β = .90) to detect a medium effect (Cohen's d = 0.50) with α = .05 in paired t-tests. Participants were 40 University of Mannheim undergraduates.

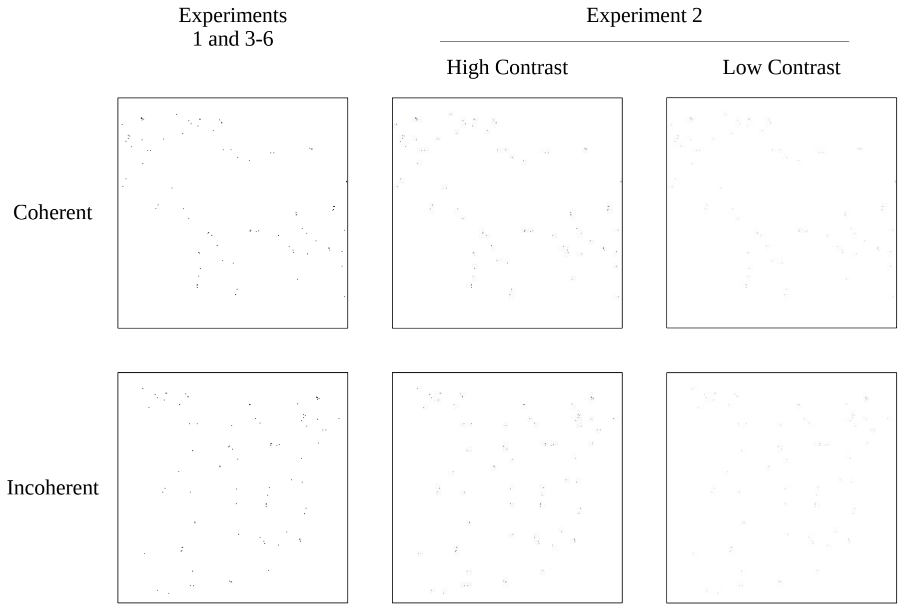
Materials

We constructed two lists of 15 coherent and 15 incoherent pictures from a stimulus set compiled by Topolinski and Strack (2009). The two lists were counterbalanced across participants such that each served equally often as target list and distractor list. All pictures were scaled to a size of 300 × 300 pixel. Examples of coherent and incoherent pictures are shown in Figure 1.

Procedure

At the outset of the experiment, participants were told that the experiment examined perceptions of abstract pictures. They were told that they will see 30 abstract black-and-white pictures and will have to judge how easy it would be for them to recognize each picture later. All trials were administered on a

Figure 1
Examples of Stimuli Used in Experiments 1-6



Note. The top row presents a coherent picture showing a moose, the bottom row presents an incoherent picture. The high-contrast picture has RGB values of R = 100, G = 100, B = 255 and the low-contrast picture has RGB values of R = 180, G = 180, B = 255.

black screen and started with a white fixation cross shown for 500 ms. Immediately afterwards, a picture was presented for 1,000 ms at the center of the screen, followed by a blank interval of 250 ms. Participants then made a JOL on a 10-point Likert scale ranging from 1 (*very difficult to recognize*) to 10 (*very easy to recognize*) using labeled keys (the keys 1 to 0 were labeled 1 to 10). The next trial followed immediately. After the last JOL, participants were asked whether they had perceived a gestalt. If so, they indicated how many pictures had depicted a gestalt and described the gestalts as detailed as possible. Participants then performed an unrelated visual search task for 5 min. Afterwards, they completed a recognition memory test with all 30 studied pictures and 30 new pictures. Each picture was again presented on a black screen, and participants clicked on buttons labeled "old" or "new". At study and test, pictures appeared in a random order for each participant except for the first two pictures that served as buffer items, were always presented first, and were discarded from all analyses.

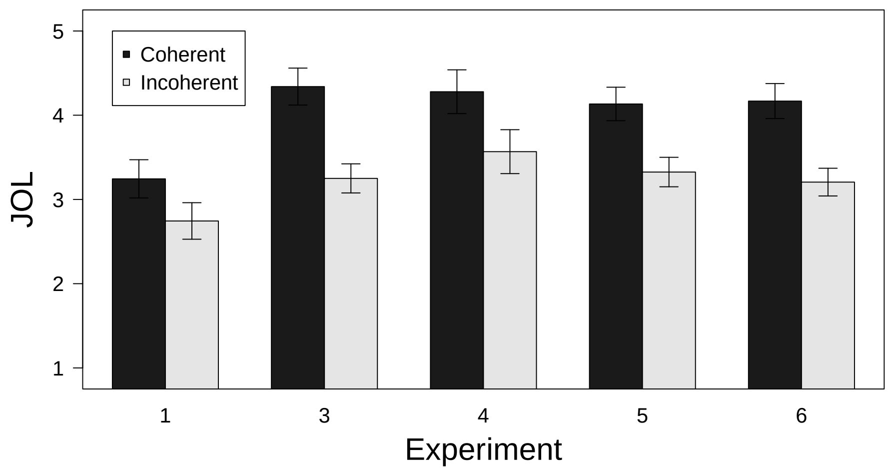
Results and Discussion

As can be seen in Figure 2, JOLs were higher for coherent than for incoherent pictures, $t(39) = 4.95$, $p < .001$, $d = 0.36$. As can be seen in Table 1, the same was true for corrected hit rates P_r, $t(39) = 2.19$, $p = .035$, $d = 0.44$. In contrast, bias index B_r did not differ between coherent and incoherent pictures, $t(39) = 0.26$, $p = .798$, $d = 0.04$. For completeness, we report correlations between JOLs and memory performance for this experiment and the following experiments in Appendix A.

To test for differences between participants who reported that they had perceived a gestalt ($n = 23$) and those who had not perceived a gestalt ($n = 17$), we conducted a mixed ANOVA on JOLs with coherence (coherent, in-

Figure 2

Mean Judgments of Learning (JOLs) in Experiment 1 and Experiments 3-6



Note. Error bars represent one standard error of the mean.

Table 1
Descriptive Statistics for Experiment 1 and 3-6

Experiment and Condition	Fluency measure	Hit rate	False alarm rate	P_r	B_r
Experiment 1					
Coherent	–	.50 (.17)	.39 (.16)	.10 (.16)	.44 (.16)
Incoherent	–	.46 (.20)	.43 (.19)	.03 (.19)	.45 (.19)
Experiment 3					
Coherent	6.75 (5.22)	.61 (.18)	.28 (.18)	.32 (.27)	.41 (.21)
Incoherent	7.56 (6.21)	.54 (.18)	.36 (.20)	.18 (.24)	.42 (.21)
Experiment 4					
Coherent	11.02 (7.46)	.49 (.18)	.33 (.15)	.16 (.21)	.40 (.15)
Incoherent	11.83 (7.61)	.42 (.20)	.36 (.19)	.07 (.20)	.38 (.18)
Experiment 5					
Coherent	2.21 (0.77)	.50 (.18)	.34 (.17)	.16 (.24)	.40 (.16)
Incoherent	2.25 (0.81)	.47 (.20)	.40 (.15)	.06 (.18)	.44 (.16)
Experiment 6					
Coherent	3.93 (1.56)	.56 (.20)	.33 (.18)	.23 (.20)	.43 (.20)
Incoherent	3.55 (1.43)	.45 (.18)	.41 (.18)	.05 (.18)	.42 (.17)

Note. Fluency measure was self-paced study time (Experiments 3 and 5), identification time (Experiment 4) and subjective fluency ratings (Experiment 6). P_r = corrected hit rate (hits - false alarms; Snodgrass & Corwin, 1988); B_r = bias index (false alarms/[1 - (hits - false alarms)]; Snodgrass & Corwin, 1988).

coherent) as a within-subjects factor and awareness (aware, unaware) as a between-subjects factor (see Appendix B for descriptive statistics). A significant main effect of coherence indicated higher JOLs for coherent pictures, $F(1,\ 38) = 22.76$, $MSE = 0.21$, $p < .001$, $\eta_p^2 = .38$. Neither the main effect of awareness nor the interaction were significant, both Fs < 1, indicating that the visual coherence effect on JOLs was similar in participants who were aware and participants who were unaware of differences in visual coherence.

Taken together, Experiment 1 showed that visual coherence impacts JOLs and memory performance. This finding was independent of whether participants were aware or unaware of differences in visual coherence, hinting that beliefs did not contribute to the visual coherence effect on JOLs. The following experiments focused on evaluating the contributions of fluency to the visual coherence effect on JOLs.

Experiment 2

Experiment 2 aimed to replicate and extend Experiment 1 to a situation where stimuli varied not only in visual coherence, but also in figure-ground contrast. Figure-ground contrast is viewed as a fluency manipulation, because it has been shown to influence participants' experiences of how easy visual stimuli are processed (e.g., Reber & Schwarz, 1999; Unkelbach, 2007). Moreover, Topolinski and Strack (2009) showed that people judged word triads more often as coherent when these were presented with a high figure-ground contrast than with a low figure-ground contrast, independent of whether word triads were actually coherent. We therefore expected higher JOLs for coherent than for incoherent pictures as well as for high-contrast than for low-contrast pictures.

Method

Participants

The sample size for this experiment was again determined by a power analysis using G*Power 3 (Faul et al., 2007). The power analysis showed that a sample size of 44 participants provided us with high power ($1 - \beta = .90$) to detect a medium effect (Cohen's $f = 0.25$) with $\alpha = .05$ and a correlation of .50 between repeated measures when conducting a 2 (Coherence: coherent, incoherent) $\times$ 2 (Contrast: high, low) within-subjects ANOVA. Participants were 44 University of Mannheim undergraduates.

Materials and Procedure

We used the same pictures as in Experiment 1 and manipulated fluency by changing each picture's figure-ground contrast. Following Unkelbach (2007) and Topolinski and Strack (2009), we created two versions of high- and low-contrast pictures to ensure that a potential effect of figure-ground contrast did not depend on a specific set of RGB (red, green, blue) values. High-contrast pictures had RGB values of R = 80, G = 80, B = 255, or R = 100, G = 100, B = 255, and low-contrast pictures had RGB values of R = 160, G = 160, B = 255, or R = 180, G = 180, B = 255. Importantly, all sets of RGB values decreased figure-ground contrast compared to the pictures used in Experiment 1 (see Figure 1). The procedure was identical to that of Experiment 1.

Results and Discussion

JOLs and recognition memory performance for coherent and incoherent pictures with high and low figure-ground contrast are reported in Figure 3 and Table 2, respectively. Neither measure depended on the pictures' specific RGB values.

JOLs were submitted to a two-way ANOVA with coherence (coherent, incoherent) and contrast (high, low) as within-subjects factors. As expected, a significant main effect of coherence revealed higher JOLs for coherent than for incoherent pictures, $F(1, 43) = 12.92$, $MSE = 0.40$, $p < .001$, $\eta_p^2 = .23$. A significant main effect of contrast revealed higher JOLs for high-contrast than for low-contrast pictures, $F(1, 43) = 35.21$, $MSE = 0.46$, $p < .001$, $\eta_p^2 = .45$. A significant interaction showed a larger effect of contrast for coherent than for incoherent pictures, $F(1, 43) = 4.11$, $MSE = 0.21$, $p = .049$, $\eta_p^2 = .09$. Post hoc t-tests revealed that JOLs were higher for coherent than for incoherent high-contrast pictures, $t(43) = 3.77$, $p < .001$, $d = 0.34$, but were not significantly higher for coherent than for incoherent low-contrast pictures, $t(43) = 1.89$, $p = .066$, $d = 0.16$.

A 2 (coherence) $\times$ 2 (contrast) ANOVA on corrected hit rates P_r revealed no significant effects, coherence: $F(1, 43) = 0.02$, $MSE = 0.06$, $p = .891$, $\eta_p^2 < .01$, contrast: $F(1, 43) = 0.04$, $MSE = 0.07$, $p = .836$, $\eta_p^2 < .01$, Coherence $\times$ Contrast:

Table 2
Descriptive Statistics for Experiment 2

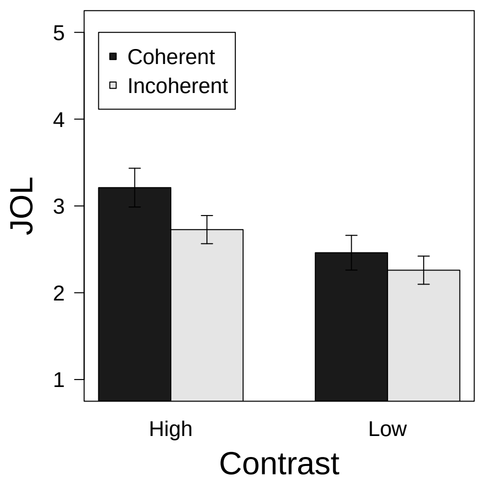
Condition	Hit rate	False alarm rate	P_r	B_r
High Contrast				
Coherent	.42 (.24)	.38 (.20)	.04 (.26)	.40 (.19)
Incoherent	.47 (.24)	.40 (.24)	.07 (.26)	.43 (.25)
Low Contrast				
Coherent	.44 (.19)	.36 (.25)	.08 (.27)	.36 (.22)
Incoherent	.47 (.22)	.43 (.27)	.04 (.22)	.44 (.25)

Note. P_r = corrected hit rate (hits - false alarms; Snodgrass & Corwin, 1988); B_r = bias index (false alarms/[1 - (hits - false alarms)]; Snodgrass & Corwin, 1988).

$F(1, 43) = 1.11$, $MSE = 0.05$, $p = .297$, $\eta_p^2 = .03$. The same was true for bias index B_r, coherence: $F(1, 43) = 3.67$, $MSE = 0.03$, $p = .062$, $\eta_p^2 = .08$, contrast: $F(1, 43) = 0.27$, $MSE = 0.04$, $p = .605$, $\eta_p^2 = .01$, Coherence $\times$ Contrast: $F(1, 43) = 1.25$, $MSE = 0.02$, $p = .271$, $\eta_p^2 = .03$.

Twenty-seven participants (61.36%) reported the perception of a gestalt, and 17 participants (38.64%) did not report the perception of a gestalt. A mixed ANOVA on JOLs with coherence (coherent, incoherent) and contrast (high, low) as within-subjects factors and awareness (aware, unaware) as a between-subjects factor revealed significant main effects of coherence, $F(1, 42) = 11.05$, $MSE = 0.41$, $p = .002$, $\eta_p^2 = .21$, and contrast, $F(1, 42) = 31.94$, $MSE = 0.47$, $p < .001$, $\eta_p^2 = .43$, and a significant Coherence $\times$ Contrast interaction,

Figure 3
Mean Judgments of Learning (JOLs) in Experiment 2



Note. Error bars represent one standard error of the mean.

$F(1, 42) = 5.75$, $MSE = 0.21$, $p = .021$, $\eta_p^2 = .12$. No other effects were significant, awareness: $F(1, 42) = 1.08$, $MSE = 5.42$, $p = .304$, $\eta_p^2 = .03$, Awareness $\times$ Coherence: $F(1, 42) = 0.46$, $MSE = 0.41$, $p = .501$, $\eta_p^2 = .01$, Awareness $\times$ Contrast: $F(1, 42) = 0.08$, $MSE = 0.47$, $p = .777$, $\eta_p^2 < .01$, Awareness $\times$ Coherence $\times$ Contrast: $F(1, 42) = 2.82$, $MSE = 0.21$, $p = .100$, $\eta_p^2 = .06$. As in Experiment 1, these results indicated that the visual coherence effect on JOLs was similar in participants who were aware and participants who were unaware of differences in visual coherence.

Replicating Experiment 1, visual coherence increased JOLs. New findings were that JOLs were higher for high-contrast than for low-contrast pictures and that the visual coherence effect on JOLs was present in high-contrast but not in low-contrast pictures. In contrast to the previous experiment, visual coherence did not affect recognition memory performance. However, as memory performance in this experiment was generally low, this finding must be interpreted with caution as it could be due to floor effects.

As in Experiment 1, the visual coherence effect on JOLs was independent of whether participants were aware or unaware of differences in visual coherence. Hence, we did not find evidence that beliefs contributed to the visual coherence effect on JOLs.

The finding that visual coherence did not significantly affect JOLs when figure-ground contrast was low suggested that some level of fluency is required for the visual coherence effect on JOLs to occur. The subsequent experiments aimed to provide more direct evidence for contributions of fluency to the visual coherence effect on JOLs. Each experiment obtained an independent measure of fluency in order to (1) test whether coherent pictures are processed more fluently than incoherent pictures and (2) examine whether fluency mediates the effect of visual coherence on JOLs.

Experiment 3

In Experiment 3, we operationally defined the fluency of processing coherent and incoherent pictures as self-paced study time, with lower self-paced study time indicating higher fluency (Koriat et al., 2006; Undorf & Erdfelder, 2011, 2015). If fluency contributes to the visual coherence effect on JOLs, self-paced study time should be shorter for coherent than for incoherent pictures. Furthermore, if fluency underlies the effect of visual coherence on JOLs, this effect should be mediated by self-paced study time.

Method

Participants, Materials and Procedure

The sample size was based on the same power analysis as in Experiment 1. Participants were 49 University of Mannheim undergraduates. Stimuli were the same as in Experiment 1. The procedure was identical to that in Experiment 1 with the following exceptions. The experiment was conducted online. Participants were instructed to study each picture for as long as they needed to recognize it in the recognition memory test towards the end of the experiment. Participants were asked to press the space bar when they were done studying a picture. Immediately afterwards, the JOL scale was presented together with ten on-screen buttons labeled 1 to 10.

Results and Discussion

Trials with study time below or above 2.5 standard deviations from each participant's mean (2.77%) were removed from the analyses (for a similar approach, see Koriat et al., 2018; Koriat et al., 2014).

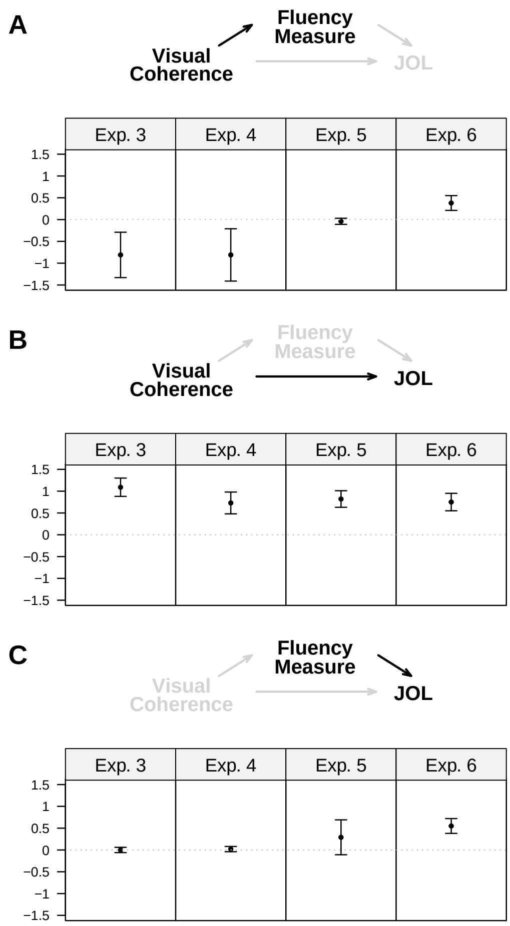
As can be seen in Table 1, self-paced study time was shorter for coherent than for incoherent pictures, $t(48) = 3.05$, $p = .004$, $d = 0.12$. Replicating the previous experiments, JOLs were higher for coherent than for incoherent pictures, $t(48) = 10.59$, $p < .001$, $d = 0.71$ (see Figure 2). Also as in Experiment 1, corrected hit rates P_r were higher for coherent than for incoherent pictures, $t(48) = 3.66$, $p < .001$, $d = 0.55$ (see Table 1). As before, bias index B_r was similar for coherent and incoherent pictures, $t(48) = 0.26$, $p = .794$, $d = 0.04$.

To examine whether the influence of visual coherence on JOLs was mediated by self-paced study time we first regressed (a) self-paced study time on dummy-coded visual coherence (0 = incoherent, 1 = coherent) and (b) JOLs on dummy-coded visual coherence and self-paced study time in two separate multilevel regression models (level 1: items, level 2: participants; cf. Kenny et al., 2003; Krull & MacKinnon, 2001). Participants were treated as random effects and visual coherence and self-paced study time were treated as fixed effects in both models. Regression analyses were conducted using the *R* packages *lme4* and *lmerTest* (Bates et al., 2015; Kuznetsova et al., 2017; R Core Team, 2020). As can be seen in Panel A of Figure 4, visual coherence reduced self-paced study time by 0.81 s, $b = -0.81$, 95% CI [-1.33, -0.29]. At the same time, visual coherence increased JOLs by 1.09 points when self-paced study time was controlled for (Panel B), $b = 1.09$, 95% CI [0.88, 1.30]. Unexpectedly, however, self-paced study time did not affect JOLs when visual coherence was controlled for (Panel C), $b = 0.00$, 95% CI [-0.06, 0.06]. Consequently, a mediation analysis using the *R* package *mediation* (Imai et al., 2010; see also Tingley et al., 2014) with 5000 simulations showed that the indirect effect was not significant (0.00, 95% CI [-0.05, 0.05], $p = .991$), confirming that self-paced study time did not mediate the effect of visual coherence on JOLs.

Forty-three participants (87.76%) reported the perception of a gestalt, and six participants (12.24%) did not report the perception of a gestalt. A mixed ANOVA on JOLs with coherence (coherent, incoherent) as a within-subjects factor and awareness (aware, unaware) as a between-subjects factor revealed a significant main effect of coherence with higher JOLs for coherent than for incoherent pictures, $F(1, 47) = 39.11$, $MSE = 0.26$ $p < .001$, $\eta_p^2 = .45$. Neither the main effect of awareness nor the interaction were significant, both Fs < 1. This indicates that the visual coherence effect on JOLs was similar in participants who were aware and participants who were unaware of differences in visual coherence.

Taken together, as in Experiment 1, visual coherence increased JOLs and memory performance. Also as in Experiment 1, the visual coherence effect on JOLs was independent of whether participants were aware or unaware of differences in visual coherence, arguing against contributions of beliefs to this effect. A new finding was that mean self-paced study time was shorter for coherent than for incoherent pictures. Inconsistent with our expectations,

Figure 4
Unstandardized Regression Coefficients for Experiments 3-6



Note. Panel A represents direct effects of visual coherence on the fluency measures. The fluency measure was expected to be inversely related to visual coherence and JOLs in Experiments 3-5 (shorter self-paced study time and shorter identification time indicate higher fluency) but to be positively related to visual coherence and JOLs in Experiment 6 (higher subjective fluency ratings indicate higher fluency). Panel B represents direct effects of visual coherence on judgments of learning (JOL). Panel C represents direct effects of the fluency measures on JOL. Error bars represent 95% confidence intervals.

however, a mediation analysis revealed that self-paced study time did not mediate the effect of visual coherence on JOLs. This finding is inconsistent with the idea that fluency underlies the visual coherence effect on JOLs. It is possible, however, that this result is due to limited validity of self-paced study time for measuring fluency (e.g., Witherby & Tauber, 2017; Yang et al., 2018; Yang et al., 2021). We therefore used another independent measure of fluency in Experiment 4.

Experiment 4

In Experiment 4, we operationally defined fluency as the time participants needed to identify the pictures in a continuous identification task, with lower identification time indicating higher fluency (e.g., Sanborn et al., 2004; Yang et al., 2018). If fluency contributes to the visual coherence effect on JOLs, identification time should be shorter for coherent than for incoherent pictures. As in Experiment 3, we used a mediation analysis to test whether fluency underlies the effect of visual coherence on JOLs.

Method

Participants, Materials and Procedure

The sample size was based on the same power analysis as in Experiment 1. Participants were 44 University of Mannheim undergraduates. We excluded one participant who stated that they interrupted the experiment despite instructions to complete it in one go. This resulted in a final sample size of 43 participants.

Stimuli were the same as in Experiment 1. The procedure was identical to that in Experiment 3 with the following exceptions. We used a continuous identification task to measure the fluency of coherent and incoherent pictures (e.g., Sanborn et al., 2004; Yang et al., 2018). Participants were asked to identify a stimulus that was initially shown very shortly (17 ms) and was backwardly masked (983 ms). In each subsequent trial, the presentation time of the stimulus increased by 17 ms (34 ms, 51 ms, 67 ms, and so on), whereas the presentation time of the mask decreased by 17 ms (966 ms, 949 ms, 932 ms, and so on). Each trial was therefore 1,000 ms long, but the stimulus became easier to identify as its presentation time increased. The mask was generated with *PsychoPy2* (Peirce et al., 2019) using a grating stimulus with the RGB values R = 129, G = 129, B = 129, an orientation of 45 degrees, a *sin* texture, and a spatial frequency of 200. Participants were instructed to stop the identification task as soon as they had studied the picture long enough.

Results and Discussion

As in Experiment 3, trials with identification time below or above 2.5 standard deviations from each participant's mean (1.91%) were removed from the analyses.

As can be seen in Table 1, identification time was shorter for coherent than for incoherent pictures, $t(42) = 2.68$, $p = .010$, $d = 0.11$. As in all previous experiments, JOLs were higher for coherent than for incoherent pictures, $t(42) = 5.62$, $p < .001$, $d = 0.42$ (see Figure 2). The same was true for corrected hit rates P_r, $t(42) = 2.48$, $p = .017$, $d = 0.44$ (see Table 1). Again, bias index B_r did not differ between coherent and incoherent pictures, $t(42) = 0.83$, $p = .410$, $d = 0.12$.

Regression analyses (see Figure 4) revealed that visual coherence reduced identification time (Panel A), $b = -0.81$, 95% CI [-1.41, -0.22]. While visual coherence increased JOLs when identification time was controlled for (Panel B), $b = 0.73$, 95% CI [0.48, 0.99], identification time did not affect JOLs when visual coherence was controlled for (Panel C), b = 0.02, 95% CI [-0.04, 0.08]. As in Experiment 3, a mediation analysis again showed that the indirect effect was not significant (-0.02, 95% CI [-0.08, 0.03], p = .420), confirming that identification time did not mediate the effect of visual coherence on JOLs.

In this experiment, 30 participants (69.77%) reported the perception of a gestalt, and 13 participants (30.23%) did not report the perception of a gestalt. A mixed ANOVA on JOLs with coherence (coherent, incoherent) as a within-subjects factor and awareness (aware, unaware) as a between-subjects factor revealed a significant main effect of coherence with higher JOLs for coherent pictures, $F(1, 41) = 21.65$, $MSE = 0.34$, $p < .001$, $\eta_p^2 = .35$. Neither the main effect of awareness nor the interaction was significant, awareness: $F(1, 41) = 0.21$, $MSE = 5.59$, $p = .652$, $\eta_p^2 < .01$, Coherence $\times$ Awareness: $F(1, 41) = 2.09$, $MSE = 0.34$, $p = .156$, $\eta_p^2 = .05$. This again indicated that the visual coherence effect on JOLs was similar in participants who were aware and participants who were unaware of differences in visual coherence.

Taken together, Experiment 4 closely replicated the previous experiments in showing that visual coherence increased JOLs and memory performance. As in Experiment 3, the visual coherence effect on JOLs was independent of whether participants were aware or unaware of differences in visual coherence. A new finding was that mean identification time was shorter for coherent than for incoherent pictures, showing that coherent pictures were processed more fluently than incoherent pictures. However, identification time did not mediate the effect of visual coherence on JOLs. This pattern of results is identical to that obtained in Experiment 3 and again argues against the idea that fluency underlies the visual coherence effect on JOLs.

One specific aspect of the results in Experiments 3 and 4, however, suggests that the validity of self-paced study time and identification time as fluency measures might have been limited. In particular, self-paced study time and identification time both were quite long (self-paced study time: M = 7.16 s; identification time: M = 11.43 s) and much longer than the fixed 1-s-presentation times used in Experiments 1 and 2. It is therefore possible that self-paced study time and identification time failed to reflect the quick and intuitive feelings of fluency that might form the basis of people's JOLs. We addressed this issue in Experiment 5.

Experiment 5

In Experiment 5, we returned to using self-paced study time as a fluency measure.[1] We encouraged participants to respond quickly, expecting that the study times would then reflect quick and intuitive feelings of fluency (e.g., Koriat & Levy-Sadot, 1999; Topolinski & Strack, 2009). If fluency contributes to the visual coherence effect on JOLs, self-paced study time should be shorter for coherent than for incoherent pictures. Furthermore, the effect of visual coherence on JOLs should be mediated by self-paced study time.

[1]Because the CID task requires a certain number of trials for stimuli to be recognized at all (see Sanborn et al., 2004; Yang et al., 2018), it was unsuited to being restricted.

Method

Participants, Materials and Procedure

The sample size was based on the same power analysis as in Experiment 1. Participants were 41 University of Mannheim undergraduates. We excluded two participants who stated that they interrupted the experiment despite instructions to complete it in one go. This resulted in a final sample size of 39 participants. Stimuli were the same as in Experiment 1.

The procedure was identical to that in Experiment 3 with the following exception. Participants were instructed to study each picture no longer than 5 s. They were told that the best strategy would be to focus on each picture as a whole rather than to concentrate on specific details. Furthermore, participants were informed that the experiment would be terminated if their study time exceeded 5 s in more than three trials, with their compensation being reduced accordingly. To acclimate participants to this requirement, a countdown timer was displayed on the screen while participants studied the first two items (buffer items).

Results and Discussion

We did not have to exclude any participant because their study time exceeded 5 s in more than three trials. Across all participants, study time exceeded 5 s in 17 trials (1.56%). As in Experiments 3 and 4, trials with study time below or above 2.5 standard deviations from each participant's mean (1.83%) were removed from the analyses.

Unlike in Experiment 3, study time was similar for coherent and incoherent pictures, $t(38) = 1.16$, $p = .254$, $d = 0.05$ (see Table 1). As in the previous experiments, JOLs were higher for coherent than for incoherent pictures, $t(38) = 8.58$, $p < .001$, $d = 0.67$ (see Figure 2). The same was true for corrected hit rates P_r, $t(38) = 2.19$, $p = .035$, $d = 0.44$ (see Table 1). Again, bias index B_r did not differ between coherent and incoherent pictures, $t(38) = 1.61$, $p = .116$, $d = 0.24$. Regression analyses (see Figure 4) revealed that visual coherence did not reduce study time (Panel A), $b = -0.04$, 95% CI [-0.11, 0.03]. While visual coherence increased JOLs when study time was controlled for (Panel B), $b = 0.82$, 95% CI [0.63, 1.01], study time did not affect JOLs when visual coherence was controlled for (Panel C), $b = 0.29$, 95% CI [-0.12, 0.69]. As in the previous experiments, a mediation analysis showed that the indirect effect was not significant (-0.01, 95% CI [-0.05, 0.01], $p = .381$), confirming that study time did not mediate the influence of visual coherence on JOLs.

In this experiment, 34 participants (87.18%) reported the perception of a gestalt, and five participants (12.82%) did not report the perception of a gestalt. A mixed ANOVA on JOLs with coherence (coherent, incoherent) as a within-subjects factor and awareness (aware, unaware) as a between-subjects factor revealed significant main effects of coherence and awareness, with higher JOLs for coherent pictures and aware participants, coherence: $F(1, 37) = 26.11$, $MSE = 0.18$, $p < .001$, $\eta_p^2 = .41$, awareness: $F(1, 37) = 9.85$, $MSE = 2.08$, $p = .003$, $\eta_p^2 = .21$. Most importantly, the interaction was not significant, $F(1, 37) = 0.65$, $MSE = 0.18$, $p = .424$, $\eta_p^2 = .02$, again indicating that the visual coherence effect on JOLs was similar in participants who were aware and participants who were unaware of differences in visual coherence.

Taken together, Experiment 5 again showed higher JOLs and memory performance for coherent than for incoherent pictures. As in the previous experiments, the visual coherence effect on JOLs was independent of whether participants were aware or unaware of differences in visual coherence.

At the same time, Experiment 5 revealed no evidence for contributions of fluency to the visual coherence effect on JOLs: Self-paced study time did not differ between coherent and incoherent pictures and, consequently, did not mediate the effect of visual coherence on JOLs. It is unclear why self-paced study time did not differ between coherent and incoherent pictures, because we have found clear differences in self-paced study time and identification time in the previous experiments. One possibility is that the 5s-time made participants focus on keeping their study time short at the expense of ignoring their feelings of fluency. This is congruent with research suggesting that objective measures of fluency sometimes fail to capture people's feelings of fluency (e.g., Wang et al., 2020; Yang et al., 2021). In these cases, it may be more appropriate to assess fluency through subjective measures (e.g., Reber, Wurtz, et al., 2004; Winkielman et al., 2003). We addressed this possibility in Experiment 6.

Experiment 6

Experiment 6 examined whether subjective fluency ratings underlie the visual coherence effect on JOLs. Prior to the study phase with JOLs, participants indicated how fluently they perceived each picture (for similar measures of subjective fluency, see Graf et al., 2018; Reber, Wurtz, et al., 2004). If subjective fluency contributes to the visual coherence effect on JOLs, ratings should be higher for coherent than for incoherent pictures. Furthermore, the effect of visual coherence on JOLs should be mediated by subjective fluency.

Method

Participants, Materials and Procedure

The sample size was based on the same power analysis as in Experiment 1. Participants were 47 University of Mannheim undergraduates. We excluded four participants who stated that they interrupted the experiment despite instructions to complete it in one go. This resulted in a final sample size of 43 participants. Stimuli were the same as in Experiment 1.

The procedure was identical to that in Experiment 1 with the following exceptions. In the beginning of the experiment, we obtained a standard measure of subjective fluency (see, e.g., Graf et al., 2018; Reber, Wurtz, et al., 2004). Participants saw each picture for 1 s and then rated the ease of perceiving it on a 9-point Likert scale ranging from 1 (*very difficult to perceive*) to 9 (*very easy to perceive*). They then completed the study phase, in which each picture was presented for 1 s.

Results and Discussion

The mean within-subject correlation between subjective fluency ratings and JOLs was moderate, $r = .36$, $t(42) = 9.16$, $p < .001$, $d = 1.40$, showing that subjective fluency ratings were related but clearly not identical to JOLs.

Correlations did not differ between coherent and incoherent pictures, coherent: $r = .40$, $t(41) = 8.57$, $p < .001$, $d = 1.32$; incoherent: $r = .29$, $t(41) = 5.53$, $p < .001$, $d = 0.85$; difference: $t(40) = 1.82$, $p = .076$, $d = 0.32$.[2]

As can be seen in Table 1, subjective fluency ratings were higher for coherent than for incoherent pictures, $t(42) = 4.34$, $p < .001$, $d = 0.25$. As in the previous experiments, JOLs were higher for coherent than for incoherent pictures, $t(42) = 8.36$, $p < .001$, $d = 0.74$ (see Figure 2). The same was true for corrected hit rates P_r, $t(42) = 5.29$, $p < .001$, $d = 0.97$ (see Table 1). Again, bias index B_r did not differ between coherent and incoherent pictures, $t(42) = 0.07$, $p = .947$, $d < 0.01$.

Regression analyses (see Figure 4) revealed that visual coherence increased subjective fluency ratings (Panel A), $b = 0.38$, 95% CI [0.21, 0.55]. Visual coherence increased JOLs when subjective fluency ratings were controlled for (Panel B), $b = 0.75$, 95% CI [0.55, 0.96]. Most importantly, subjective fluency ratings increased JOLs when visual coherence was controlled for (Panel C), $b = 0.55$, 95% CI [0.38, 0.71]. A mediation analysis revealed a significant indirect effect, 0.21, 95% CI [0.10, 0.33], $p < .001$, showing that subjective fluency partially mediated the visual coherence effect on JOLs. The proportion of the total effect of visual coherence on JOLs that was mediated by subjective fluency was 0.22, 95% CI [0.11, 0.33].

In this experiment, 28 participants (65.12%) reported the perception of a gestalt, and 15 participants (34.88%) did not report the perception of a gestalt. A mixed ANOVA on JOLs with coherence (coherent, incoherent) as a within-subjects factor and awareness (aware, unaware) as a between-subjects factor revealed significant main effects of coherence and awareness, with higher JOLs for coherent pictures and aware participants, coherence: $F(1, 41) = 59.20$, $MSE = 0.25$, $p < .001$, $\eta_p^2 = .59$, awareness: $F(1, 41) = 4.79$, $MSE = 2.51$, $p = .034$, $\eta_p^2 = .11$. Unlike in all previous experiments, a significant interaction, $F(1, 41) = 6.28$, $MSE = 0.25$, $p = .016$, $\eta_p^2 = .13$, indicated that the visual coherence effect on JOLs was more pronounced in participants who were aware than for participants who were unaware of differences in visual coherence.

Taken together, Experiment 6 again showed higher JOLs and memory performance for coherent than for incoherent pictures. Like the objective fluency measures obtained in Experiments 3 and 4, subjective fluency ratings were higher for coherent than for incoherent pictures. In contrast to the previously obtained objective fluency measures, subjective fluency ratings partially mediated the visual coherence effect on JOLs. Experiment 6 thus indicated that fluency contributed to the visual coherence effect on JOLs.

Unlike in the previous experiments, a more pronounced visual coherence effect on JOLs in participants who were aware of differences in visual coherence suggested that beliefs also contributed to the visual coherence effect on JOLs. Because this finding was inconsistent with the previous experiments, we conducted a joint analysis of all experiments to take a closer look at the contribution of beliefs to the visual coherence effect on JOLs.

[2]Degrees of freedom vary because one participant assigned the same JOL to all coherent pictures and one participant assigned the same JOL all incoherent pictures.

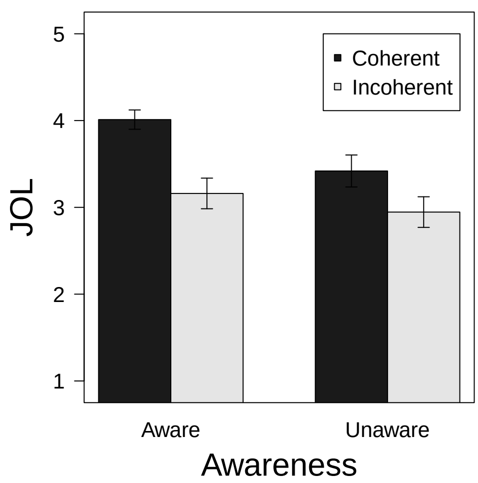
Contribution of Beliefs: A Joint Analysis

Experiments 1 to 5 showed that the visual coherence effect on JOLs was independent of whether participants were aware or unaware of differences in visual coherence, indicating that beliefs did not contribute to this effect. In contrast, in Experiment 6, the visual coherence effect on JOLs for was more pronounced in participants who were aware of differences in visual coherence. One reason for the inconsistent results might be that the numbers of participants who were unaware of differences in visual coherence were relatively small (n = 17 [42.50%], n = 17 [38.64%], n = 6 [12.24%], n = 13 [30.23%], n = 5 [12.82%], n = 15 [34.88%] in Experiments 1-6, respectively). This limited the statistical power for detecting potential contributions of beliefs to the visual coherence effect on JOLs in each separate experiment. We addressed this possible limitation with a joint analysis of all experiments.

Across all experiments, n = 185 participants were aware of differences in visual coherence and n = 73 participants were unaware of differences in visual coherence. A mixed ANOVA on JOLs with coherence (coherent, incoherent) as a within-subjects factor and awareness (aware, unaware) as a between-subjects factor revealed a significant main effect of coherence, indicating higher JOLs for coherent pictures, $F(1, 256) = 175.09$, $MSE = 0.26$, $p < .001$, $\eta_p^2 = .41$. The main effect of awareness was also significant, indicating higher JOLs for participants who were aware of differences in visual coherence, $F(1, 256) = 4.49$, $MSE = 3.79$, $p = .035$, $\eta_p^2 = .02$. Most importantly, a significant Coherence $\times$ Awareness interaction, $F(1, 256) = 14.22$, $MSE = 0.26$, $p < .001$, $\eta_p^2 = .05$, revealed that the visual coherence effect on JOLs was more pronounced in participants who were aware of differences in visual coherence (see Figure 5). Post hoc

Figure 5

Mean Judgments of Learning (JOLs) for Participants Who Were Aware and Unaware of Differences in Visual Coherence in Experiments 1-6



Note. Error bars represent one standard error of the mean.

t-tests showed that both aware and unaware participants made higher JOLs for coherent than for incoherent pictures, aware: $t(184) = 15.13$, $p < .001$, $d = 0.58$, unaware: $t(72) = 6.66$, $p < .001$, $d = 0.31$.

This joint analysis of Experiments 1-6 showed that the visual coherence effect on JOLs was more pronounced in participants who were aware than in participants who were unaware of differences in visual coherence. This result indicates that beliefs contribute to the visual coherence effect on JOLs.

General Discussion

The present study investigated the effect of visual coherence on JOLs. In all six experiments, JOLs were higher for coherent than for incoherent pictures, indicating that participants used visual coherence as a cue when making JOLs. Memory performance was higher for coherent than for incoherent pictures in five of six experiments, showing that visual coherence usually was a valid cue for predicting future memory performance. Concerning the processes underlying the visual coherence effect on JOLs, the current study provided evidence that both fluency and beliefs contribute to this effect, supporting the dual-basis view of metacognition (Kelley & Jacoby, 1996; Koriat, 1997, 2007).

Contributions of Fluency and Beliefs

Consistent with the notion that fluency contributes to the visual coherence effect on JOLs, JOLs were higher for pictures presented with a high than with a low figure-ground contrast (Experiment 2). Previous research showed that high figure-ground contrast increases perceptual fluency (Reber & Schwarz, 1999, 2001; Reber, Zimmermann, et al., 2004; Unkelbach, 2007) and that word triads presented with a high figure-ground contrast are more likely to be judged as coherent, irrespective of its actual coherence (Topolinski & Strack, 2009). In the current study, both figure-ground contrast and visual coherence impacted JOLs, hinting that experience-based, non-analytic processes contribute to the visual coherence effect on JOLs.

To provide a more direct test for contributions of experience-based, non-analytic processes to the visual coherence effect on JOLs, we obtained independent measures of fluency in Experiments 3 to 6. We operationally defined fluency as self-paced study time in Experiments 3 and 5, as identification time from a continuous identification task in Experiment 4, and as subjective fluency ratings in Experiment 6. Except for Experiment 5, all experiments indicated that coherent pictures were processed more fluency than incoherent pictures. Unexpectedly, however, neither self-paced study time nor identification time mediated the visual coherence effect on JOLs. In contrast, subjective fluency ratings partially mediated the effect of visual coherence on JOLs.

Why did subjective fluency but not objective fluency mediate the visual coherence effect on JOLs? Prior research demonstrated fluency differences between coherent and incoherent pictures when pictures were presented for 400 ms or 1 s (Topolinski & Strack, 2009; Volz & von Cramon, 2006). These quick and intuitive feelings of fluency presumably form the basis of people's JOLs (see Undorf & Zander, 2017) but were not well captured by the much longer self-paced study times and identification times obtained in Experiments 3 and 4 (self-paced study time: $M = 7.16$ s; identification time: $M = 11.43$ s). In contrast, the subjective fluency measure was obtained after presenting coherent and incoherent pictures for 1 s and therefore probably reflected the quick and intuitive feelings of fluency underlying JOLs. Consistent with this idea, research has shown that subjective fluency measures can capture the ease of processing at various stages of perceptual processes (e.g., Reber, Wurtz, et al., 2004; Winkielman et al., 2003). Further research is needed to test this speculation.

Concerning the contributions of beliefs, we found that the size of the visual coherence effect on JOLs did not differ between participants who were aware and participants who were unaware of differences in visual coherence when analyzing Experiments 1 to 5 separately. However, results from Experiment 6 as well as a joint analysis of all experiments showed that the visual coherence effect on JOLs was more pronounced in participants who were aware of differences in visual coherence. This provides evidence that theory-based, analytic processes contribute to the visual coherence effect on JOLs. The mixed results found in the separate experiments are presumably due to limited statistical power.

Importantly, however, the robust finding of a visual coherence effect on JOLs for participants who were unaware of differences in visual coherence showed that theory-based, analytic processes alone cannot explain this effect. Furthermore, the belief measure used in the current study might have overestimated the contribution of theory-based, analytic processes to the visual coherence effect on JOLs. The reason for this is that we categorized all participants who perceived a gestalt during the study phase as basing their JOLs on theory-based, analytic processes. It is possible, however, that some participants perceived a gestalt during the study phase but did not use this information to form a relevant belief. Future research is therefore needed to consider the role of beliefs in the visual coherence effect on JOLs more closely. Another avenue for future research is to explore potential contributions of subjective fluency to other cue effects in metamemory.

Coherence Effects in Metamemory

The current study demonstrates similarities between the visual coherence effect on JOLs and the semantic coherence effect on JOLs (Undorf & Zander, 2017), that is, higher JOLs for coherent word triads that share a common associate than for incoherent word triads that have no common associate. Specifically, both coherence effects produces higher JOLs for coherent than for incoherent stimuli and both effects rely on experience-based, non-analytic processes as well as on theory-based, analytic processes. As mentioned in the introduction, these similarities are not trivial because pictorial materials differ from verbal materials regarding memory encoding, and, presumably, metacognitive monitoring (Koriat, 1997; Mintzer & Snodgrass, 1999; Nelson et al., 1976). Similar coherence effects on JOLs for pictorial and verbal materials therefore suggests that coherence plays an important role in metamemory.

However, the current study also demonstrates a critical difference between the effects of semantic and visual coherence on JOLs. Specifically, Undorf and Zander (2017) found that reading time, an independent measure of objective fluency, mediated the semantic coherence effect on JOLs (Experiment 2: p = .033, Experiment 3: $p < .001$). In contrast, two independent measures of objective fluency, self-paced study time and identification time, did not mediate the visual coherence effect on JOLs. This difference across materials suggests that objective fluency measures might not be appropriate for assessing the experience-based, non-analytic processes that contribute to the visual coherence effect on JOLs.

Conclusion

Taken together, the present study showed that visual coherence impacts JOLs. Because visual coherence also impacts actual memory performance, it can be added to the list of valid cues that guide human metamemory and, potentially, self-regulated learning. We found evidence for contributions of experience-based, non-analytic processes and theory-based, analytic processes to the visual coherence effect on JOLs. Together with prior studies showing that both types of processes contribute to various cue effects on JOLs (Jia et al., 2016; Mendes et al., 2020, 2021; Mueller et al., 2013; Undorf & Erdfelder, 2015; Undorf & Zander, 2017), the present findings argue against the idea that effects on JOLs are mainly or even solely due to metacognitive beliefs (Jia et al., 2016; Li et al., 2016; Mueller & Dunlosky, 2017; Witherby & Tauber, 2017) or fluency (Begg et al., 1989; Hertzog et al., 2003; Koriat et al., 2004), but support the dual-basis view of metacognition (Kelley & Jacoby, 1996; Koriat, 1997, 2007).

References

Ball, B. H., Klein, K. N., & Brewer, G. A. (2014). Processing fluency mediates the influence of perceptual information on monitoring learning of educationally relevant materials. *Journal of Experimental Psychology: Applied*, *20*(4), 336–348. https://doi.org/10.1037/xap0000023

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Begg, I., Duft, S., Lalonde, P., Melnick, R., & Sanvito, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory and Language*, *28*(5), 610–632. https://doi.org/10.1016/0749-596X(89)90016-8

Besken, M. (2016). Picture-perfect is not perfect for metamemory: Testing the perceptual fluency hypothesis with degraded images. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(9), 1417–1433. https://doi.org/10.1037/xlm0000246

Bolte, A., & Goschke, T. (2008). Intuition in the context of object perception: intuitive gestalt judgments rest on the unconscious activation of semantic representations. *Cognition*, *108*(3), 608–616. https://doi.org/10.1016/j.cognition.2008.05.001

Bowers, K. S., Regehr, G., Balthazard, C., & Parker, K. (1990). Intuition in the context of discovery. *Cognitive Psychology*, *22*(1), 72–110. https://doi.org/10.1016/0010-0285(90)90004-N

Carpenter, S. K., & Olson, K. M. (2012). Are pictures good for learning new vocabulary in a foreign language? Only if you think they are not. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(1), 92–101. https://doi.org/10.1037/a0024828

Dunlosky, J., & Metcalfe, J. (2009). *Metacognition*. Sage.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/bf03193146

Frank, D. J., & Kuhlmann, B. G. (2017). More than just beliefs: Experience and beliefs jointly contribute to volume effects on metacognitive judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(5), 680–693. https://doi.org/10.1037/xlm0000332

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

Hertzog, C., Dunlosky, J., Robinson, A. E., & Kidder, D. P. (2003). Encoding fluency is a cue used for judgments about learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 22–34. https://doi.org/10.1037/0278-7393.29.1.22

Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods*, *15*(4), 309–334. https://doi.org/10.1037/a0020761

Jia, X., Li, P., Li, X., Zhang, Y., Cao, W., Cao, L., & Li, W. (2016). The effect of word frequency on judgments of learning: Contributions of beliefs and processing fluency. *Frontiers in Psychology*, *6*, 1995. https://doi.org/10.3389/fpsyg.2015.01995

Kelley, C. M., & Jacoby, L. L. (1996). Adult egocentrism: Subjective experience versus analytic bases for judgment. *Journal of Memory and Language*, *35*(2), 157–175. https://doi.org/10.1006/jmla.1996.0009

Kenny, D. A., Korchmaros, J. D., & Bolger, N. (2003). Lower level mediation in multilevel models. *Psychological Methods*, *8*(2), 115–128. https://doi.org/10.1037/1082-989X.8.2.115

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*(4), 349–370. https://doi.org/10.1037/0096-3445.126.4.349

Koriat, A. (2007). Metacognition and consciousness. In P. D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *The Cambridge handbook of consciousness* (pp. 289–325). Cambridge University Press. https://doi.org/10.1017/CBO9780511816789.012

Koriat, A., Adiv-Mashinsky, S., Undorf, M., & Schwarz, N. (2018). The prototypical majority effect under social influence. *Personality & Social Psychology Bulletin*, *44*(5), 670–683. https://doi.org/10.1177/0146167217744527

Koriat, A., Bjork, R. A., Sheffer, L., & Bar, S. K. (2004). Predicting one's own forgetting: The role of experience-based and theory-based processes. *Journal of Experimental Psychology: General*, *133*(4), 643–656. https://doi.org/10.1037/0096-3445.133.4.643

Koriat, A., & Levy-Sadot, R. (1999). Processes underlying metacognitive judgments: Information-based and experience-based monitoring of one's own knowledge. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 483–502). The Guilford Press.

Koriat, A., Ma'ayan, H., & Nussinson, R. (2006). The intricate relationships between monitoring and control in metacognition: Lessons for the cause-and-effect relation between subjective experience and behavior. *Journal of Experimental Psychology: General*, *135*(1), 36–69. https://doi.org/10.1037/0096-3445.135.1.36

Koriat, A., Nussinson, R., & Ackerman, R. (2014). Judgments of learning depend on how learners interpret study effort. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(6), 1624–1637. https://doi.org/10.1037/xlm0000009

Krull, J. L., & MacKinnon, D. P. (2001). Multilevel modeling of individual and group level mediated effects. *Multivariate Behavioral Research*, *36*(2), 249–277. https://doi.org/10.1207/S15327906MBR3602{\textunderscore}06

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Li, P., Jia, X., Li, X., & Li, W. (2016). The effect of animacy on metamemory. *Memory & Cognition*, *44*(5), 696–705. https://doi.org/10.3758/s13421-016-0598-7

Luna, K., Martín-Luengo, B., & Albuquerque, P. B. (2018). Do delayed judgements of learning reduce metamemory illusions? A meta-analysis. *Quarterly Journal of Experimental Psychology*, *71*, 1626–1636. https://doi.org/10.1080/17470218.2017.1343362

Mendes, P. S., Luna, K., & Albuquerque, P. B. (2020). Experience matters: Effects of (in)congruent prompts about word frequency on judgments of learning. *Zeitschrift für Psychologie*, *228*(4), 254–263. https://doi.org/10.1027/2151-2604/a000423

Mendes, P. S., Luna, K., & Albuquerque, P. B. (2021). Word frequency effects on judgments of learning: More than just beliefs. *The Journal of General Psychology*, *148*(2), 124–148. https://doi.org/10.1080/00221309.2019.1706073

Mintzer, M. Z., & Snodgrass, J. G. (1999). The picture superiority effect: Support for the distinctiveness model. *The American Journal of Psychology*, *112*(1), 113–146. https://doi.org/10.2307/1423627

Mueller, M. L., & Dunlosky, J. (2017). How beliefs can impact judgments of learning: Evaluating analytic processing theory with beliefs about fluency. *Journal of Memory and Language*, *93*, 245–258. https://doi.org/10.1016/j.jml.2016.10.008

Mueller, M. L., Dunlosky, J., & Tauber, S. K. (2016). The effect of identical word pairs on people's metamemory judgments: What are the contributions of processing fluency and beliefs about memory? *Quarterly Journal of Experimental Psychology*, *69*(4), 781–799. https://doi.org/10.1080/17470218.2015.1058404

Mueller, M. L., Dunlosky, J., Tauber, S. K., & Rhodes, M. G. (2014). The font-size effect on judgments of learning: Does it exemplify fluency effects or reflect people's beliefs about memory? *Journal of Memory and Language*, *70*, 1–12. https://doi.org/10.1016/j.jml.2013.09.007

Mueller, M. L., Tauber, S. K., & Dunlosky, J. (2013). Contributions of beliefs and processing fluency to the effect of relatedness on judgments of learning. *Psychonomic Bulletin & Review*, *20*, 378–384. https://doi.org/10.3758/s13423-012-0343-6

Nelson, D. L., Reed, V. S., & Walling, J. R. (1976). Pictorial superiority effect. *Journal of Experimental Psychology: Human Learning and Memory*, *2*(5), 523–528. https://doi.org/10.1037/0278-7393.2.5.523

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

R Core Team. (2020). R: A language and environment for statistical computing. https://www.R-project.org/

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

Reber, R., & Schwarz, N. (2001). The hot fringes of consciousness. *Consciousness & Emotion*, *2*(2), 223–231. https://doi.org/10.1075/ce.2.2.03reb

Reber, R., Wurtz, P., & Zimmermann, T. D. (2004). Exploring "fringe" consciousness: The subjective experience of perceptual fluency and its objective bases. *Consciousness and Cognition*, *13*(1), 47–60. https://doi.org/10.1016/S1053-8100(03)00049-7

Reber, R., Zimmermann, T. D., & Wurtz, P. (2004). Judgments of duration, figure-ground contrast, and size for words and nonwords. *Consciousness and Cognition*, *66*(7), 1105–1114. https://doi.org/10.3758/BF03196839

Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology: General*, *137*(4), 615–625. https://doi.org/10.1037/a0013684

Rhodes, M. G., & Castel, A. D. (2009). Metacognitive illusions for auditory information: Effects on monitoring and control. *Psychonomic Bulletin & Review*, *16*(3), 550–554. https://doi.org/10.3758/PBR.16.3.550

Sanborn, A. N., Malmberg, K. J., & Shiffrin, R. M. (2004). High-level effects of masking on perceptual identification. *Vision Research*, *44*(12), 1427–1436. https://doi.org/10.1016/j.visres.2004.01.004

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, *117*(1), 34–50. https://doi.org/10.1037/0096-3445.117.1.34

Susser, J. A., & Mulligan, N. W. (2015). The effect of motoric fluency on metamemory. *Psychonomic Bulletin & Review*, *22*(4), 1014–1019. https://doi.org/10.3758/s13423-014-0768-1

Susser, J. A., Panitz, J., Buchin, Z., & Mulligan, N. W. (2017). The motoric fluency effect on metamemory. *Journal of Memory and Language*, *95*, 116–123. https://doi.org/10.1016/j.jml.2017.03.002

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). mediation: R package for causal mediation analysis. *Journal of Statistical Software*, *59*(5), 1–38. https://doi.org/10.18637/jss.v059.i05

Topolinski, S., & Strack, F. (2009). The architecture of intuition: Fluency and affect determine intuitive judgments of semantic and visual coherence and judgments of grammaticality in artificial grammar learning. *Journal of Experimental Psychology: General*, *138*(1), 39–63. https://doi.org/10.1037/a0014678

Undorf, M. (2020). Fluency illusions in metamemory. In A. M. Cleary & B. L. Schwartz (Eds.), *Memory quirks* (pp. 150–174). Routledge. https://doi.org/10.4324/9780429264498-12

Undorf, M., & Bröder, A. (2021). Metamemory for pictures of naturalistic scenes: Assessment of accuracy and cue utilization. *Memory & Cognition*, *49*(7), 1405–1422. https://doi.org/10.3758/s13421-021-01170-5

Undorf, M., & Erdfelder, E. (2011). Judgments of learning reflect encoding fluency: conclusive evidence for the ease-of-processing hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(5), 1264–1269. https://doi.org/10.1037/a0023719

Undorf, M., & Erdfelder, E. (2015). The relatedness effect on judgments of learning: A closer look at the contribution of processing fluency. *Memory & Cognition*, *43*(4), 647–658. https://doi.org/10.3758/s13421-014-0479-x

Undorf, M., Söllner, A., & Bröder, A. (2018). Simultaneous utilization of multiple cues in judgments of learning. *Memory & Cognition*, *46*, 507–519. https://doi.org/10.3758/s13421-017-0780-6

Undorf, M., & Zander, T. (2017). Intuition and metacognition: The effect of semantic coherence on judgments of learning. *Psychonomic Bulletin & Review*, *24*(4), 1217–1224. https://doi.org/10.3758/s13423-016-1189-0

Undorf, M., & Zimdahl, M. F. (2019). Metamemory and memory for a wide range of font sizes: What is the contribution of perceptual fluency?

Journal of Experimental Psychology: Learning, Memory, and Cognition, *45*(1), 97–109. https://doi.org/10.1037/xlm0000571

Undorf, M., Zimdahl, M. F., & Bernstein, D. M. (2017). Perceptual fluency contributes to effects of stimulus size on judgments of learning. *Journal of Memory and Language*, *92*, 293–304. https://doi.org/10.1016/j.jml.2016.07.003

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

Volz, K. G., & von Cramon, D. Y. (2006). What neuroscience can tell about intuitive processes in the context of perceptual discovery. *Journal of Cognitive Neuroscience*, *18*(12), 2077–2087. https://doi.org/10.1162/jocn.2006.18.12.2077

Wang, Z., Yang, C., Zhao, W., & Jiang, Y. (2020). Perceptual fluency affects judgments of learning non-analytically and analytically through beliefs about how perceptual fluency affects memory. *Frontiers in Psychology*, *11*, 552824. https://doi.org/10.3389/fpsyg.2020.552824

Winkielman, P., Schwarz, N., Fazendeiro, T. A., & Reber, R. (2003). The hedonic marking of processing fluency: Implications for evaluative judgment. In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 189–217). Lawrence Erlbaum Associates Publishers.

Witherby, A. E., & Tauber, S. K. (2017). The concreteness effect on judgments of learning: Evaluating the contributions of fluency and beliefs. *Memory & Cognition*, *45*(4), 639–650. https://doi.org/10.3758/s13421-016-0681-0

Yang, C., Huang, T. S.-T., & Shanks, D. R. (2018). Perceptual fluency affects judgments of learning: The font size effect. *Journal of Memory and Language*, *99*, 99–110. https://doi.org/10.1016/j.jml.2017.11.005

Yang, C., Yu, R., Hu, X., Luo, L., Huang, T. S.-T., & Shanks, D. R. (2021). How to assess the contributions of processing fluency and beliefs to the formation of judgments of learning: methods and pitfalls. *Metacognition and Learning*, *16*(2), 319–343. https://doi.org/10.1007/s11409-020-09254-4

Appendix A

Table A1 shows within-person gamma correlations between JOLs and memory performance in Experiments 1-6. Correlations were calculated across all pictures (overall) and separately for coherent and incoherent pictures and, in Experiment 2, for pictures with high and low contrast.

Table A1

Means (and Standard Deviations) of Gamma Correlations in Experiments 1-6

Experiment	Coherent	Incoherent	Overall
Experiment 1	.19* (.49)	-.02 (.52)	.11 (.45)
Experiment 2			
High Contrast	.25* (.62)	.15 (.63)	.25*** (.44)
Low Contrast	.26* (.69)	.17 (.74)	.21* (.64)
Overall	.25** (.53)	.21* (.59)	.24*** (.41)
Experiment 3	.34*** (.38)	.15* (.41)	.28*** (.29)
Experiment 4	.19* (.51)	.17* (.52)	.22*** (.39)
Experiment 5	.28*** (.43)	.10 (.41)	.24*** (.31)
Experiment 6	.23** (.45)	-.02 (.52)	.16** (.38)

* $p < .05$, ** $p < .01$, *** $p < .001$.

Appendix B

Table B1 shows JOLs separately for participants who reported that they had perceived a gestalt (aware) and those who did not perceive a gestalt (unaware) in Experiments 1-6.

Table B1

Means (and Standard Deviations) of Judgments of Learning (JOLs) for Participants Who Were Aware and Unaware of Differences in Visual Coherence in Experiments 1-6

Experiment	Aware		Unaware	
	Coherent	Incoherent	Coherent	Incoherent
Experiment 1	3.17 (1.29)	2.63 (1.18)	3.34 (1.65)	2.90 (1.62)
Experiment 2				
High Contrast	3.06 (1.44)	2.61 (1.01)	3.45 (1.56)	2.91 (1.42)
Low Contrast	2.38 (1.09)	2.03 (0.95)	2.60 (1.66)	2.62 (1.19)
Experiment 3	4.38 (1.57)	3.26 (1.19)	4.04 (1.31)	3.20 (1.44)
Experiment 4	4.42 (1.66)	3.58 (1.54)	3.97 (1.84)	3.53 (2.13)
Experiment 5	4.35 (1.07)	3.51 (0.98)	2.70 (1.52)	2.09 (1.10)
Experiment 6	4.54 (1.22)	3.38 (0.99)	3.47 (1.38)	2.88 (1.19)

MEMORY
2021, VOL. 29, NO. 5, 559–572
https://doi.org/10.1080/09658211.2021.1919144

Hindsight bias in metamemory: outcome knowledge influences the recollection of judgments of learning

Malte F. Zimdahl and Monika Undorf

Department of Psychology, School of Social Sciences, University of Mannheim, Mannheim, Germany

ABSTRACT
Hindsight bias describes people's tendency to overestimate how accurately they have predicted an event's outcome after obtaining knowledge about it. Outcome knowledge has been shown to influence various forms of judgments, but it is unclear whether outcome knowledge also produces a hindsight bias on Judgments of Learning (JOLs). Three experiments tested whether people overestimated the accuracy of their memory predictions after obtaining knowledge about their actual memory performance. In all experiments, participants studied 60 cue-target word pairs, made a JOL for each word pair, and tried to recall the targets in a cued-recall test. In Experiments 1a and 1b, people recollected their original JOLs after attempting to recall each target, that is, after they obtained outcome knowledge for all items. In Experiments 2 and 3, people recollected their original JOLs in a separate phase after attempting to recall half the targets so that they had outcome knowledge for some but not all items. In all experiments, recollected JOLs were closer to actual memory performance than original JOLs for items with outcome knowledge only. Thus, outcome knowledge produced a hindsight bias on JOLs. Our results demonstrate that people overestimate the accuracy of their memory predictions in hindsight.

ARTICLE HISTORY
Received 14 August 2020
Accepted 14 April 2021

KEYWORDS
Metamemory; metacognition; judgments of learning; hindsight bias; outcome knowledge

When people obtain knowledge about the outcome of an event, their recollected judgments about the outcome's probability are typically closer to the outcome than their original judgments made before the outcome was known. This influence of outcome knowledge has been called *hindsight bias* (Fischhoff, 1975) and is among the most studied judgment biases in the literature (e.g., Roese & Vohs, 2012). The hindsight bias was demonstrated in a variety of domains, including medical diagnoses (Arkes et al., 1988), legal decisions (Giroux et al., 2016), political elections (Blank et al., 2003), financial investments (Biais & Weber, 2009), and sport events (Roese & Maniar, 2016). The hindsight bias exists across cultures (Pohl et al., 2002), across the lifespan (Bayen et al., 2007; Bernstein et al., 2011), and it is evident in visual (Harley et al., 2004), auditory (Bernstein et al., 2012), and gustatory judgments (Pohl et al., 2003b). Thus, the hindsight bias is a very robust cognitive illusion (Christensen-Szalanski & Willham, 1991; Guilbault et al., 2004; Hoffrage & Pohl, 2003).

The current study investigated whether the hindsight bias is also evident in another type of judgment, that is, *judgments of learning* (JOLs). As one of the most popular metacognitive judgments, JOLs assess the subjective probabilities of remembering recently studied items in an upcoming memory test. Numerous prior studies found that JOLs are moderately accurate with regard to resolution or relative accuracy, that is, within-participant gamma correlations between JOLs and actual memory performance (for a review, see Rhodes, 2016). Moreover, there is evidence for systematic dissociations between JOLs and memory performance. For example, people erroneously predict that their chances of remembering words increase with increasing font size (Luna et al., 2018; Mueller et al., 2014; Rhodes & Castel, 2008; Undorf & Zimdahl, 2019; but see Halamish, 2018). Systematic dissociations between JOLs and memory performance are incompatible with the idea that people base their JOLs directly on the strength of memory traces but rather suggest that people infer their JOLs from cues available during study (e.g., Koriat, 1997; Undorf et al., 2018).

First evidence for effects of outcome knowledge effects on JOLs comes from studies showing that people use outcome knowledge as a cue for JOLs in multiple study-test trials on the same material (Finn & Metcalfe, 2008; Serra & Ariel, 2014). Several studies found that participants base their JOLs in a second study phase on their memory performance in the first test phase (memory-for-past-test-heuristic; Ariel & Dunlosky, 2011; Finn & Metcalfe, 2007, 2008). This resulted in improved resolution of JOLs in later study-test cycles compared to the first study-test

CONTACT Malte F. Zimdahl malte.zimdahl@psychologie.uni-mannheim.de Department of Psychology, School of Social Sciences, University of Mannheim, A5, 6, 68159 Mannheim, Germany

cycle. Moreover, Kornell and Rhodes (2013) and Sitzman et al. (2016) found that the type of outcome knowledge impacted the resolution of JOLs. JOL resolution was higher when participants received no explicit feedback at test than when they were provided with the correct answer, presumably because they had difficulty to predict at an item-by-item level which errors would be corrected by feedback. Despite these effects of outcome knowledge on JOLs, no study to date has investigated the influence of outcome knowledge on people's *memory* for their previous JOLs. Thus, it is unclear how people *recollect* their JOLs after their memory is tested, that is, once they have obtained outcome knowledge about their test performance.

Investigating the question how JOLs are recollected with outcome knowledge is especially important for understanding how people use metacognitive judgments to monitor their learning progress (Benjamin et al., 1998; Nelson & Dunlosky, 1991). Moreover, metacognitive judgments influence learning behaviour, such as how learners choose what to study (Metcalfe & Finn, 2008) and how to allocate their study time (Hines et al., 2009; Undorf & Ackerman, 2017). In such situations, accurate metacognitive judgments are essential for successful learning. Consider therefore a student preparing for an upcoming exam: during learning, she monitors her studying by making JOLs. Afterwards, she tests herself and receives outcome knowledge about her memory performance. If she is able to correctly recollect her previous JOLs, testing will also provide her with feedback about the accuracy of her JOLs and, consequently, will allow her to improve her monitoring accuracy and optimise her learning. A hindsight bias on recollected JOLs would result in the student overestimating the accuracy of her original JOLs. This may prevent her from understanding that her monitoring was of limited accuracy, which is a prerequisite for improved monitoring and better learning in the future. At the same time, basing control processes on biased recollected JOLs that have high resolution might make learning more effective (see Ackerman et al., 2020 for a similar conclusion).

So far, little is known about how people recollect JOLs. To our best knowledge, only one experiment (Finn & Metcalfe, 2008, Experiment 5a) has ever investigated memory for JOLs. In that experiment, a first study phase in which participants learned 48 word pairs and made JOLs was followed by a second phase, in which they remembered their previous JOLs. A mean correlation of $G = .59$ between original and recollected JOLs showed that participants remembered their original JOLs quite accurately. Notably, participants were not tested in between, that is, they had no outcome knowledge. Thus, it is an open question whether obtaining outcome knowledge biases the recollection of JOLs and produces a hindsight bias.

If there is a hindsight bias on JOLs, recollected JOLs (RJOLs) made after the outcome is known – in this case whether an item is recalled in the memory test – should be closer to the outcome than original JOLs (OJOLs) made before the outcome is known. Therefore, evidence for a hindsight bias on JOLs should be found in three different measures: first, given that JOLs usually differentiate between recalled and not-recalled items, both types of JOLs should be higher for recalled than for not-recalled items. However, a hindsight bias on JOLs should increase the difference between recalled and not-recalled items in RJOLs as compared to OJOLs, producing an interaction in an ANOVA with the repeated measures factors JOL type (OJOL, RJOL) and recall success (yes, no). Second, a hindsight bias on JOLs should produce higher resolution of RJOLs than of OJOLs, resulting in higher gamma correlations for RJOLs than for OJOLs. Third, a hindsight bias on JOLs should also be evident in ΔHB, a standard measure from the hindsight bias literature (Pohl, 1992). ΔHB is defined as

$$\Delta HB = |OJ - CJ| - |ROJ - CJ| \qquad (1)$$

where OJ denotes the original judgment, CJ the correct judgment, and ROJ the recollected judgment. In our case, OJ equals OJOL, CJ equals the correct JOL (0% for not-recalled items and 100% for recalled items), and ROJ equals RJOL. ΔHB is positive when the recollected judgment is closer to the correct judgment than the original judgment, therefore indicating a hindsight bias. ΔHB is equal to 0 when the recollected judgment is identical to the original judgment, and negative when the recollected judgment is more distant from the correct judgment than the original judgment. In sum, a hindsight bias on JOLs should reveal itself in an interaction between JOL type and recall success, higher resolution of RJOLs than of OJOLs, and a significantly positive ΔHB.

Two of these measures have been used in studies on the hindsight bias with another type of metacognitive judgments, that is, confidence judgments. Although people are quite good at recollecting their original confidence judgments (Sitzman et al., 2014, Experiment 2), there is evidence for a hindsight bias when recollection fails. In one of the earliest studies on the hindsight bias by Fischhoff (1977), participants were provided with two-alternative forced-choice trivia questions and judged the subjective probability that the experimenter-chosen answer was correct. Afterwards, participants learned about the actual correctness of the answers and recollected their original confidence judgments. Results showed an interaction between judgment type and correctness of the answer with higher recollected than original confidence judgments for correct answers and lower recollected than original confidence judgments for incorrect answers. Notably, resolution was not calculated in this study. In a similar study by Ackerman et al. (2020), participants answered general-knowledge questions with fixed-interval responses (e.g., "In what year was Barack Obama first elected president?" Required interval: 5 years) and made confidence judgments for their answers. Afterwards, they were provided with feedback about the correctness of their answers and recollected their original confidence judgments. Results again showed an interaction between judgment type and correctness of the answer. However, recollected confidence judgments were higher than original confidence judgments for correct answers and similar to original confidence judgments for incorrect answers. Furthermore, resolution of recollected confidence judgments was higher than resolution of original confidence judgments. Granhag et al. (2000) examined confidence judgments in the domain of eyewitness memory. Participants answered two-alternative forced-choice questions about a film clip depicting a kidnapping event, made confidence judgments for their answers, and recollected them after they had received feedback. As in the previous studies, Granhag et al. (2000) found an interaction between judgment type and correctness of the answer. However, they neither found higher recollected than original confidence judgments for correct answers nor lower recollected than original confidence judgments for incorrect answers and did not report resolution. Taken together, these studies provide evidence for a hindsight bias on confidence judgments. However, only one prior study (Ackerman et al., 2020) examined the influence of outcome knowledge on the resolution of metacognitive judgments, and no study to date has used ΔHB to quantify the size of the hindsight bias on metacognitive judgments. Moreover, it is an open question whether the same results would emerge with JOLs.

The aim of the present study was to investigate whether outcome knowledge produces a hindsight bias on JOLs, which should be reflected in an interaction between JOL type and recall success, higher resolution of RJOLs than of OJOLs, and a significantly positive ΔHB. As an additional question, we were interested in whether biased recollection of JOLs would be evident in both recalled and not-recalled items. Therefore, we used planned comparisons to test whether RJOLs are higher than OJOLs when recall is successful and lower than OJOLs when recall is unsuccessful. However, mixed prior results on confidence judgments suggest that results on JOLs might not show a clear pattern regarding this question (Ackerman et al., 2020; Fischhoff, 1977; Granhag et al., 2000).

Experiment 1a

In Experiment 1a, participants studied and judged 60 cue-target word pairs and tried to recall the targets in a cued-recall test. Also, participants recollected their JOLs immediately after attempting to recall each target.

Method

Participants

An a priori power analysis using G*Power 3 (Faul et al., 2007) revealed that 54 participants are required to detect a medium effect (Cohen's $f = 0.25$) with $\alpha = .05$ and a statistical power of $1 - \beta = .95$ when conducting a within-subjects ANOVA assuming a correlation of .50 between repeated measures. Participants were 59 University of Mannheim undergraduates. Participants completed the study online at a place and time of their choosing and were compensated with course credit. Because a lack of variability in recall performance produced missing values in several analyses, we excluded participants with zero recall ($n = 2$) or perfect recall ($n = 1$) in this and all following experiments (for a similar approach, see Ackerman et al., 2020). Furthermore, we excluded participants who reported that they had engaged in other tasks during the experiment ($n = 2$) or that they had copied words (none in this experiment). Therefore, the final sample included 54 participants (48 female, 6 male) with a mean age of 20.93 years ($SD = 3.04$). All participants indicated to be native speakers of or fluent in German.

Materials

Stimuli were 58 weakly associated word pairs (e.g., *tent–rain*) from the Noun Associations for German database (Melinger & Weber, 2006) with a mean number of 5.27 letters ($SD = 1.19$) and a mean association of 0.05 ($SD = 0.01$). Two additional word pairs served as primacy buffers and were not included in the analysis.

Procedure

The experiment consisted of a study phase, a filler task, and a cued recall test. Instructions informed participants that they would study 60 word pairs and would be asked to recall the second word of each word pair with the first word given as a retrieval cue in a final memory test. Participants were also told that they should, immediately after studying each word pair, estimate the probability of remembering the target in the test phase. At study, each trial started with a 500-ms fixation cross presented at the top of the screen, followed by a randomly selected word pair displayed for 2 s at the same location. Immediately after the word pair disappeared, the JOL prompt "The chance to recall (0%–100%)?" was presented at the bottom of the screen. Participants indicated their JOL by entering a number between 0 and 100 with the keyboard in a self-paced manner. A 500-ms blank screen preceded the next study trial. After the study phase, participants performed an unrelated filler task for 90 s, which consisted of items from the Wiener Matrizen Test (Formann & Piswanger, 1979). In the cued recall test, all word pairs were shown in a new random order. On each trial, the cue was presented at the top of the screen and participants typed in the target. Immediately afterwards, the correct target was added beneath the participant's response to provide feedback about whether they correctly recalled the target. After 1 s, the JOL prompt "What chance of recall (0%–100%) did you assign during study?" was added to the screen and participants indicated their RJOL self-paced. Afterwards, participants completed a post-experimental questionnaire in which they indicated whether they had engaged in other tasks during the experiment or had copied words. At the end of the study, participants were debriefed and thanked for their participation.

Results and discussion

Figure 1 (left panel) shows mean OJOLs and RJOLs for recalled and not-recalled items. As expected, OJOLs were higher for recalled than for not-recalled items, $t(53) = 5.90$, $p < .001$, $d = 0.58$.

To test whether there was a hindsight bias on JOLs, we first conducted a 2 (JOL type: OJOL, RJOL) × 2 (recall success: yes, no) within-subjects ANOVA. A significant main effect of recall success indicated higher JOLs for recalled than for not-recalled items, $F(1, 53) = 60.42$, $MSE = 99.51$, $p < .001$, $\eta_p^2 = .53$. The main effect of JOL type was not significant, $F < 1$. Most importantly, a significant interaction revealed that the JOL difference between recalled and not-recalled items was larger for RJOLs than for OJOLs, suggesting a hindsight bias on JOLs, $F(1, 53) = 12.16$, $MSE = 21.79$, $p < .001$, $\eta_p^2 = .19$.

As can be seen from Table 1, resolution was positive for OJOLs and RJOLs. More importantly, resolution of RJOLs was higher than of OJOLs, again suggesting a hindsight bias on JOLs, $t(53) = 4.95$, $p < .001$, $d = 0.63$. A significantly positive ΔHB further supported the conclusion that there was a hindsight bias on JOLs, $t(53) = 3.21$, $p = .002$, $d = 0.44$.[1]

We conducted planned comparisons to examine whether a biased recollection of JOLs was evident in both recalled and not-recalled items. For recalled items, RJOLs were numerically but not significantly higher than OJOLs, $t(53) = 1.67$, $p = .102$, $d = 0.11$. For not-recalled items, RJOLs were lower than OJOLs, $t(53) = 2.35$, $p = .023$, $d = 0.20$, suggesting that the biased recollection of JOLs was stronger for not-recalled than for recalled items.

In summary, Experiment 1a showed clear evidence for a hindsight bias on JOLs. RJOLs were biased towards actual memory performance, resulting in a significant interaction between JOL type and recall success, higher resolution of RJOLs than of OJOLs, and a significantly positive ΔHB.

Experiment 1b

The purpose of Experiment 1b was to replicate Experiment 1a with a non-student sample. We expected the same results as in Experiment 1a.

Method

Participants, materials, and procedure

As in Experiment 1a, we aimed at a sample size of 54 participants to obtain a statistical power of $1 - \beta = .95$ in a repeated-measures ANOVA with $\alpha = .05$, Cohen's $f = 0.25$, and a correlation of .50 between repeated measures.[2] Participants were 53 Amazon's Mechanical Turk workers. We included only workers with a HIT Approval Rate greater than 95% who were located in Germany. Participants completed the study online at a place and time of their choosing and were compensated with a payment of $3.00. The data of four participants were excluded after applying the same exclusion criteria used in Experiment 1a (all of them because they reported that they had engaged in other tasks during the experiment), resulting in a sample size of 49 participants (7 female, 41 male, 1 diverse) with a mean age of 27.84 years ($SD = 9.24$). All participants were native speakers of or fluent in German. Stimuli and procedure were the same as in Experiment 1a.

Results and discussion

Figure 1 (right panel) shows mean OJOLs and RJOLs for recalled and not-recalled items. As in Experiment 1a,

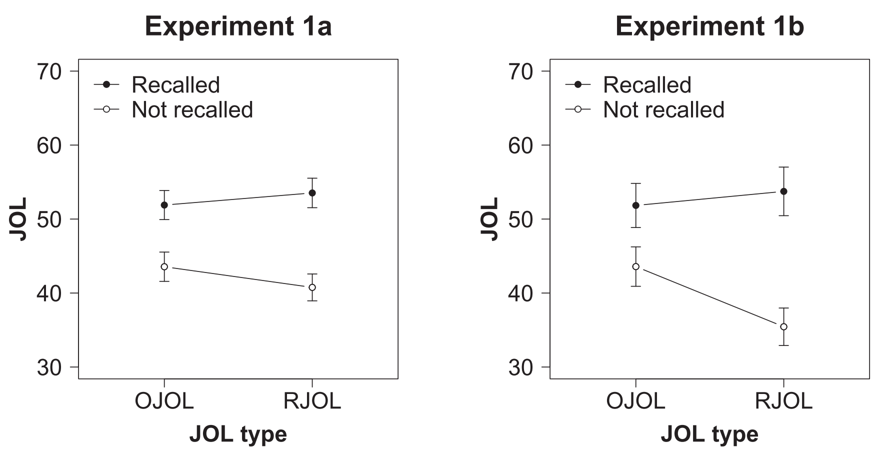


Figure 1. The figure shows mean original judgments of learning (OJOL) and recollected judgments of learning (RJOL) in Experiment 1a (left panel) and Experiment 1b (right panel). Filled symbols represent recalled items and open symbols represent not-recalled items. Error bars represent one standard error of the mean.

OJOLs were higher for recalled than for not-recalled items, $t(48) = 5.27$, $p < .001$, $d = 0.41$.

A 2 (JOL type: OJOL, RJOL) × 2 (recall success: yes, no) within-subjects ANOVA revealed a significant main effect of recall success, indicating higher JOLs for recalled than for not-recalled items, $F(1, 48) = 68.24$, $MSE = 126.67$, $p < .001$, $\eta_p^2 = .59$. The main effect of JOL type was also significant, indicating that OJOLs were higher than RJOLs, $F(1, 48) = 8.31$, $MSE = 57.25$, $p = .006$, $\eta_p^2 = .15$. A significant interaction between JOL type and recall success showed that the JOL difference between recalled and not-recalled items was larger for RJOLs than for OJOLs, suggesting a hindsight bias on JOLs, $F(1, 48) = 34.72$, $MSE = 35.50$, $p < .001$, $\eta_p^2 = .42$.

Gamma correlations (Table 1) showed that resolution of RJOLs was again higher than of OJOLs, $t(48) = 6.72$, $p < .001$, $d = 0.93$. As in the previous experiment, a positive ΔHB revealed a hindsight bias on JOLs, $t(48) = 5.23$, $p < .001$, $d = 0.75$.

Planned comparisons revealed that for recalled items, RJOLs were numerically but not significantly higher than OJOLs, $t(48) = 1.40$, $p = .168$, $d = 0.09$. For not-recalled items, RJOLs were lower than OJOLs, $t(48) = 5.83$, $p < .001$, $d = 0.45$, again suggesting that the biased recollection of JOLs was stronger for not-recalled than for recalled items.

In summary, Experiment 1b closely replicated Experiment 1a in showing clear evidence for a hindsight bias on JOLs. As in the previous experiment, RJOLs were biased towards actual memory performance, resulting in a significant interaction between JOL type and recall success, higher resolution of RJOLs than of OJOLs, and a significantly positive ΔHB.

Experiment 2

In Experiment 1, outcome knowledge produced a hindsight bias on JOLs. However, participants obtained outcome knowledge for all word pairs because they recollected their JOLs after attempting to recall the respective target. Therefore, it is possible that similar results would emerge when people recollect JOLs without obtaining outcome knowledge. We tested this possibility in Experiment 2 where participants recollected their JOLs in a separate JOL recollection phase after the first half of the cued recall test. This provided participants with outcome knowledge for some but not all items. If outcome knowledge underlies the hindsight bias on JOLs, items for which participants have outcome knowledge should reveal a hindsight bias, characterised by a significant interaction between JOL type and recall success, higher resolution of RJOLs than of OJOLs, and a significantly positive ΔHB.[3] In contrast, items for which participants do not have outcome knowledge should reveal no hindsight bias, that is, no interaction between JOL type and recall success, similar resolution of RJOLs and of OJOLs, and an insignificant ΔHB. Furthermore, we also expected memory performance to be higher in the second half of the cued recall test because the word pairs were presented for a second time in the JOL recollection phase.

Table 1. Means (and Standard Deviations) of Recall, Gamma Correlations, and Pohl's (1992) Difference Measure of Hindsight Bias (ΔHB) for Experiments 1 and 2.

	Exp. 1a	Exp. 1b	Exp. 2	
			With outcome knowledge	Without outcome knowledge
Recall	.69 (.19)	.56 (.23)	.58 (.22)	.70 (.23)
G(OJOL, Recall)	.21 (.25)	.19 (.35)	.26 (.35)	.30 (.41)
G(RJOL, Recall)	.38 (.28)	.51 (.34)	.49 (.34)	.31 (.42)
ΔHB	2.03**	4.30***	3.56***	−0.72

Note: OJOL = original judgment of learning, RJOL = recollected judgment of learning. ** $p < .01$, *** $p < .001$.

Method

Participants and materials

We set a sample size about twice as large as in the previous experiment because we manipulated an additional within-subjects factor. Participants were 105 Amazon's Mechanical Turk workers. As before, we included only workers with a HIT Approval Rate greater than 95% who were located in Germany. Participants completed the study online at a place and time of their choosing and were compensated with a payment of $3.00. As in Experiment 1, we excluded participants with perfect recall ($n = 4$), participants who had engaged in other tasks ($n = 2$), and participants who had copied words ($n = 3$), resulting in a sample size of 96 participants (22 female, 74 male) with a mean age of 28.34 ($SD = 9.23$) years.[4] Most participants indicated to be native speakers of or fluent in German (93.75%). Materials were the same as in Experiment 1.

Procedure

The procedure was similar to that of Experiment 1 with the exception that JOLs were recollected in a separate phase in the middle of the cued recall test, that is, after participants had attempted to recall 30 randomly selected targets. In the JOL recollection phase, all 60 word pairs were presented in a new random order, and participants indicated their RJOLs in a self-paced manner. Immediately after this phase, they completed the second half of the cued recall test in which they attempted to recall the remaining 30 targets.

Results and discussion

As expected, memory performance (Table 1) was lower in the first than in the second part of the cued recall test, $t(95) = 8.35$, $p < .001$, $d = 0.51$. Replicating Experiment 1, OJOLs were higher for recalled than for not-recalled items. This was true for items with outcome knowledge,

$t(95) = 6.75$, $p < .001$, $d = 0.55$, and without outcome knowledge, $t(95) = 6.80$, $p < .001$, $d = 0.61$.

Figure 2 shows mean OJOLs and RJOLs for recalled and not-recalled items with outcome knowledge (left panel) and without outcome knowledge (right panel).

To test whether there was a hindsight bias on JOLs, we conducted a 2 (JOL type: OJOL, RJOL) × 2 (recall success: yes, no) × 2 (outcome knowledge: yes, no) within-subjects ANOVA. As expected, a significant main effect of recall success indicated higher JOLs for recalled than for not-recalled items, $F(1, 95) = 95.69$, $MSE = 356.47$, $p < .001$, $\eta_p^2 = .50$. Significant interactions between JOL type and recall success and among JOL type, recall success, and outcome knowledge suggested a hindsight bias on JOLs that was limited to items with outcome knowledge, $F(1, 95) = 19.15$, $MSE = 38.17$, $p < .001$, $\eta_p^2 = .17$, and $F(1, 95) = 21.18$, $MSE = 36.12$, $p < .001$, $\eta_p^2 = .18$. No other effects were significant, all $Fs < 3.52$, all $ps > .064$.

To test more directly if there was a hindsight bias on JOLs for items with outcome knowledge but not for items without outcome knowledge, we conducted two separate 2 (JOL type: OJOL, RJOL) × 2 (recall success: yes, no) within-subjects ANOVAs for items with and without outcome knowledge. For items with outcome knowledge, a significant main effect of recall success indicated higher JOLs for recalled than for not-recalled items, $F(1, 95) = 91.16$, $MSE = 234.21$, $p < .001$, $\eta_p^2 = .49$. More importantly, the interaction between JOL type and recall success was significant, suggesting a hindsight bias on JOLs for items with outcome knowledge, $F(1, 95) = 44.39$, $MSE = 33.68$, $p < .001$, $\eta_p^2 = .32$. The main effect of JOL type was not significant, $F < 1$. For items without outcome knowledge, a main effect of recall success indicated higher JOLs for recalled than for not-recalled items, $F(1, 95) = 51.07$, $MSE = 259.25$, $p < .001$, $\eta_p^2 = .35$. A significant main effect of JOL type revealed that RJOLs were lower than OJOLs, $F(1, 95) = 5.18$, $MSE = 68.45$, $p = .025$, $\eta_p^2 = .05$. Because this unexpected finding replicated in Experiment 3, we will return to it in the General Discussion. Most importantly, the interaction between JOL type and recall success was not significant, $F < 1$, suggesting that there was no hindsight bias on JOLs for items without outcome knowledge.

As can be seen in Table 1, gamma correlations revealed that, for items with outcome knowledge, resolution of RJOLs was higher than of OJOLs, $t(95) = 7.49$, $p < .001$, $d = 0.68$. In contrast, for items without outcome knowledge, resolution of RJOLs and of OJOLs was similar, $t(95) = 0.31$, $p = .760$, $d = 0.03$. These results again suggested that the hindsight bias was limited to items with outcome knowledge. Furthermore, ΔHB was significantly positive for items with outcome knowledge, $t(95) = 5.63$, $p < .001$, $d = 0.58$, but not for items without outcome knowledge, $t(95) = 1.37$, $p = .175$, $d = 0.14$. This further supported the interpretation that there was a hindsight bias on JOLs for items with outcome knowledge only.

Planned comparisons revealed that for items with outcome knowledge, RJOLs were higher than OJOLs for recalled items, $t(95) = 4.22$, $p < .001$, $d = 0.18$, and lower than OJOLs for not-recalled items, $t(95) = 4.43$, $p < .001$, $d = 0.22$. In contrast, for items without outcome knowledge, RJOLs were lower than OJOLs for recalled items, $t(95) = 2.61$, $p = .011$, $d = 0.11$, and similar to OJOLs for not-recalled items, $t(95) = 1.44$, $p = .154$, $d = 0.09$.

In summary, results from Experiment 2 underline the crucial role of outcome knowledge for a hindsight bias on JOLs. RJOLs were biased towards actual memory performance for items with outcome knowledge only, resulting in a significant interaction between JOL type and recall success, higher resolution of RJOLs than of OJOLs, and a significantly positive ΔHB. In contrast, for items without outcome knowledge, there was no interaction between JOL type and recall success, similar resolution of RJOLs and OJOLs, and an insignificant ΔHB. Overall, these results converge on the conclusion that the hindsight bias on JOLs occurred only when participants had outcome knowledge.

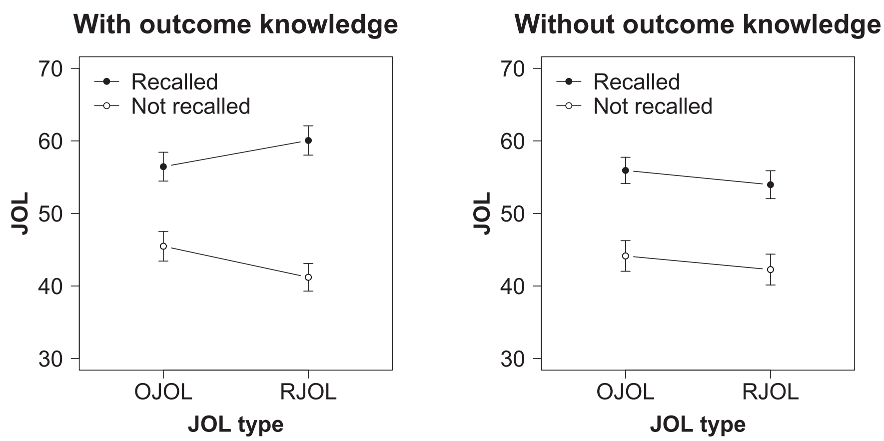


Figure 2. The figure shows mean original judgments of learning (OJOL) and recollected judgments of learning (RJOL) for items with outcome knowledge (left panel) and without outcome knowledge (right panel) in Experiment 2. Filled symbols represent recalled items and open symbols represent not-recalled items. Error bars represent one standard error of the mean.

Experiment 3

The first two experiments reported here showed that RJOLs were biased towards actual memory performance when the outcome was known. However, in both experiments, mean OJOLs were located near the middle of the JOL scale (around 50%) for recalled and not-recalled items. Presumably, this maximised the chance of obtaining a hindsight bias on JOLs. Provided that OJOLs differentiate between items that will later be recalled and not-recalled, OJOLs around 50% allow RJOLs to be biased both upwards and downwards by up to 50%. In contrast, the maximum upward bias for later recalled items is reduced when OJOLs are located considerably above 50% and the maximum downward bias for later not-recalled items is reduced when OJOLs are located considerably below 50%. Manipulating the level of OJOLs therefore allows to test the robustness of the hindsight bias on JOLs across different JOL levels.

In Experiment 3, we used an anchoring procedure to manipulate the level of OJOLs. The anchoring procedure was similar to this used by England and Serra (2012) and Yang et al. (2018): Prior to the study phase, participants were told that most people find the task to be extremely easy and recall about 80% of the items (high-anchor group) or, alternatively, that most people find the task to be extremely difficult and recall about 20% of the items (low-anchor group). Importantly, materials and tasks were identical for the two groups.

We expected that the anchor manipulation would result in higher OJOLs in the high-anchor group and lower OJOLs in the low-anchor group. For items with outcome knowledge, we expected to see a hindsight bias in both anchor groups, characterised by a significant interaction between JOL type and recall success, higher resolution of RJOLs than of OJOLs, and a significantly positive ΔHB. For items without outcome knowledge, in contrast, we expected that neither anchor group would show a hindsight bias, that is, no interaction between JOL type and recall success, similar resolution of RJOLs and of OJOLs, and an insignificant ΔHB.

Method

Participants

We set a sample size about twice as large as in the previous experiment because we manipulated an additional between-subjects factor. Participants were 258 Amazon's Mechanical Turk workers. We included only workers with a HIT Approval Rate greater than 95%. Because only a small proportion of MTurk workers is located in Germany (0.27%; Difallah et al., 2018) and because we excluded all workers that had participated in Experiments 1b and 2, we used workers located in the United States. Participants completed the study online at a place and time of their choosing and were compensated with a payment of \$2.00. As in the previous experiments, we excluded participants with zero recall ($n = 15$), perfect recall ($n = 6$), participants who had engaged in other tasks ($n = 17$), and participants who had copied words ($n = 8$), resulting in a sample size of 212 participants (89 female, 123 male) with a mean age of 37.29 years ($SD = 11.53$). Most participants indicated to be native speakers of or fluent in English (99.53%). The participants were randomly assigned to either the high-anchor group ($n = 104$) or to the low-anchor group ($n = 108$).

Materials

Stimuli were 60 weakly associated word pairs (e.g., *hero – saint*) from the University of South Florida free association, rhyme, and word fragment norms (Nelson et al., 2004) with a mean number of 6.55 letters ($SD = 2.10$) and a mean association of 0.02 ($SD = 0.01$).

Procedure

The procedure was identical to that of Experiment 2 with two exceptions. First, before the study phase, participants in the high-anchor group read "Most participants find the present task to be extremely easy and recall about 80% of the items", whereas participants in the low-anchor group read "Most participants find the present task to be extremely difficult and recall about 20% of the items". Second, all instructions were in English.

Results and discussion

As in the previous experiment, memory performance (Table 2) was lower in the first than in the second part of the cued recall test, $t(211) = 5.61$, $p < .001$, $d = 0.23$. Somewhat surprisingly, memory performance was higher in the high-anchor than in the low-anchor group, $t(210) = 2.28$, $p = .024$, $d = 0.31$. Replicating Experiments 1 and 2, OJOLs were higher for recalled than for not-recalled items. This was true in both anchor groups for items with outcome knowledge, high-anchor group: $t(103) = 6.99$, $p < .001$, $d = 0.31$; low-anchor group: $t(107) = 7.74$, $p < .001$, $d = 0.36$, and without outcome knowledge, high-anchor group: $t(103) = 4.55$, $p < .001$, $d = 0.22$; low-anchor group: $t(107) = 8.04$, $p < .001$, $d = 0.36$.

Figure 3 shows mean OJOLs and RJOLs for recalled and not-recalled items in the high-anchor group (top panels)

Table 2. Means (and Standard Deviations) of Recall, Gamma Correlations, and Pohl's (1992) Difference Measure of Hindsight Bias (ΔHB) for Experiment 3.

	High-anchor group		Low-anchor group	
	With outcome knowledge	Without outcome knowledge	With outcome knowledge	Without outcome knowledge
Recall	.49 (.28)	.55 (.28)	.41 (.23)	.47 (.24)
G(OJOL, Recall)	.26 (.39)	.16 (.46)	.27 (.39)	.26 (.34)
G(RJOL, Recall)	.40 (.46)	.24 (.42)	.38 (.48)	.25 (.40)
ΔHB	2.07*	1.05	2.28***	0.52

Note: OJOL = original judgment of learning, RJOL = recollected judgment of learning. * $p < .05$, *** $p < .001$.

and the low-anchor group (bottom panels). Left panels depict items with outcome knowledge; right panels depict items without outcome knowledge.

To test whether there was a hindsight bias on JOLs, we conducted a 2 (JOL type: OJOL, RJOL) × 2 (recall success: yes, no) × 2 (outcome knowledge: yes, no) × 2 (anchor group: high, low) mixed ANOVA with JOL type, recall success, and outcome knowledge as within-subjects factors and anchor group as a between-subjects factor. As expected, a significant main effect of anchor group indicated higher JOLs in the high-anchor than in the low-anchor group, $F(1, 210) = 27.87$, $MSE = 4{,}001.11$, $p < .001$, $\eta_p^2 = .12$. Replicating the previous experiments, a significant main effect of recall success indicated higher JOLs for recalled than for not-recalled items, $F(1, 210) = 196.07$, $MSE = 144.73$, $p < .001$, $\eta_p^2 = .48$. Furthermore, a significant main effect of outcome knowledge indicated higher JOLs for items with than without outcome knowledge, $F(1, 210) = 22.37$, $MSE = 83.21$, $p < .001$, $\eta_p^2 = .10$. Significant interactions between JOL type and recall success and among JOL type, recall success, and outcome knowledge suggested a hindsight bias on JOLs that was limited to items with outcome knowledge, $F(1, 210) = 10.17$, $MSE = 47.25$, $p = .002$, $\eta_p^2 = .05$, and $F(1, 210) = 11.38$, $MSE = 39.55$, $p < .001$, $\eta_p^2 = .05$. Moreover, the interactions between JOL type and outcome knowledge and between recall success and outcome knowledge were significant, $F(1, 210) = 34.38$, $MSE = 39.19$, $p < .001$, $\eta_p^2 = .14$, and $F(1, 210) = 13.24$, $MSE = 71.83$, $p < .001$, $\eta_p^2 = .06$. No other effects were significant, all Fs < 3.24, all ps $> .074$.

To test more directly if the hindsight bias on JOLs was limited to items with outcome knowledge, we conducted two separate 2 (JOL type: OJOL, RJOL) × 2 (recall success: yes, no) × 2 (anchor group: high, low) mixed ANOVAs for items with and without outcome knowledge. As expected, for items with outcome knowledge, a significant main effect of anchor group indicated higher JOLs in the high-anchor than in the low-anchor group, $F(1, 210) = 26.47$, $MSE = 1{,}997.64$, $p < .001$, $\eta_p^2 = .11$. Again, a significant main effect of recall success indicated higher JOLs for recalled than for not-recalled items, $F(1, 210) = 159.58$, $MSE = 124.45$, $p < .001$, $\eta_p^2 = .43$. Most importantly, the interaction between JOL type and recall success was significant, indicating a hindsight bias on JOLs, $F(1, 210) = 15.71$, $MSE = 59.21$, $p < .001$, $\eta_p^2 = .07$. No other effects were significant, all Fs < 1.

For items without outcome knowledge, a significant main effect of anchor group indicated higher JOLs in the high-anchor than in the low-anchor group, $F(1, 210) = 28.13$, $MSE = 2{,}086.67$, $p < .001$, $\eta_p^2 = .12$. Again, a significant main effect of recall success indicated higher JOLs for recalled than for not-recalled items, $F(1, 210) = 102.80$, $MSE = 92.11$, $p < .001$, $\eta_p^2 = .33$. A significant main effect of JOL type revealed that RJOLs were lower than OJOLs, $F(1, 210) = 15.12$, $MSE = 123.39$, $p < .001$, $\eta_p^2 = .07$ (see General Discussion). Most importantly, the interaction between JOL type and recall success was not significant, indicating that there was no hindsight bias on JOLs, $F < 1$.[5] No other effects were significant, all Fs < 2.01, all ps $> .158$.

For the sake of completeness, we report the results from four separate 2 (JOL type: OJOL, RJOL) × 2 (recall success: yes, no) within-subjects ANOVAs for items with and without outcome knowledge in the high- and low-anchor group in the Appendix.

As can be seen in Table 2, for items with outcome knowledge, resolution of RJOLs was higher than that of OJOLs in both anchor groups, high-anchor group: $t(103) = 2.85$, $p = .005$, $d = 0.31$; low-anchor group: $t(107) = 2.12$, $p = .037$, $d = 0.25$. In contrast, for items without outcome knowledge, resolution of RJOLs and of OJOLs was similar in both anchor groups, high-anchor group: $t(103) = 1.71$, $p = .090$, $d = 0.19$; low-anchor group: $t(107) = 0.16$, $p = .877$, $d = 0.02$.

Furthermore, ΔHB supported the interpretation that there was a hindsight bias for items with outcome knowledge in both anchor groups, high-anchor group: $t(103) = 2.20$, $p = .030$, $d = 0.22$; low-anchor group: $t(107) = 3.59$, $p < .001$, $d = 0.35$. In contrast, ΔHB was not significantly different from 0 for items without outcome knowledge in both anchor groups, high-anchor group: $t(103) = 1.47$, $p = .145$, $d = 0.14$; low-anchor group: $t(107) = 1.12$, $p = .266$, $d = 0.11$.

Planned comparisons showed that for items with outcome knowledge, RJOLs were higher than OJOLs for recalled items in the low-anchor group, $t(107) = 2.76$, $p = .007$, $d = 0.14$, but not in the high-anchor group, $t(103) = 1.59$, $p = .116$, $d = 0.09$. RJOLs were similar to OJOLs for not-recalled items in both anchor groups, high-anchor group: $t(103) = 1.36$, $p = .178$, $d = 0.07$; low-anchor group: $t(107) = 1.11$, $p = .268$, $d = 0.05$. For items without outcome knowledge, RJOLs were lower than OJOLs for recalled items in both anchor groups, high-anchor group: $t(103) = 2.73$, $p = .008$, $d = 0.14$; low-anchor group: $t(107) = 2.18$, $p = .031$, $d = 0.10$. RJOLs were lower than OJOLs for not-recalled items in the high-anchor group, $t(103) = 3.21$, $p = .002$, $d = 0.17$, but not in the low-anchor group, $t(107) = 1.58$, $p = .116$, $d = 0.07$.

In summary, Experiment 3 replicated the hindsight bias on JOLs for items with outcome knowledge across different levels of JOLs. Both in the high- and in the low-anchor group, RJOLs were biased towards actual memory performance for items with outcome knowledge only, resulting in a significant interaction between JOL type and recall success, higher resolution of RJOLs than of OJOLs, and a significantly positive ΔHB. In contrast, there was no indication for a hindsight bias on JOLs for items without outcome knowledge. These findings emphasise the robustness of the hindsight bias on JOLs by showing that it occurs with JOLs clustered across different points of the JOL scale. Notably, results were also similar between participants from the U.S. (this experiment) and from Germany (Experiments 1 and 2), demonstrating that the hindsight bias on JOLs is robust across different materials and populations.

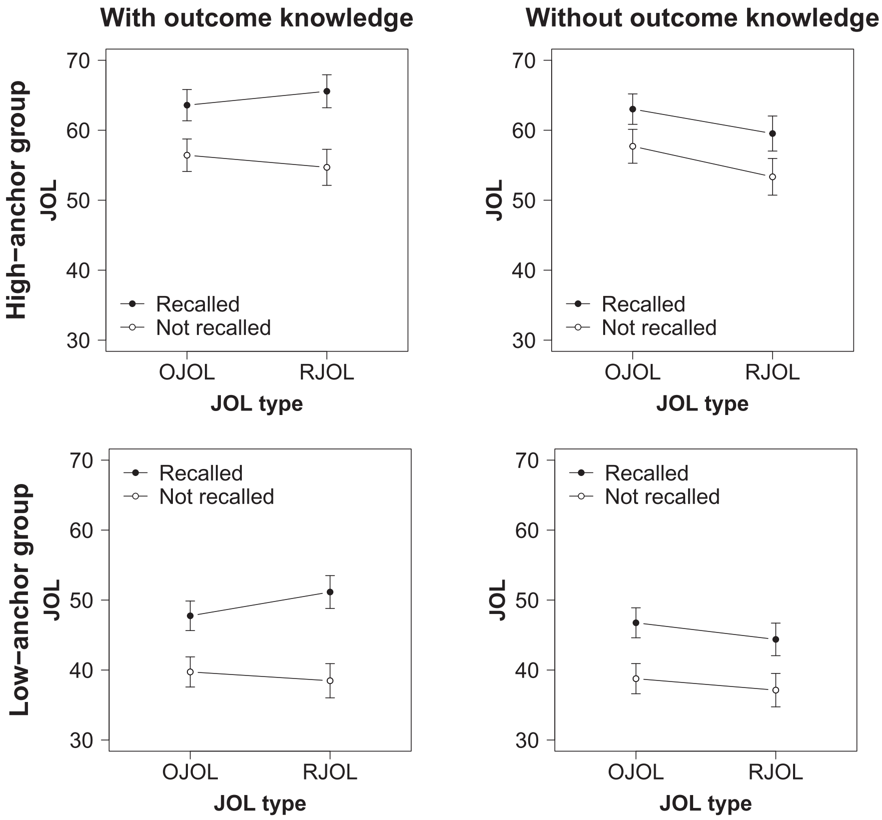


Figure 3. The figure shows mean original judgments of learning (OJOL) and recollected judgments of learning (RJOL) for items with outcome knowledge (left panels) and without outcome knowledge (right panels) in the high-anchor group (top panels) and the low-anchor group (bottom panels) in Experiment 3. Filled symbols represent recalled items and open symbols represent not-recalled items. Error bars represent one standard error of the mean.

Somewhat surprisingly, we found better memory performance in the high- than in the low-anchor group, whereas England and Serra (2012) and Yang et al. (2018) found no effect of anchor level on memory performance. This difference in memory performance might be simply spurious given the small effect size and the high statistical power of the current experiment. Alternatively, the expectation of recalling few items lowered motivation of participants in the low-anchor group. However, this explanation is mere speculation and the found difference in memory performance requires replication. More importantly for present purposes, the observed difference in memory performance shows that the hindsight bias occurs with different levels of actual memory performance.

General Discussion

This study investigated whether one of the most investigated judgment biases, the hindsight bias (for a review, see Bernstein et al., 2016), also afflicts JOLs. Three experiments clearly demonstrated a hindsight bias on JOLs with recollected JOLs being biased towards the outcome of whether or not items were remembered at test when compared to original JOLs made before the outcome was known. Furthermore, outcome knowledge increased the resolution of RJOLs as compared to OJOLs. A standard measure from the hindsight bias literature (ΔHB; Pohl, 1992) also confirmed that there was a hindsight bias on JOLs. Crucially, the hindsight bias occurred only when participants had outcome knowledge, that is, only for items that were tested before people recollected their JOLs. Moreover, the hindsight bias was demonstrated with JOLs clustered across different points of the JOL scale.

Hindsight bias for various judgments

The current study shows that JOLs may be added to the long list of judgments that are afflicted by hindsight bias. Prior studies found a hindsight bias when participants judged the truth of given statements and rated the confidence in their judgments (Campbell & Tesser, 1983; Fischhoff, 1977), responded to general knowledge questions (Erdfelder & Buchner, 1998; Hell et al., 1988), or estimated the probability of possible event outcomes (Fischhoff & Beyth, 1975). Hindsight bias thus occurred with various materials, although with some differences in its magnitude (Christensen-Szalanski & Willham, 1991; Guilbault et al., 2004).

There is a crucial difference between JOLs and other judgments used to study the hindsight bias: JOLs are predictions about one's own memories, whereas other kinds of judgments used to study the hindsight bias are estimates of given, non-changeable facts that are set outside the participant's influence. In these contexts, the correct answer is established before original judgments are made. In contrast, JOLs are judgments about a learning process whose outcome is unknown until participants complete the memory test. Despite this fundamental difference, there are similarities between JOLs and other judgments used to study the hindsight bias. First, either type of judgment is made under uncertainty, which is a necessary precondition to observe the hindsight bias (Pohl & Erdfelder, 2017). Furthermore, there is consensus that both type of judgments are based on probabilistic cues (e.g., Bröder & Undorf, 2019; Koriat, 1997, 2015; Pohl et al., 2003a). Overall, the current research demonstrates that JOLs are afflicted by hindsight bias as are other kinds of judgments.

The hindsight bias on JOLs reported in this study resembles the hindsight bias on another type of metacognitive judgments, that is, confidence judgments. In the present experiments, we found an interaction between JOL type and recall success for items with outcome knowledge, which replicates previously demonstrated interactions between confidence judgment type and correctness of the answer (Ackerman et al., 2020; Fischhoff, 1977; Granhag et al., 2000). Moreover, resolution of recollected JOLs as measured by gamma correlations was higher than of original JOLs for items with outcome knowledge, resembling similar results with confidence judgments (Ackerman et al., 2020). Above and beyond, the current work also showed that a hindsight bias on metacognitive judgments was reflected in ΔHB (Pohl, 1992), a standard measure of hindsight bias magnitude.

The hindsight bias on JOLs reported here is also consistent with findings related to the *forgetting bias* (Castel et al., 2012; Rhodes et al., 2017; Witherby et al., 2019). In studies on the forgetting bias, participants studied items with varying point values that denoted their importance. After trying to recall the items, participants were asked to recollect their point values. Results showed that participants assigned lower values to items they had forgotten than to items they had remembered. Rhodes et al. (2017) showed that the forgetting bias is most likely be due to participants' beliefs that their memory is adaptive and that it stores important information but forgets unimportant information (see also, e.g., Schooler & Hertwig, 2005). In view of evidence that people give higher JOLs to important information (e.g., Soderstrom & McCabe, 2011; Undorf & Ackerman, 2017), it is possible that lower recollected than original JOLs for not-recalled items in the current study were related to people's belief that forgotten items are less important and had therefore received lower original JOLs. In any case, the experiments reported here take it a step further by showing that people also recollected higher JOLs for recalled items. Thus, the observed hindsight bias on JOLs cannot be explained by the forgetting bias alone.

Potential limitations

In Experiments 2 and 3, we found an unexpected outcome: For items without outcome knowledge, RJOLs were lower than OJOLs rather than similar. At first glance, this finding seems problematic, because a decline from OJOLs to RJOLs might contribute to lower RJOLs than OJOLs for not-recalled items with outcome knowledge, thereby explaining parts of the hindsight bias on JOLs. However, the hindsight bias as documented in the current study is characterised by an interaction between JOL type and recall success that cannot be explained by a general decline from OJOLs to RJOLs. What is more, several findings from the current experiments argue against a general decline from OJOLs to RJOLs. In Experiments 1a and 1b, OJOLs and RJOLs were similar for recalled items. In Experiment 2, OJOLs were lower than RJOLs for recalled items with outcome knowledge. Finally, in Experiment 3, RJOLs were similar to or higher than OJOLs for items with outcome knowledge. Consequently, lower RJOLs than OJOLs for items without outcome knowledge in Experiments 2 and 3 do not undermine the finding of a hindsight bias on JOLs. Nevertheless, exploring the reasons for lower RJOLs than OJOLs might be an interesting avenue for future research.

A potential concern is that, in the JOL recollection phase, participants may have made new judgments instead of recollecting their previous judgments. However, we consider this unlikely because participants were explicitly instructed to recollect their OJOLs rather than to make new JOLs. Moreover, the JOL prompt clearly stated that participants should recollect their OJOL. Apart from that, participants in each experiment correctly recollected a substantial number of OJOLs, Experiment 1a: 26.20%, Experiment 1b: 25.80%, Experiment 2: 29.22%, Experiment 3: 28.15%. This finding suggests that, when instructed to make RJOLs, participants indeed attempted to recollect their OJOLs.

Another potential concern might be related to the online samples recruited via Amazon's Mechanical Turk used in Experiments 1b, 2, and 3. In recent years, there have been reports on bots and fraudulent responses that may challenge the quality of data obtained with this recruitment method (e.g., Chmielewski & Kucker, 2020) and suggestions how to achieve high data quality (Aust et al., 2013; Buchanan & Scofield, 2018). In the current study, we ensured high data quality by including only workers with a HIT Approval Rate greater than 95%, which some researchers see as "a sufficient condition for obtaining high-quality data" (Peer et al., 2014, p. 1030). Furthermore, we excluded participants who reported in the post-experimental questionnaire that they had engaged in other tasks or had copied words (see Procedure of Experiment 1a). Importantly, it was clearly stated that participants would receive their reward regardless of their answers to these questions. Finally, very similar results across Experiment 1a (undergraduates) and

Experiment 1b (MTurk workers) also argue in favour of high data quality. Taken together, we believe that the results we obtained in this study from online samples recruited via Amazon's Mechanical Turk are as valid as results from student samples (for a similar conclusion, see Goodman et al., 2013).

Implications and future directions

The current study has theoretical implications for research on the memory-for-past-test heuristic. The memory-for-past-test heuristic assumes that people who study and remember items across multiple study-test cycles base their JOLs in later study phases on their memory performance in the immediately preceding test phase (Ariel & Dunlosky, 2011; Finn & Metcalfe, 2007, 2008). The current study expands this logic by suggesting that past test performance impacts not only new JOLs but also the recollection of previous JOLs. However, future research will be needed to investigate the impact of biased recollection of prior JOLs on new JOLs made in subsequent study phases.

To illustrate the implications of the current findings for learning, we return to the example of a student who is preparing for an upcoming exam. During learning, she monitors her studying by making JOLs and afterwards tests herself on the materials. Our results indicate that when she receives feedback (outcome knowledge) about her memory performance at test, she will overestimate the accuracy of her JOLs. Consequently, she might think that she "knew-it-all-along" (Hawkins & Hastie, 1990), preventing her from identifying causes why her previous JOLs were not accurate. In this case, the hindsight bias on JOLs would hinder future learning. Interestingly, however, the hindsight bias on JOLs might also foster rather than impair learning (as proposed by Ackerman et al., 2020). In particular, when basing future JOLs on the recollection of JOLs, unduly low recollected JOLs for material that was not recalled at test might result in the student allocating more attention or study time to this material. Similarly, unduly high recollected JOLs for material that was recalled at test might help the student to focus on not yet learned materials. However, more research will be needed to understand the role of the hindsight bias on JOLs in learning.

It may also be worthwhile to test the generalizability of the hindsight bias on JOLs when modifying the procedure used here in two theoretically important ways. First, it might be informative to change the experimental design. The influence of outcome knowledge can be assessed with two designs (Pohl & Erdfelder, 2017). In studies using the *memory design* (including the current study), participants recollect judgments they had previously made after obtaining outcome knowledge for some of the items. In studies using the *hypothetical design,* participants first obtain outcome knowledge for some items and then make hypothetical judgments ("What would you have estimated before you knew the answer?"; e.g., Hom & Ciaramitaro, 2001) for these items and original judgments for the remaining items. Further studies should test whether the hindsight bias on JOLs can also be found in a hypothetical design. Second, it seems worthwhile to examine whether the time between providing participants with outcome knowledge and JOL recollection would affect the hindsight bias on JOLs. Some prior studies have used a procedure similar to that of Experiment 1: People recollected their original judgments immediately after receiving outcome knowledge (e.g., Calvillo, 2013; Pohl et al., 2018). Other studies were similar to Experiments 2 and 3 in that obtaining outcome knowledge and recollecting original judgments took place in different phases of the experiment (e.g., Erdfelder & Buchner, 1998). While the two procedures usually yield similar results (Guilbault et al., 2004), there is some evidence to suggest that the hindsight bias is stronger when it comes to elapsed times of days or weeks between the event and the recollection of the original judgments. This was found in studies investigating reactions to the verdict in the O. J. Simpson criminal trial (Bryant & Brockway, 1997), the Clinton impeachment verdict (Bryant & Guilbault, 2002), and the outcomes of political elections (Blank et al., 2003). Examining whether the size of the hindsight bias on JOLs is affected by the delay between the memory test and JOL recollection might be an interesting avenue for future research.

In summary, this study demonstrated that providing outcome knowledge in the form of completing a memory test produces a hindsight bias on JOLs.

Notes

1. In many studies on the hindsight bias, correctly recollected judgments (ROJ = OJ) are excluded when computing ΔHB (see Pohl, 2007; Pohl & Erdfelder, 2017). In the experiments reported here, all results were unchanged when excluding correctly recollected judgments.
2. Notably, the three relevant effect sizes for the hindsight bias found in Experiment 1a were $\eta_p^2 = .19$ for the interaction between JOL type and recall success, $d = 0.63$ for the difference in resolution, and $d = 0.44$ for ΔHB. To ensure sufficient power, we assumed a medium effect to be able to detect effects in all three measures.
3. We did not specify predictions for ΔHB in the pre-registrations for Experiments 2 and 3. For reasons of consistency to Experiments 1a and 1b, we included ΔHB as a standard measure of hindsight bias magnitude (Pohl, 1992; Pohl & Erdfelder, 2017).
4. In the pre-registrations for Experiments 2 and 3, we did not specify that we planned to exclude participants with zero or perfect recall performance. For reasons of consistency, we applied the same exclusion criteria used in Experiments 1a and 1b. Importantly, including participants with zero or perfect recall performance from Experiments 2 and 3 did not change the reported results.
5. In the pre-registration, we specified that RJOLs would be a compromise between the anchor level and actual memory performance. More specifically, we expected lower RJOLs than OJOLs in the high-anchor group and higher RJOLs than OJOLs in the low-anchor group of items without outcome knowledge, resulting in an interaction between JOL type and

recall success, that is qualitatively different from the hindsight bias interaction between JOL type and recall. Although this prediction was not supported, the absence of the interaction here underlines the interpretation that the hindsight bias on JOLs is limited to items with outcome knowledge.

Disclosure statement

No potential conflict of interest was reported by the author(s).

The data for all experiments are available at https://osf.io/6jpcy and Experiments 2 and 3 were preregistered (https://osf.io/nehf2; https://osf.io/mrkzq).

Funding

This work was supported by the University of Mannheim's Graduate School of Economic and Social Sciences and by a grant from the Deutsche Forschungsgemeinschaft (DFG) to Monika Undorf (UN 345/1-3).

ORCID

Malte F. Zimdahl http://orcid.org/0000-0001-5808-2967
Monika Undorf http://orcid.org/0000-0002-0118-824X

References

Ackerman, R., Bernstein, D. M., & Kumar, R. (2020). Metacognitive hindsight bias. *Memory & Cognition*, *48*(5), 731–744. https://doi.org/10.3758/s13421-020-01012-w

Ariel, R., & Dunlosky, J. (2011). The sensitivity of judgment-of-learning resolution to past test performance, new learning, and forgetting. *Memory & Cognition*, *39*(1), 171–184. https://doi.org/10.3758/s13421-010-0002-y

Arkes, H. R., Faust, D., Guilmette, T. J., & Hart, K. (1988). Eliminating the hindsight bias. *Journal of Applied Psychology*, *73*(2), 305–307. https://doi.org/10.1037/0021-9010.73.2.305

Aust, F., Diedenhofen, B., Ullrich, S., & Musch, J. (2013). Seriousness checks are useful to improve data validity in online research. *Behavior Research Methods*, *45*(2), 527–535. https://doi.org/10.3758/s13428-012-0265-2

Bayen, U. J., Pohl, R. F., Erdfelder, E., & Auer, T.-S. (2007). Hindsight bias across the life span. *Social Cognition*, *25*(1), 83–97. https://doi.org/10.1521/soco.2007.25.1.83

Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General*, *127*(1), 55–68. https://doi.org/10.1037/0096-3445.127.1.55

Bernstein, D. M., Aßfalg, A., Kumar, R., & Ackerman, R. (2016). Looking backward and forward on hindsight bias. In J. Dunlosky, & S. K. Tauber (Eds.), *The Oxford handbook of metamemory* (pp. 289–304). Oxford University Press.

Bernstein, D. M., Erdfelder, E., Meltzoff, A. N., Peria, W., & Loftus, G. R. (2011). Hindsight bias from 3 to 95 years of age. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(2), 378–391. https://doi.org/10.1037/a0021971

Bernstein, D. M., Wilson, A. M., Pernat, N. L. M., & Meilleur, L. R. (2012). Auditory hindsight bias. *Psychonomic Bulletin & Review*, *19*(4), 588–593. https://doi.org/10.3758/s13423-012-0268-0

Biais, B., & Weber, M. (2009). Hindsight bias, risk perception, and investment performance. *Management Science*, *55*(6), 1018–1029. https://doi.org/10.1287/mnsc.1090.1000

Blank, H., Fischer, V., & Erdfelder, E. (2003). Hindsight bias in political elections. *Memory (Hove, England)*, *11*(4-5), 491–504. https://doi.org/10.1080/09658210244000513

Bröder, A., & Undorf, M. (2019). Metamemory viewed through the judgment lens. *Acta Psychologica*, *197*, 153–165. https://doi.org/10.1016/j.actpsy.2019.04.011

Bryant, F. B., & Brockway, J. H. (1997). Hindsight bias in reaction to the verdict in the O. J. Simpson criminal trial. *Basic and Applied Social Psychology*, *19*(2), 225–241. https://doi.org/10.1207/s15324834basp1902_5

Bryant, F. B., & Guilbault, R. L. (2002). "I knew It All along" eventually: The Development of hindsight bias in Reaction to the Clinton Impeachment verdict. *Basic and Applied Social Psychology*, *24*(1), 27–41. https://doi.org/10.1207/S15324834BASP2401_3

Buchanan, E. M., & Scofield, J. E. (2018). Methods to detect low quality data and its implication for psychological research. *Behavior Research Methods*, *50*(6), 2586–2596. https://doi.org/10.3758/s13428-018-1035-6

Calvillo, D. P. (2013). Rapid recollection of foresight judgments increases hindsight bias in a memory design. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(3), 959–964. https://doi.org/10.1037/a0028579

Campbell, J. D., & Tesser, A. (1983). Motivational interpretations of hindsight bias: An individual difference analysis. *Journal of Personality*, *51*(4), 605–620. https://doi.org/10.1111/j.1467-6494.1983.tb00868.x

Castel, A. D., Rhodes, M. G., McCabe, D. P., Soderstrom, N. C., & Loaiza, V. M. (2012). The fate of being forgotten: Information that is initially forgotten is judged as less important. *Quarterly Journal of Experimental Psychology*, *65*(12), 2281–2287. https://doi.org/10.1080/17470218.2012.739183

Chmielewski, M., & Kucker, S. C. (2020). An MTurk crisis? Shifts in data quality and the impact on study results. *Social Psychological and Personality Science*, *11*(4), 464–473. https://doi.org/10.1177/1948550619875149

Christensen-Szalanski, J. J., & Willham, C. F. (1991). The hindsight bias: A meta-analysis. *Organizational Behavior and Human Decision Processes*, *48*(1), 147–168. https://doi.org/10.1016/0749-5978(91)90010-Q

Difallah, D., Filatova, E., & Ipeirotis, P. (2018). Demographics and dynamics of Mechanical Turk workers. In Y. Chang, C. Zhai, Y. Liu, & Y. Maarek (Eds.), *Proceedings of the eleventh ACM international conference on web search and data mining* (pp. 135–143). ACM. https://doi.org/10.1145/3159652.3159661

England, B. D., & Serra, M. J. (2012). The contributions of anchoring and past-test performance to the underconfidence-with-practice effect. *Psychonomic Bulletin & Review*, *19*(4), 715–722. https://doi.org/10.3758/s13423-012-0237-7

Erdfelder, E., & Buchner, A. (1998). Decomposing the hindsight bias: A multinomial processing tree model for separating recollection and reconstruction in hindsight. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(2), 387–414. https://doi.org/10.1037/0278-7393.24.2.387

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Finn, B., & Metcalfe, J. (2007). The role of memory for past test in the underconfidence with practice effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(1), 238–244. https://doi.org/10.1037/0278-7393.33.1.238

Finn, B., & Metcalfe, J. (2008). Judgments of learning are influenced by memory for past test. *Journal of Memory and Language*, *58*(1), 19–34. https://doi.org/10.1016/j.jml.2007.03.006

Fischhoff, B. (1975). Hindsight is not equal to foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and*

Performance, *1*(3), 288–299. https://doi.org/10.1037/0096-1523.1.3.288

Fischhoff, B. (1977). Perceived informativeness of facts. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(2), 349–358. https://doi.org/10.1037/0096-1523.3.2.349

Fischhoff, B., & Beyth, R. (1975). I knew it would happen: Remembered probabilities of once-future things. *Organizational Behavior and Human Performance*, *13*(1), 1–16. https://doi.org/10.1016/0030-5073(75)90002-1

Formann, A. K., & Piswanger, K. (1979). *Wiener Matrizen-Test [Vienna matrices test]: Ein rasch-skalierter sprachfreier Intelligenztest*. Beltz.

Giroux, M. E., Coburn, P. I., Harley, E. M., Connolly, D. A., & Bernstein, D. M. (2016). Hindsight bias and law. *Zeitschrift für Psychologie*, *224*(3), 190–203. https://doi.org/10.1027/2151-2604/a000253

Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of mechanical Turk samples. *Journal of Behavioral Decision Making*, *26*(3), 213–224. https://doi.org/10.1002/bdm.1753

Granhag, P. A., Strömwall, L. A., & Allwood, C. M. (2000). Effects of reiteration, hindsight bias, and memory on realism in eyewitness confidence. *Applied Cognitive Psychology*, *14*(5), 397–420. https://doi.org/10.1002/1099-0720(200009)14:5<397::AID-ACP658>3.0.CO;2-M

Guilbault, R. L., Bryant, F. B., Brockway, J. H., & Posavac, E. J. (2004). A meta-analysis of research on hindsight bias. *Basic and Applied Social Psychology*, *26*(2-3), 103–117. https://doi.org/10.1207/s15324834basp2602&3_1

Halamish, V. (2018). Can very small font size enhance memory? *Memory & Cognition*, *46*(6), 979–993. https://doi.org/10.3758/s13421-018-0816-6

Harley, E. M., Carlsen, K. A., & Loftus, G. R. (2004). The "saw-it-all-along" effect: Demonstrations of visual hindsight bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(5), 960–968. https://doi.org/10.1037/0278-7393.30.5.960

Hawkins, S. A., & Hastie, R. (1990). Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin*, *107*(3), 311–327. https://doi.org/10.1037/0033-2909.107.3.311

Hell, W., Gigerenzer, G., Gauggel, S., Mall, M., & Müller, M. (1988). Hindsight bias: An interaction of automatic and motivational factors? *Memory & Cognition*, *16*(6), 533–538. https://doi.org/10.3758/BF03197054

Hines, J. C., Touron, D. R., & Hertzog, C. (2009). Metacognitive influences on study time allocation in an associative recognition task: An analysis of adult age differences. *Psychology and Aging*, *24*(2), 462–475. https://doi.org/10.1037/a0014417

Hoffrage, U., & Pohl, R. F. (2003). Research on hindsight bias: A rich past, a productive present, and a challenging future. *Memory (Hove, England)*, *11*(4-5), 329–335. https://doi.org/10.1080/09658210344000080

Hom, H. L., & Ciaramitaro, M. (2001). GTIDHNIHS: I knew-it-all-along. *Applied Cognitive Psychology*, *15*(5), 493–507. https://doi.org/10.1002/acp.716

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*(4), 349–370. https://doi.org/10.1037/0096-3445.126.4.349

Koriat, A. (2015). Metacognition: Decision making processes in self-monitoring and self-regulation. In G. Keren, & G. Wu (Eds.), *The Wiley Blackwell handbook of judgment and decision making* (pp. 356–379). Wiley-Blackwell.

Kornell, N., & Rhodes, M. G. (2013). Feedback reduces the metacognitive benefit of tests. *Journal of Experimental Psychology: Applied*, *19*(1), 1–13. https://doi.org/10.1037/a0032147

Luna, K., Martín-Luengo, B., & Albuquerque, P. B. (2018). Do delayed judgements of learning reduce metamemory illusions? A meta-analysis. *Quarterly Journal of Experimental Psychology*, *71*(7), 1626–1636. https://doi.org/10.1080/17470218.2017.1343362

Melinger, A., & Weber, A. (2006). Database of Noun Associations for German. http://www.coli.uni-saarland.de/projects/nag/

Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review*, *15*(1), 174–179. https://doi.org/10.3758/PBR.15.1.174

Mueller, M. L., Dunlosky, J., Tauber, S. K., & Rhodes, M. G. (2014). The font-size effect on judgments of learning: Does it exemplify fluency effects or reflect people's beliefs about memory? *Journal of Memory and Language*, *70*, 1–12. https://doi.org/10.1016/j.jml.2013.09.007

Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The 'delayed-JOL effect.'. *Psychological Science*, *2*(4), 267–271. https://doi.org/10.1111/j.1467-9280.1991.tb00147.x

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 402–407. https://doi.org/10.3758/BF03195588

Peer, E., Vosgerau, J., & Acquisti, A. (2014). Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior Research Methods*, *46*(4), 1023–1031. https://doi.org/10.3758/s13428-013-0434-y

Pohl, R. (1992). Der Rückschau-Fehler: Systematische Verfälschung der Erinnerung bei Experten und Novizen [Hindsight bias: Systematic distortions of recall in experts and novices]. *Kognitionswissenschaft*, *3*, 38–44.

Pohl, R. F. (2007). Ways to assess hindsight bias. *Social Cognition*, *25*(1), 14–31. https://doi.org/10.1521/soco.2007.25.1.14

Pohl, R. F., Bayen, U. J., Arnold, N., Auer, T.-S., & Martin, C. (2018). Age differences in processes underlying hindsight bias: A life-span study. *Journal of Cognition and Development*, *19*(3), 278–300. https://doi.org/10.1080/15248372.2018.1476356

Pohl, R. F., Bender, M., & Lachmann, G. (2002). Hindsight bias around the world. *Experimental Psychology*, *49*(4), 270–282. https://doi.org/10.1027//1618-3169.49.4.270

Pohl, R. F., Eisenhauer, M., & Hardt, O. (2003a). SARA: A cognitive process model to simulate the anchoring effect and hindsight bias. *Memory (Hove, England)*, *11*(4-5), 337–356. https://doi.org/10.1080/09658210244000487

Pohl, R. F., & Erdfelder, E. (2017). Hindsight bias. In R. F. Pohl (Ed.), *Cognitive illusions: Intriguing phenomena in thinking, judgment and memory* (2nd ed., pp. 424–445). Routledge.

Pohl, R. F., Schwarz, S., Sczesny, S., & Stahlberg, D. (2003b). Hindsight bias in gustatory judgments. *Experimental Psychology*, *50*(2), 107–115. https://doi.org/10.1026//1618-3169.50.2.107

Rhodes, M. G. (2016). Judgments of learning: Methods, data, and theory. In J. Dunlosky, & S. K. Tauber (Eds.), *The Oxford handbook of metamemory* (pp. 65–80). Oxford University Press.

Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology: General*, *137*(4), 615–625. https://doi.org/10.1037/a0013684

Rhodes, M. G., Witherby, A. E., Castel, A. D., & Murayama, K. (2017). Explaining the forgetting bias effect on value judgments: The influence of memory for a past test. *Memory & Cognition*, *45*(3), 362–374. https://doi.org/10.3758/s13421-016-0674-z

Roese, N. J., & Maniar, S. D. (2016). Perceptions of purple: Counterfactual and hindsight judgments at Northwestern Wildcats football games. *Personality and Social Psychology Bulletin*, *23*(12), 1245–1253. https://doi.org/10.1177/01461672972312002

Roese, N. J., & Vohs, K. D. (2012). Hindsight bias. *Perspectives on Psychological Science*, *7*(5), 411–426. https://doi.org/10.1177/1745691612454303

Schooler, L. J., & Hertwig, R. (2005). How forgetting aids heuristic inference. *Psychological Review*, *112*(3), 610–628. https://doi.org/10.1037/0033-295X.112.3.610

Serra, M. J., & Ariel, R. (2014). People use the memory for past-test heuristic as an explicit cue for judgments of learning. *Memory & Cognition*, *42*(8), 1260–1272. https://doi.org/10.3758/s13421-014-0431-0

Sitzman, D. M., Rhodes, M. G., & Kornell, N. (2016). The influence of feedback on predictions of future memory performance. *Memory & Cognition*, *44*(7), 1102–1113. https://doi.org/10.3758/s13421-016-0623-x

Sitzman, D. M., Rhodes, M. G., & Tauber, S. K. (2014). Prior knowledge is more predictive of error correction than subjective confidence. *Memory & Cognition*, *42*(1), 84–96. https://doi.org/10.3758/s13421-013-0344-3

Soderstrom, N. C., & McCabe, D. P. (2011). The interplay between value and relatedness as bases for metacognitive monitoring and control: Evidence for agenda-based monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(5), 1236–1242. https://doi.org/10.1037/a0023548

Undorf, M., & Ackerman, R. (2017). The puzzle of study time allocation for the most challenging items. *Psychonomic Bulletin & Review*, *24*(6), 2003–2011. https://doi.org/10.3758/s13423-017-1261-4

Undorf, M., Söllner, A., & Bröder, A. (2018). Simultaneous utilization of multiple cues in judgments of learning. *Memory & Cognition*, *46*(4), 507–519. https://doi.org/10.3758/s13421-017-0780-6

Undorf, M., & Zimdahl, M. F. (2019). Metamemory and memory for a wide range of font sizes: What is the contribution of perceptual fluency? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(1), 97–109. https://doi.org/10.1037/xlm0000571

Witherby, A. E., Tauber, S. K., Rhodes, M. G., & Castel, A. D. (2019). Aging and forgetting: Forgotten information is perceived as less important than is remembered information. *Psychology and Aging*, *34*(2), 228–241. https://doi.org/10.1037/pag0000322

Yang, C., Sun, B., & Shanks, D. R. (2018). The anchoring effect in metamemory monitoring. *Memory & Cognition*, *46*(3), 384–397. https://doi.org/10.3758/s13421-017-0772-6

Appendix

To further explore the influence of the anchor manipulation on the recollection of JOLs, we conducted four separate 2 (JOL type: OJOL, RJOL) × 2 (recall success: yes, no) within-subjects ANOVAs for items with and without outcome knowledge in the high- and low-anchor group. All four ANOVAs revealed significant main effects of recall success, $F(1, 103) = 70.94$, $MSE = 119.07$, $p < .001$, $\eta_p^2 = .41$, and $F(1, 107) = 89.23$, $MSE = 129.63$, $p < .001$, $\eta_p^2 = .46$, for items with outcome knowledge in the high- and low-anchor group, respectively, and $F(1, 103) = 32.09$, $MSE = 107.17$, $p < .001$, $\eta_p^2 = .24$, and $F(1, 107) = 80.75$, $MSE = 77.64$, $p < .001$, $\eta_p^2 = .43$, for items without outcome knowledge in the high- and low-anchor group, respectively. The two ANOVAs on items without outcome knowledge revealed significant main effects of JOL type, $F(1, 103) = 10.45$, $MSE = 153.52$, $p = .002$, $\eta_p^2 = .09$, and $F(1, 107) = 4.61$, $MSE = 94.38$, $p = .034$, $\eta_p^2 = .04$, for the high- and low-anchor group, respectively. The main effects of JOL type were not significant in the two ANOVAs on items with outcome knowledge, both Fs < 1.34, both ps > .250. Most importantly, the interaction between JOL type and recall success was significant in the two ANOVAs on items with outcome knowledge, $F(1, 103) = 6.07$, $MSE = 59.31$, $p = .015$, $\eta_p^2 = .06$, and $F(1, 107) = 9.92$, $MSE = 59.11$, $p = .002$, $\eta_p^2 = .09$, for the high- and low-anchor group, respectively, but not in the two ANOVAs on items without outcome knowledge, both Fs < 1. These findings again indicated that the hindsight bias on JOLs was limited to items with outcome knowledge.

Warnings and Incentives Do Not Reduce Hindsight Bias on Judgments of Learning

Malte F. Zimdahl and Monika Undorf

Department of Psychology, School of Social Sciences, University of Mannheim

Abstract

People often overestimate in hindsight what they knew in foresight. Recent evidence shows that hindsight bias also afflicts people's predictions of their own future memory performance (judgments of learning, JOLs). The current study investigated the robustness of hindsight bias on JOLs by testing whether it can be reduced with two different debiasing methods: warnings and incentives. In three experiments, participants learned word pairs and made JOLs. Afterwards, participants were tested on half the word pairs, which provided them with outcome knowledge for these items. Participants then recollected their JOLs before being tested on the other half of the word pairs. In Experiment 1, one group of participants received detailed information about hindsight bias on JOLs and instructions to avoid it before recollecting their JOLs, whereas another group did not receive any warning. In Experiments 2 and 3, participants received a high monetary incentive for correctly recollecting their JOLs to some items and a low monetary incentive for correctly recollecting their JOLs to other items. All three experiments revealed hindsight bias on JOLs for word pairs with outcome knowledge only. Neither warnings nor incentives reduced hindsight bias on JOLs, with Bayesian analyses providing evidence in favor of similar hindsight biases across the warning and no-warning group (Experiment 1) and across high- and low-incentive items (Experiments 2 and 3). These results demonstrate that hindsight bias on JOLs is a robust phenomenon that persists despite people's awareness of the bias or increased motivation to avoid it.

Keywords: hindsight bias, judgments of learning, metamemory, warnings, incentives

In hindsight, people often overestimate what they knew in foresight (Fischhoff, 1975; Hawkins & Hastie, 1990; Wood, 1978). For example, in medical decision making, physicians overestimate their ability to correctly diagnose

---

This research was supported by the University of Mannheim's Graduate School of Economic and Social Sciences and by grant UN 345/1-3 from the German Research Foundation (DFG) to Monika Undorf.

Materials, data, and analysis code for all experiments can be found on the Open Science Framework (https://osf.io/jznyr).

Correspondence concerning this article should be addressed to Malte F. Zimdahl, Department of Psychology, School of Social Sciences, University of Mannheim, A5, 6, 68159 Mannheim, Germany. Email: malte.zimdahl@psychologie.uni-mannheim.de

a patient's condition after they have learned what he actually suffers from (Arkes et al., 1981). In legal decision making, mock jurors overestimate the foreseeability of a case's outcome after they have acquired knowledge about it (Hastie et al., 1999). In economic decision making, management students overestimate the accuracy of their initial value ratings for different companies after they have found out how the companies developed (Louie, 1999). These and other biased recollections of previously made judgments towards actual outcomes have been summarized under the term *hindsight bias* (Fischhoff, 1975). Hindsight bias phenomena have been addressed in a large number of studies throughout the last 45 years (see, e.g., Bernstein et al., 2016; Pohl & Erdfelder, 2022; Roese & Vohs, 2012, for reviews). They were demonstrated using various materials (e.g., Bernstein et al., 2012; Pohl, Schwarz, et al., 2003) and in people of various ages (Bayen et al., 2007; Bernstein et al., 2011) and from various cultural backgrounds (Pohl et al., 2002). Overall, hindsight bias phenomena afflict a variety of judgments and have considerable real-world consequences (see, e.g., Giroux et al., 2016; Louie et al., 2007; Pezzo, 2011).

Hindsight Bias on JOLs

In a recent study, it could be shown that hindsight bias also afflicts *judgments of learning* (JOLs, Zimdahl & Undorf, 2021). JOLs are metamemory judgments that assess people's ability to predict their memory performance in a future test. In the mentioned study, participants studied word pairs and predicted their chances of remembering each pair in a cued recall test where they were provided with the first word of each pair and asked to recall the second. Participants then recollected their JOLs either before or after their memory for the word pairs was tested. Results showed that people's recollected JOLs (RJOLs) were similar to their original JOLs (OJOLs) when recollection took place prior to the memory test (without outcome knowledge), but systematically biased when recollection took place after the memory test (with outcome knowledge).

This hindsight bias on JOLs was apparent in three different measures. First, the difference in JOLs between recalled and not-recalled word pairs was larger for RJOLs than for OJOLs, showing that participants remembered their original judgments as being closer to actual recall than they were. Second, RJOLs had a higher resolution than OJOLs, as indicated by higher within-subjects Goodman-Kruskal gamma correlations between JOLs and recall performance (Goodman & Kruskal, 1954). Third, a standard measure of hindsight bias magnitude (ΔHB, Pohl, 1992) also indicated that RJOLs were biased towards actual recall. Importantly, all three measures showed that RJOLs were unbiased when obtained before the memory test. This finding demonstrated that outcome knowledge produced hindsight bias in metamemory (Zimdahl & Undorf, 2021).

The current study investigated the robustness of hindsight bias on JOLs. This is important for several reasons. First, replicating the finding of hindsight bias on JOLs demonstrated by Zimdahl and Undorf (2021) would emphasize the significance of this novel effect (for an overview of the importance of replication for good science, see Asendorpf et al., 2013). Second, while there is much evidence that hindsight bias is a robust phenomenon (e.g., Pohl, 2007; Pohl & Erdfelder, 2022), several biases that afflict JOLs can be reduced or eliminated (see, e.g., Undorf et al., 2022). It is therefore an open question whether hindsight bias on JOLs is robust against established debiasing methods. Third, attempts to debias hindsight bias on JOLs have the potential to shed light on the theoretical basis of this bias. More specifically, investigating whether hindsight bias on JOLs can be reduced by (1) warning participants about hindsight bias and instructing them to avoid it and (2) incentivizing them for correct JOL recollections allow us to test whether hindsight bias on JOLs is due to a lack of awareness of the bias or of the motivation to avoid it. Conversely, discovering that neither warnings nor incentives reduce hindsight bias on JOLs would suggest that it is due to memory distortions that are hard to avoid (see Blank et al., 2008).

The Effects of Warnings

Several studies have attempted to reduce hindsight biases with warnings. For example, Fischhoff (1977) gave one group of participants a detailed description of hindsight bias together with the instruction to avoid it when recollecting probabilities that an experimenter-provided answer to a two-alternative forced choice question was correct. Results showed that hindsight bias in this group was similar to that in a group that received no information about hindsight bias. Pohl and Hell (1996) replicated this finding when informing participants about hindsight bias before the experiment started or immediately after their original judgments. Harley et al. (2004) found that providing warnings about hindsight bias immediately before participants recollected their original judgments did not reduce hindsight bias. Together, these findings indicate that hindsight biases are very robust and often unaffected by warnings.

Warnings have, however, been successful at reducing biases that afflict JOLs. For example, Koriat and Bjork (2006) found that warnings reduced foresight bias on JOLs (see Koriat & Bjork, 2005). Foresight bias refers to inflated JOLs for material that seems easy to remember at study but is actually difficult to remember (e.g., *cheddar* - *cheese* appears semantically related and highly memorable at study, but when *cheddar* is presented alone at test, several other related words including *mouse*, *milk*, or *cracker* interfere with the correct response and reduce memory performance). Importantly, informing participants about foresight bias and asking them to keep this information in mind when making JOLs reduced this bias (Koriat & Bjork, 2006). In a similar vein, Rhodes and Castel (2008) found that warnings reduced the font-size illusion, that is, higher JOLs for study words printed in a larger font than for study words printed in a smaller font despite similar memory performance across font sizes. Furthermore, Carpenter and Olson (2012) found that warnings to avoid overconfidence prior to the study phase reduced overconfidence in JOLs for learning words in a foreign language by studying the word paired with a picture (e.g., a picture of a dog paired with the Swahili word *kelb*). In contrast, Yan et al. (2016) reported that warnings failed to reduce the interleaving illusion, that is, people's erroneous belief that blocking study materials is more effective than spacing study materials. Taken together, warnings were found to reduce or even eliminate some but not all biases that afflict JOLs. It is therefore an open question whether warnings reduce hindsight bias on JOLs.

The Effects of Incentives

Apart from warnings, studies have attempted to reduce hindsight biases with incentives. For example, Camerer et al. (1989) presented participants with business cases that included the earnings of various companies and asked them to estimate the earnings predicted by participants who were unaware of the actual earnings. Rewarding participants with $1.00 if their estimates were within 10% of the naïve participants' predictions did not reduce hindsight bias as compared to control participants. In a similar study, Hell et al. (1988) combined an incentive manipulation (high vs. low monetary bonus for correctly recalling numerical responses to questions) with whether or not participants had to give reasons for their original judgments, and whether they received outcome knowledge immediately after making the original judgment or immediately before recollecting it. Results showed that hindsight bias was smaller when high incentives were combined with giving reasons and receiving outcome knowledge immediately after the original response than when low incentives were combined with not giving reasons and receiving outcome knowledge immediately before recollecting judgments. This study suggests that hindsight biases can be reduced using incentives.

Very little is known about whether incentives can alleviate biased JOLs. To our knowledge, the only study to date (Koriat et al., 2004) has examined the influence of incentives on people underestimating future forgetting. When different groups of participants make JOLs for tests after retention intervals of 10 minutes, 1 day, or 1 week, JOLs are typically insensitive to the retention interval, whereas memory performance declines with longer retention intervals. This bias was unaffected by monetary incentives for correct predictions. Taken together, there is too little evidence to evaluate whether incentives reduce or eliminate biases that afflict JOLs. It is therefore an open question whether incentives reduce hindsight bias on JOLs.

The Current Study

In the current study, we investigated whether warnings and incentives reduce hindsight bias on JOLs. In Experiment 1, one group of participants received detailed information about hindsight bias on JOLs and was asked to avoid this bias before recollecting JOLs, whereas another group did not receive any information about hindsight bias. In Experiments 2 and 3, correct JOL recollections were incentivized by a monetary bonus, which was 10 times (Experiment 2) or 30 times (Experiment 3) higher for high-value items than for low-value items.

As in previous research, hindsight bias on JOLs should become apparent in three different measures (see Zimdahl & Undorf, 2021). First, hindsight bias on JOLs should result in a larger difference in JOLs between recalled and not-recalled items for RJOLs than for OJOLs, showing that participants' RJOLs are closer to the outcome than their OJOLs. Second, hindsight bias on JOLs should produce higher resolution (as measured by within-subjects Goodman-Kruskal gamma correlations) of RJOLs than of OJOLs. Third, the standard measure of hindsight bias magnitude ΔHB (Pohl, 1992) should be significantly positive for items with outcome knowledge but insignificant for items without outcome knowledge. ΔHB is defined as

$$\Delta HB = |OJOL - CJOL| - |RJOL - CJOL| \quad (1)$$

and indicates to what extent the RJOL is closer to the correct outcome (CJOL) than the OJOL. CJOL corresponds to 100 if an item is recalled, and to 0 if an item is not recalled. ΔHB is positive when the RJOL is closer to the CJOL than the OJOL, therefore indicating hindsight bias on JOLs. An insignificant ΔHB suggests that there is no hindsight bias on JOLs.

If warnings reduce hindsight bias on JOLs in Experiment 1, then the warning group should reveal a smaller difference in RJOLs between recalled and not-recalled items, a smaller increase in resolution between OJOLs and RJOLs, and a smaller ΔHB than the no-warning group. If incentives reduce hindsight bias on JOLs in Experiments 2 and 3, then the high-value items should reveal a smaller difference in RJOLs between recalled and not-recalled items, a smaller increase in resolution between OJOLs and RJOLs, and a smaller ΔHB than the low-value items. Because reduced hindsight bias on JOLs could also manifest as an increase in correct JOL recollections, we compared the proportions of correct RJOLs across the warning and no-warning groups in Experiment 1 and across high- and low-value items in Experiments 2 and 3.

Experiment 1

Experiment 1 examined whether warnings reduce hindsight bias on JOLs. Participants studied 60 word pairs and made JOLs. After being tested on half of the items, they recollected their JOLs and were then tested on the other half of the items. Immediately before recollecting their JOLs, participants in the warning group received detailed information about hindsight bias on JOLs and instructions to avoid it. Participants in the no-warning group received no information about hindsight bias on JOLs.

For both groups, we expected no hindsight bias on JOLs for items that were tested after JOLs were recollected (without outcome knowledge). For the no-warning group, we expected hindsight bias on JOLs for items that were tested before JOLs were recollected (with outcome knowledge). Hindsight bias on JOLs should become evident in a significant interaction between JOL type and recall success, higher resolution of RJOLs than of OJOLs, and a significantly positive ΔHB. To test whether warnings reduce hindsight bias on JOLs, we compared all three hindsight bias measures across the warning- and no-warning groups. With regard to memory performance, we expected higher memory performance for items without outcome knowledge because these items were tested after word pairs were presented in the JOL recollection phase.

Materials, data, and analysis code for this and the subsequent experiments can be found on the Open Science Framework (https://osf.io/jznyr).

Method

Design

The experiment used a 2 (JOL type: OJOL, RJOL) $\times$ 2 (Recall success: yes, no) $\times$ 2 (Outcome knowledge: yes, no) $\times$ 2 (Warning group: warning, no-warning) mixed design with JOL type, recall success, and outcome knowledge as within-subjects factors and warning group as a between-subjects factor.

Participants

An a priori power analysis using G*Power 3 (Faul et al., 2007) revealed that 64 participants are required to detect medium sized effects (Cohen's f = 0.25) with α = .05 and a statistical power of 1 - β = .95 when conducting a mixed ANOVA assuming a correlation of .50 between repeated measures. We recruited 94 participants from the University of Mannheim community. Participants completed the study online at a place and time of their choosing and psychology undergraduates were compensated with course credit. We used the following planned exclusion criteria in this and the subsequent experiment (for a similar approach, see Zimdahl & Undorf, 2021): We excluded participants who reported that they had engaged in other tasks during the experiment (n = 2), participants who reported that they had copied words (none in this experiment), participants with zero recall (n = 10), and participants with perfect recall (none in this experiment). Thus, the final sample included 82 participants (68 female, 11 male, 1 diverse, 2 who did not indicate their gender) with a mean age of 24.56 years (SD = 8.23). All participants were fluent in German. Participants were randomly assigned to either the warning group (n = 41) or to the no-warning group (n = 41).

Materials

Stimuli were 56 unrelated word pairs consisting of German 4-8 letter nouns. Words were taken from the Berlin Affective Word List Reloaded (Võ et al., 2009) and were of neutral valence (M = 0.19, SD = 0.46), moderate arousal (M = 2.47, SD = 0.39), and moderate imageability (M = 3.56, SD = 0.94). Four additional unrelated word pairs served as primacy buffers and were not included in the analysis.

Procedure

The experiment consisted of a study phase, a filler task, and a cued recall test with an interpolated JOL recollection phase. Instructions at the beginning of the experiment informed participants that they would study 60 word pairs and would be asked to recall the second word of each pair in response to the first word in a later memory test. Participants were also told that they should, immediately after studying each word pair, estimate the probability of remembering the second word in the cued recall test. At study, each trial started with a 500-ms fixation cross presented at the top of the screen, followed by a word pair displayed for 2 s at the same location. Immediately after the word pair, the JOL prompt "The chance to recall (0%-100%)?" appeared at the bottom of the screen. Participants indicated their JOL in a self-paced manner by entering a number between 0 and 100 on the keyboard. A 500-ms blank screen preceded the next study trial. After the study phase, participants performed an unrelated filler task for 90 s, which consisted of items from the Wiener Matrizen Test (Formann & Piswanger, 1979). In the cued recall test, the first word of each word pair was presented at the top of the screen and participants were asked to type in the second word. Immediately afterwards, the correct word appeared beneath the participant's response for 2 s to provide feedback about whether it was correctly recalled. After being tested on half (30) of the word pairs, participants recollected their JOLs they had made during study. Word pairs were shown for 2 s before the JOL prompt "What chance of recall (0%-100%) did you assign during study?" was added to the screen and participants indicated their RJOL in a self-paced manner. Prior to the JOL recollection phase, participants in the warning group received detailed information about hindsight bias on JOLs (the verbatim instructions, translated into English, can be found in the Appendix). It was explained to participants that outcome knowledge had influenced participants' RJOLs in former studies. They were told that RJOLs made with outcome knowledge were higher than OJOLs for recalled items but lower than OJOLs for not-recalled items. Participants were instructed to avoid hindsight bias on JOLs. Additionally, their understanding of hindsight bias on JOLs was tested using three two-alternative forced-choice questions (see Appendix). Participants were only allowed to start with the JOL recollection phase after they had correctly responded to all three questions. Participants in the no-warning group did not receive any information about hindsight bias on JOLs. After the JOL recollection phase, participants completed the second half of the cued recall test (remaining 30 word pairs). In the study phase, the cued recall test, and the JOL recollection phase, word pairs were presented in a new random order for each participant. At the end of the experiment, participants completed a questionnaire about their commitment to the study and possible cheating. Afterwards, they were debriefed and thanked for their participation. The whole experiment lasted about 30 minutes.

Results and Discussion

Memory performance was analyzed using a 2 (Outcome knowledge: yes, no) $\times$ 2 (Warning group: warning, no warning) mixed ANOVA with outcome knowledge as a within-subjects factor and warning group as a between-subjects factor. As expected, a main effect of outcome knowledge revealed lower memory performance in the first than in the second half of the cued recall test, $F(1, 80) = 75.46$, $MSE = 0.01$, $p < .001$, $\eta_p^2 = .49$ (see Table 1). The main effect of warning group and the interaction were not significant, Fs < 1.

To test whether warnings influenced hindsight bias on JOLs, we conducted a 2 (JOL type: OJOL, RJOL) $\times$ 2 (Recall success: yes, no) $\times$ 2 (Outcome knowledge: yes, no) $\times$ 2 (Warning group: warning, no-warning) mixed ANOVA (see Figure 1). A main effect of recall success indicated higher JOLs for recalled than for not-recalled items, $F(1, 80) = 147.55$, $MSE = 202.49$, $p < .001$, $\eta_p^2 = .65$. A main effect of JOL type revealed higher OJOLs than RJOLs, $F(1, 80) = 5.85$, $MSE = 54.95$, $p = .018$, $\eta_p^2 = .07$. A main effect of outcome knowledge showed higher JOLs for items with than without outcome knowledge, $F(1, 80) = 11.10$, $MSE = 129.75$, $p = .001$, $\eta_p^2 = .12$. The interaction between JOL type and outcome knowledge was significant, indicating that the difference between OJOLs and RJOLs was larger for items without than with outcome knowledge, $F(1, 80) = 19.19$, $MSE = 23.66$, $p < .001$, $\eta_p^2 = .19$. An interaction between JOL type and recall success revealed hindsight bias on JOLs, $F(1, 80) = 11.32$, $MSE = 24.60$, $p = .001$, $\eta_p^2 = .12$, with an interaction among JOL type,

Table 1
Means (and Standard Deviations) of Recall, Gamma Correlations Between JOLs and Recall, ΔHB, and Correct JOL Recollections for Experiment 1

	Outcome knowledge		No outcome knowledge	
	Warning	No warning	Warning	No warning
p(Recall)	.26 (.21)	.30 (.17)	.39 (.24)	.43 (.16)
G(OJOL, Recall)	.41 (.32)	.40 (.31)	.35 (.34)	.34 (.32)
G(RJOL, Recall)	.57 (.27)	.62 (.25)	.40 (.33)	.41 (.35)
ΔHB	2.36**	1.96*	1.21	0.99
p(RJOL = OJOL)	.30 (.19)	.32 (.15)	.30 (.18)	.36 (.16)

Note. *p*(Recall) = percentage of correctly recalled items; OJOL = original judgment of learning; RJOL = recollected judgment of learning; ΔHB = Pohl's (1992) difference measure of hindsight bias; *p*(RJOL = OJOL) = percentage of correct JOL recollections.
* $p < .05$, ** $p < .01$.

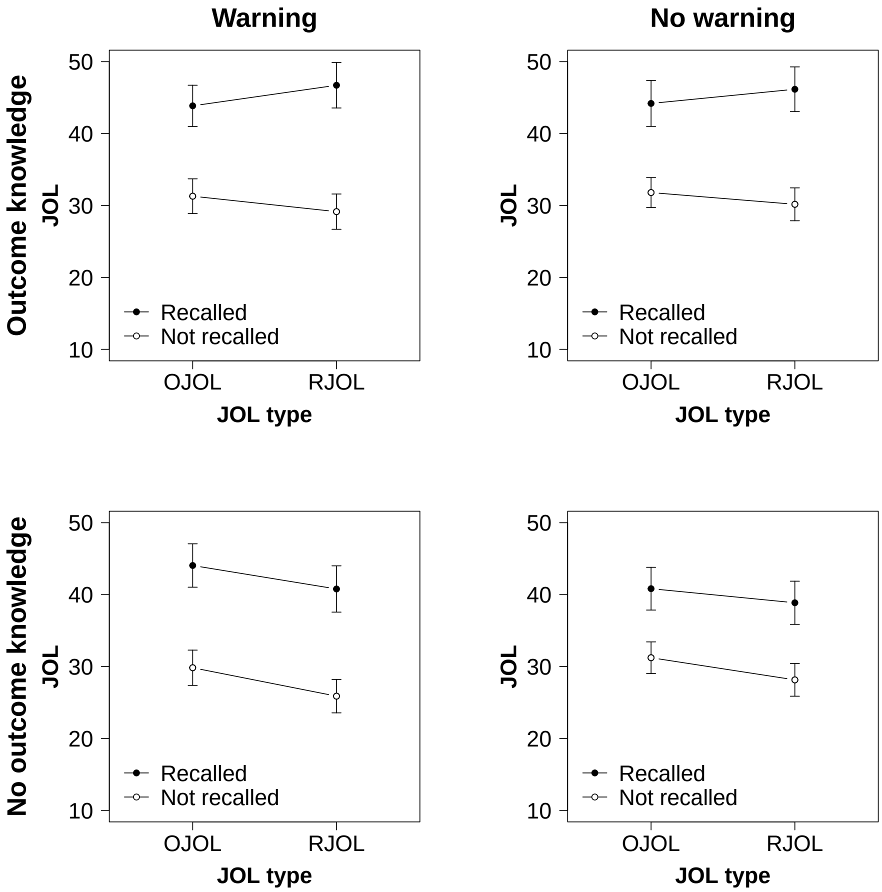
recall success, and outcome knowledge suggesting that it was limited to the items with outcome knowledge, $F(1, 80) = 7.79$, $MSE = 15.24$, $p = .007$, $\eta_p^2 = .09$. Most importantly, there was no main effect of warning group, nor any interactions involving warning group, all *F*s < 1.41, all *p*s > .239. This finding indicated that warning participants did not have any effect on hindsight bias on JOLs. No other effects were significant, all *F*s < 2.13, all *p*s > .148.

To test whether there was positive evidence that hindsight bias on JOLs was unaffected by warnings, we ran a Bayesian ANOVA using default priors as suggested in Rouder et al. (2017) using the *R* package *BayesFactor* (Morey & Rouder, 2015). Specifically, we tested the full model described above against a model without the four-way interaction (JOL type × Recall success × Outcome knowledge × Warning group) that is indicative of potential effects of warnings on hindsight bias on JOLs. A Bayes Factor (BF) of 0.243 indicated that the data were 4.12 times more likely under the model without the four-way interaction than under the full model, providing moderate evidence for the hypothesis that warnings did not influence hindsight bias on JOLs.

As can be seen in Table 1, for items with outcome knowledge, resolution of RJOLs was higher than resolution of OJOLs in either group, warning group: $t(40) = 3.76$, $p < .001$, $d = 0.57$; no-warning group: $t(40) = 4.32$, $p < .001$, $d = 0.78$. Importantly, the size of this increase in resolution was similar across groups, $t(80) = 0.76$, $p = .452$, $d = 0.17$, again indicating that hindsight bias on JOLs was unaffected by warnings. This was further supported by a Bayesian *t*-test revealing a BF of 0.295 that provided moderate evidence for the null hypothesis according to which the increase in resolution was equal across groups (i.e., the data are 3.39 times more likely under the null hypothesis). For items without outcome knowledge, resolution of RJOLs did not differ from that of OJOLs, warning group: $t(40) = 1.36$, $p = .181$, $d = 0.16$; no-warning group: $t(40) = 1.99$, $p = .053$, $d = 0.20$.

The standard measure of hindsight bias magnitude ΔHB demonstrated hindsight bias on JOLs for items with outcome knowledge in either group, warning group: $t(40) = 3.37$, $p = .002$, $d = 0.53$; no-warning group: $t(40) = 2.63$, $p = .012$, $d = 0.41$ (see Table 1). Crucially, ΔHB did not differ between groups,

Figure 1
Mean Original Judgments of Learning (OJOLs) and Recollected Judgments of Learning (RJOLs) in Experiment 1



Note. Top panels present items with outcome knowledge and bottom panels present items without outcome knowledge in the warning group (left panels) and the no-warning group (right panels) in Experiment 1. Filled symbols represent recalled items and open symbols represent not-recalled items. Error bars represent one standard error of the mean.

$t(80) = 0.39$, $p = .699$, $d = 0.09$, again suggesting that warning participants did not reduce hindsight bias on JOLs. A Bayesian t-test provided moderate evidence (BF = 0.246; 4.07 times more likely) for the null hypothesis according to which ΔHB was equal across groups. ΔHB was not significantly different from zero for items without outcome knowledge in either group, warning group: $t(40) = 1.89$, $p = .066$, $d = 0.30$; no-warning group: $t(40) = 1.88$, $p = .068$, $d = 0.29$.

Proportions of correctly recollected JOLs (see Table 1) did not differ between groups, with outcome knowledge: $t(80) = 0.69$, $p = .491$, $d = 0.15$, BF = 0.283; without outcome knowledge: $t(80) = 1.75$, $p = .085$, $d = 0.39$, BF = 0.857, arguing against the possibility that warning participants increased the number of correct JOL recollections.

In summary, Experiment 1 revealed hindsight bias on JOLs for items with outcome knowledge only. Hindsight bias on JOLs was evident in a significant interaction between JOL type and recall success, in higher resolution of RJOLs than of OJOLs, and in a significantly positive ΔHB. Most importantly, frequentist as well as Bayesian analyses demonstrated that warnings did not reduce hindsight bias on JOLs.

Experiment 2

Experiment 2 examined whether incentives reduce hindsight bias on JOLs. During the JOL recollection phase, items were assigned either a high or a low value, indicating that participants would receive 10 points (corresponding to £0.10) or 1 point (corresponding to £0.01) for correctly recollecting the item's JOL. As in Experiment 1, we expected unbiased JOL recollections for items without outcome knowledge, but hindsight bias on JOLs for items with outcome knowledge. To test whether incentives reduce hindsight bias on JOLs, we compared three hindsight bias measures (i.e., significant interaction between JOL type and recall success, higher resolution of RJOLs than of OJOLs, and a significantly positive ΔHB) across high- and low-value items.

Method

Design

The experiment used a 2 (JOL type: OJOL, RJOL) × 2 (Recall success: yes, no) × 2 (Outcome knowledge: yes, no) × 2 (Value: high, low) within-subjects design.

Participants

An a priori power analysis using G*Power 3 (Faul et al., 2007) revealed that 54 participants are required to detect medium sized effects (Cohen's f = 0.25) with α = .05 and a statistical power of 1 - β = .95 when conducting a within-subjects ANOVA assuming a correlation of .50 between repeated measures. We recruited 64 participants from the online platform Prolific (www.prolific.co). We included only participants with Approval Rates greater than 95% and German as their first language and nationality. Furthermore, participants' current country of residence and country of birth had to be Germany. All participants confirmed that they were fluent in German. Participants completed the study online at a place and time of their choosing and were compensated with a payment of £3.75 (equals an hourly rate of £7.50). We used the same planned exclusion criteria as in Experiment 1 and excluded participants who had engaged in other tasks during the experiment (n = 1), participants who had copied words (none in this experiment), participants with zero recall (none in this experiment), and participants with perfect recall (n = 11). This resulted in a sample size of 52 participants (13 female, 39 male) with a mean age of 31.50 years (SD = 11.58). A post-hoc power analysis confirmed that this sample size provided us with a power of 1 - β > .94 to detect medium sized effects.

Materials

Because memory performance was relatively low in Experiment 1, we used 56 weakly associated word pairs from the Noun Associations for German database (Melinger & Weber, 2006). Word pairs consisted of German 3-8 letter nouns and had a mean association of 0.05 (*SD* = 0.01). Words were of neutral valence (*M* = 0.36, *SD* = 1.03), moderate arousal (*M* = 2.45, *SD* = 0.74), and high imageability (*M* = 5.69, *SD* = 0.80), with word characteristics being taken from Võ et al. (2009) or Lahl et al. (2009). Four additional weakly associated word pairs served as primacy buffers and were not included in the analysis.

Procedure

The procedure was identical to that of the no-warning group of Experiment 1 with the exception that, prior to the JOL recollection phase, participants were told that they would gain points for correct JOL recollections. Instructions explained that participants would gain 1 point for correct RJOLs for some items and 10 points for correct RJOLs for the remaining items. Furthermore, they were told that the five participants who scored the most points would get a bonus payment where 1 p (£0.01) corresponded to each point. This amounted to a maximum bonus of £3.30 (30 * £0.01 + 30 * £0.10). In the JOL recollection phase, each item's value was displayed below the word pair. Value was randomly assigned to items, with the constraint that half the items with and without outcome knowledge were assigned high and low values.

Results and Discussion

Memory performance was analyzed using a 2 (Outcome knowledge: yes, no) $\times$ 2 (Value: high, low) within-subjects ANOVA. Replicating Experiment 1, a main effect of outcome knowledge showed lower memory performance in the first than in the second half of the test, $F(1, 51) = 15.15$, $MSE = 0.02$, $p < .001$, $\eta_p^2 = .23$ (see Table 2). The main effect of value and the interaction were not significant, *F*s < 1.

To test whether incentives influenced hindsight bias on JOLs, we used a 2 (JOL type: OJOL, RJOL) $\times$ 2 (Recall success: yes, no) $\times$ 2 (Outcome knowledge: yes, no) $\times$ 2 (Value: high, low) within-subjects ANOVA. As can be seen in Figure 2, a main effect of recall success indicated higher JOLs for recalled than for not-recalled items, $F(1, 51) = 58.61$, $MSE = 138.06$, $p < .001$, $\eta_p^2 = .54$. The interaction between JOL type and outcome knowledge was significant, indicating a larger difference between OJOLs and RJOLs for items without than with outcome knowledge, $F(1, 51) = 6.89$, $MSE = 47.98$, $p = .011$, $\eta_p^2 = .12$. Furthermore, an interaction between recall success and outcome knowledge showed a larger difference in JOLs between recalled and not-recalled items for word pairs with than without outcome knowledge, $F(1, 51) = 8.31$, $MSE = 144.16$, $p = .006$, $\eta_p^2 = .14$. More importantly, an interaction between JOL type and recall success showed hindsight bias on JOLs, $F(1, 51) = 9.55$, $MSE = 51.37$, $p = .003$, $\eta_p^2 = .16$, with an interaction among JOL type, recall success, and outcome knowledge suggesting that it was limited to the items with outcome knowledge, $F(1, 51) = 8.21$, $MSE = 42.06$, $p = .006$, $\eta_p^2 = .14$. Most importantly, there was no main effect of value, nor any interactions involving

Table 2

Means (and Standard Deviations) of Recall, Gamma Correlations Between JOLs and Recall, ΔHB, and Correct JOL Recollections for Experiment 2

	Outcome knowledge		No outcome knowledge	
	High value	Low value	High value	Low value
p(Recall)	.58 (.19)	.58 (.21)	.65 (.22)	.66 (.21)
G(OJOL, Recall)	.18 (.44)	.17 (.37)	.10 (.48)	.16 (.43)
G(RJOL, Recall)	.47 (.37)	.39 (.39)	.18 (.51)	.16 (.44)
ΔHB	2.37**	2.90***	0.51	-0.80
p(RJOL = OJOL)	.25 (.16)	.25 (.17)	.28 (.18)	.29 (.16)

Note. *p*(Recall) = percentage of correctly recalled items; OJOL = original judgment of learning; RJOL = recollected judgment of learning; ΔHB = Pohl's (1992) difference measure of hindsight bias; *p*(RJOL = OJOL) = percentage of correct JOL recollections.
** $p < .01$, *** $p < .001$.

value, all Fs < 1.69, all ps > .199. This finding indicated that incentivizing correct JOL recollections did not affect hindsight bias on JOLs. No other effects were significant, all Fs < 1.

As in Experiment 1, we ran a Bayesian ANOVA to test whether there was positive evidence for the hypothesis that hindsight bias on JOLs was comparable for high- and low-value items. We again tested the full model against a model without the four-way interaction (JOL type × Recall success × Outcome knowledge × Value) that is indicative of effects of incentives on hindsight bias. A BF of 0.239 provided moderate evidence for the hypothesis that incentives did not influence hindsight bias on JOLs (i.e., the data are 4.18 times more likely under the model without the four-way interaction than under the full model).

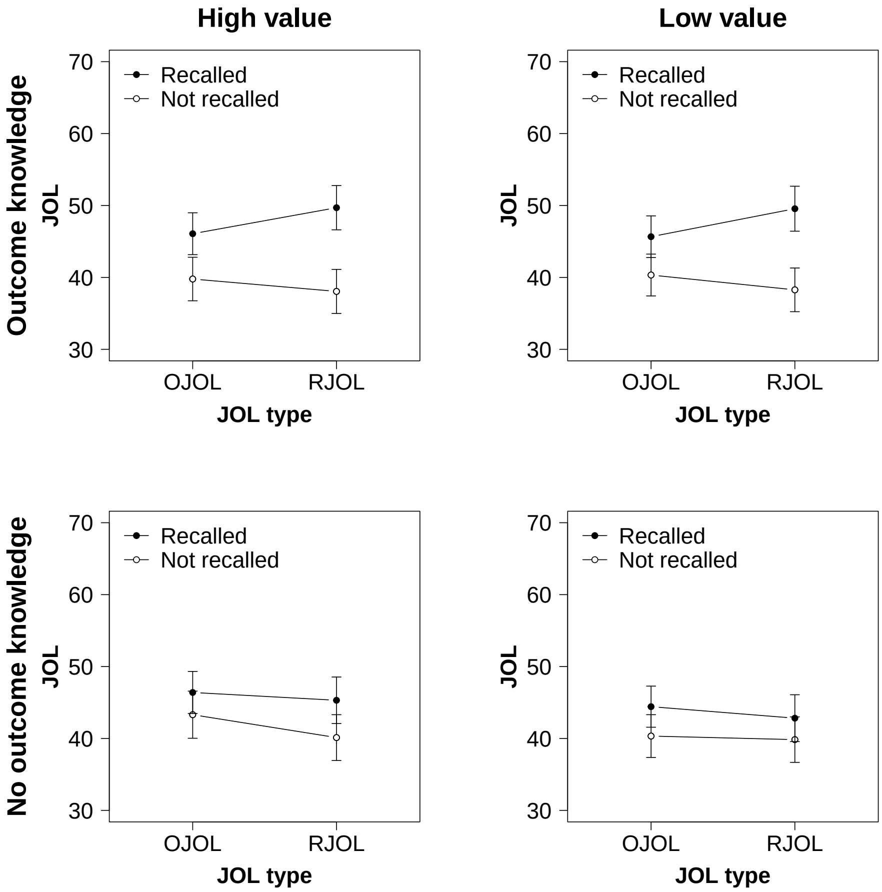
Replicating Experiment 1, for high- and low-value items with outcome knowledge, resolution of RJOLs was higher than that of OJOLs, high value: $t(51) = 4.78$, $p < .001$, $d = 0.71$; low value: $t(51) = 3.89$, $p < .001$, $d = 0.58$ (see Table 2). Importantly, the increase in resolution from OJOLs to RJOLs did not differ between high- and low-value items, $t(51) = 0.88$, $p = .384$, $d = 0.16$, BF = 0.218, indicating that hindsight bias on JOLs was unaffected by incentivizing participants for correct JOL recollections. For items without outcome knowledge, resolution of RJOLs did not differ from that of OJOLs for high- and low-value items, high value: $t(51) = 1.36$, $p = .179$, $d = 0.15$; low value: $t(51) = 0.05$, $p = .962$, $d < 0.01$.

ΔHB revealed hindsight bias for high- and low-value items with outcome knowledge, high value: $t(51) = 3.12$, $p = .003$, $d = 0.43$; low value: $t(51) = 3.56$, $p < .001$, $d = 0.49$ (see Table 2), that did not differ across high- and low-value items, $t(51) = 0.60$, $p = .552$, $d = 0.09$, BF = 0.179. For high- and low-value items without outcome knowledge, ΔHB was not significantly different from zero, high value: $t(51) = 0.73$, $p = .469$, $d = 0.10$; low value: $t(51) = 0.99$, $p = .327$, $d = 0.14$.

As can be seen in Table 2, proportions of correct JOL recollections did not differ between high- and low-value items, with outcome knowledge: $t(51) = 0.19$, $p = .847$, $d = 0.03$, BF = 0.154; without outcome knowledge: $t(51) = 0.39$,

Figure 2

Mean Original Judgments of Learning (OJOLs) and Recollected Judgments of Learning (RJOLs) in Experiment 2



Note. Top panels present items with outcome knowledge and bottom panels present items without outcome knowledge that were assigned a high value (left panels) or a low value (right panels) in Experiment 2. Filled symbols represent recalled items and open symbols represent not-recalled items. Error bars represent one standard error of the mean.

p = .698, d = 0.06, BF = 0.162, showing that high incentives did not promote correct JOL recollections.

As expected, we found hindsight bias on JOLs for items with outcome knowledge only. More importantly, hindsight bias on JOLs was not reduced for high-value items as compared to low-value items. Experiment 2 therefore showed that hindsight bias on JOLs is robust against monetary incentives. An alternative possibility, however, was that the incentive manipulation was too weak to impact people's motivation to correctly recollect their JOLs. We addressed this possibility in Experiment 3.

Experiment 3

Experiment 3 provided a stronger test of whether incentives reduce hindsight bias on JOLs. To this end, we increased the monetary incentive for correct JOL recollections of high-value items to £0.30 per item. This increased the value difference between high- and low-value items and, at the same time, increased the maximum monetary bonus to £9.30. We also assessed participants' motivation to correctly recollect JOLs they had assigned to high- and low-value items.

Method

Design, Materials, and Procedure

Design and materials were the same as in Experiment 2. The procedure was also the same with the following exceptions. First, the monetary bonus for high-value items was increased to £0.30, resulting in a maximum bonus of £9.30 (30 * £0.01 + 30 * £0.30). Second, at the end of the experiment, we asked participants about their motivation when recollecting their JOLs. Specifically, participants indicated on a 10-point scale whether they were more motivated to recollect JOLs for items with a monetary bonus of *£0.01* (scale point 1) or to recollect JOLs for items with a monetary bonus of *£0.30* (scale point 10).

Participants

Sample size was based on the same power analysis as in Experiment 2. We recruited 75 participants via Prolific applying the same inclusion criteria as in the previous experiment (all participants confirmed that they were fluent in German). Participants were again compensated with a payment of £3.75 and had the chance to receive a monetary bonus of up to £9.30 if they were among the five participants who scored the most points. As in the previous experiments, we excluded participants who had engaged in other tasks during the experiment ($n = 2$), participants who had copied words (none in this experiment), participants with zero recall (none in this experiment), and participants with perfect recall ($n = 21$). This resulted in a sample size of 52 participants (21 female, 30 male, 1 diverse) with a mean age of 31.02 years ($SD = 10.85$).

Results and Discussion

Participants were more motivated to recollect their JOLs assigned to high- than to low-value items. The mean rating was 7.48 ($SD = 2.05$) and significantly higher than the middle of the scale, $t(51) = 6.96$, $p < .001$, $d = 0.97$. Forty-eight participants (92.31%) gave a rating of 6 or higher, binomial test: $p < .001$.

With regard to memory performance, a 2 (Outcome knowledge: yes, no) × 2 (Value: high, low) within-subjects ANOVA revealed a main effect of outcome knowledge, showing that memory performance was lower in the first than in the second half of the cued recall test, $F(1, 51) = 26.10$, $MSE = 0.02$, $p < .001$, $\eta_p^2 = .34$ (see Table 3). The main effect of value and the interaction were not significant, Fs < 1.

Table 3
Means (and Standard Deviations) of Recall, Gamma Correlations Between JOLs and Recall, ΔHB, and Correct JOL Recollections for Experiment 3

	Outcome knowledge		No outcome knowledge	
	High value	Low value	High value	Low value
p(Recall)	.63 (.19)	.62 (.21)	.70 (.17)	.72 (.17)
G(OJOL, Recall)	.09 (.48)	.21 (.46)	.08 (.49)	.12 (.50)
G(RJOL, Recall)	.28 (.50)	.41 (.40)	.21 (.48)	.13 (.49)
ΔHB	2.94**	1.87*	0.46	-0.16
p(RJOL = OJOL)	.28 (.18)	.26 (.14)	.27 (.18)	.28 (.18)

Note. *p*(Recall) = percentage of correctly recalled items; OJOL = original judgment of learning; RJOL = recollected judgment of learning; ΔHB = Pohl's (1992) difference measure of hindsight bias; *p*(RJOL = OJOL) = percentage of correct JOL recollections.
* $p < .05$, ** $p < .01$.

A 2 (JOL type: OJOL, RJOL) $\times$ 2 (Recall success: yes, no) $\times$ 2 (Outcome knowledge: yes, no) $\times$ 2 (Value: high, low) within-subjects ANOVA revealed a main effect of recall success with higher JOLs for recalled than for not-recalled items, $F(1, 51) = 38.59$, $MSE = 294.08$, $p < .001$, $\eta_p^2 = .43$ (see Figure 3). A significant interaction between JOL type and outcome knowledge indicated that the difference between OJOLs and RJOLs was larger for items without than with outcome knowledge, $F(1, 51) = 16.01$, $MSE = 55.92$, $p < .001$, $\eta_p^2 = .24$. An interaction between JOL type and recall success showed hindsight bias on JOLs, $F(1, 51) = 6.44$, $MSE = 42.77$, $p = .014$, $\eta_p^2 = .11$. Unlike in the previous experiments and unexpectedly, the interaction among JOL type, recall success, and outcome knowledge was not significant, $F(1, 51) = 2.57$, $MSE = 41.71$, $p = .115$, $\eta_p^2 = .05$. Most importantly, there was no main effect of value, nor any interactions involving value, all *F*s < 1.98, all *p*s > .165, indicating that incentivizing correct JOL recollections did not have any effect on hindsight bias on JOLs. No other effects were significant, all *F*s < 3.21, all *p*s > .079.

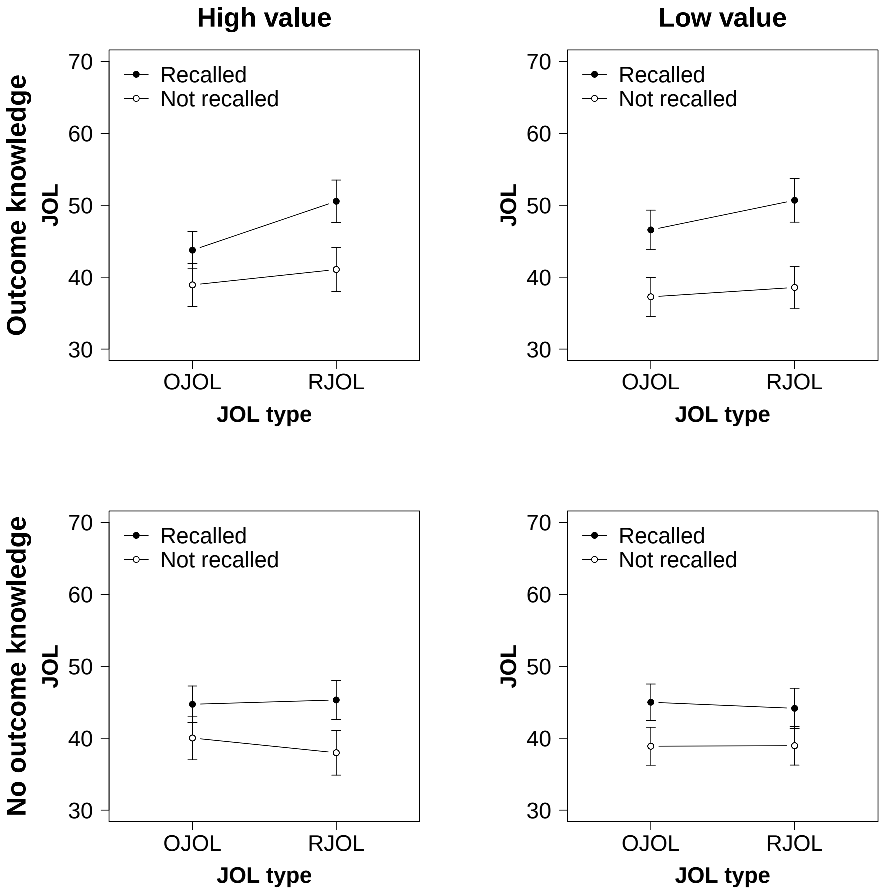
A Bayesian ANOVA provided moderate evidence for the hypothesis that incentives did not influence hindsight bias on JOLs (BF = 0.176; data are 5.68 times more likely under the model without the four-way interaction than under the full model).

Replicating the two previous experiments, for high- and low-value items with outcome knowledge, resolution of RJOLs was higher than that of OJOLs, high value: $t(51) = 2.88$, $p = .006$, $d = 0.39$; low value: $t(51) = 2.92$, $p = .005$, $d = 0.47$ (see Table 3). Importantly, the increase in resolution from OJOLs to RJOLs did not differ between high- and low-value items, $t(51) = 0.14$, $p = .892$, $d = 0.02$, BF = 0.152, again indicating that hindsight bias on JOLs was unaffected by incentivizing participants for correct JOL recollections. Unexpectedly, for items without outcome knowledge, resolution of RJOLs was higher than of OJOLs for high-value items, $t(51) = 2.16$, $p = .036$, $d = 0.28$; but not for low-value items, $t(51) = 0.17$, $p = .867$, $d = 0.03$.

Again, ΔHB revealed hindsight bias for high- and low-value items with outcome knowledge, high value: $t(51) = 3.44$, $p = .001$, $d = 0.48$; low value: $t(51) = 2.01$, $p = .049$, $d = 0.28$ (see Table 3), with no differences between high- and

Figure 3

Mean Original Judgments of Learning (OJOLs) and Recollected Judgments of Learning (RJOLs) in Experiment 3



Note. Top panels present items with outcome knowledge and bottom panels present items without outcome knowledge that were assigned a high value (left panels) or a low value (right panels) in Experiment 3. Filled symbols represent recalled items and open symbols represent not-recalled items. Error bars represent one standard error of the mean.

low-value items, $t(51) = 1.02$, $p = .315$, $d = 0.17$, BF = 0.246. As in the previous experiments, for high- and low-value items without outcome knowledge, ΔHB was not significantly different from zero, high value: $t(51) = 0.67$, $p = .508$, $d = 0.09$; low value: $t(51) = 0.21$, $p = .834$, $d = 0.03$.

Replicating Experiment 2, proportions of correct JOL recollections did not differ between high- and low-value items, with outcome knowledge: $t(51) = 0.90$, $p = .371$, $d = 0.11$, BF = 0.222; without outcome knowledge: $t(51) = 0.57$, $p = .569$, $d = 0.07$, BF = 0.177, showing that high incentives did not promote correct JOL recollections (see Table 3).

Overall, despite large differences in monetary incentives between high- and low-value items and despite higher participant motivation to correctly recollect the JOLs assigned to high- than to low-value items, Experiment 3 replicated Experiment 2 in showing that incentives do not reduce hindsight bias on JOLs.

General Discussion

The current study demonstrates the robustness of hindsight bias on JOLs by showing that neither warning participants (Experiment 1) nor incentivizing correct JOL recollections (Experiments 2 and 3) reduced this bias. As in previous research, biased recollection of JOLs due to outcome knowledge was apparent in three different measures (see Zimdahl & Undorf, 2021). First, there was a larger JOL difference between recalled and not-recalled word pairs in RJOLs than in OJOLs, showing that participants remembered their OJOLs as being closer to the outcome than they actually were. Second, resolution of RJOLs was higher than resolution of OJOLs. Third, we obtained a significantly positive hindsight bias measure ΔHB (Pohl, 1992).

Neither warning participants about hindsight bias nor incentivizing correct JOL recollections reduced hindsight bias on JOLs. Across three experiments, we found no main effects or interactions involving warnings or incentives indicating that either manipulation affected hindsight bias on JOLs. What is more, Bayesian analyses provided positive evidence for the hypothesis that hindsight bias on JOLs was equal across the warning and no-warning groups (Experiment 1) and across high- and low-value items (Experiments 2 and 3). The finding that neither warnings nor higher incentives produced more correct recollections of JOLs provided further evidence against reduced hindsight bias on JOLs.

Theoretical Implications

This study extends prior research on hindsight bias in metamemory (Zimdahl & Undorf, 2021) by demonstrating its robustness against two different debiasing methods. Hindsight biases affecting judgments about the external world are robust phenomena that resisted extensive debiasing attempts (e.g., Bernstein et al., 2016; Blank et al., 2007; Hell et al., 1988; Pohl & Erdfelder, 2022; Roese & Vohs, 2012). In contrast, previous research on reducing or eliminating biases that afflict JOLs with warnings and incentives has produced mixed results (e.g., Carpenter & Olson, 2012; Koriat & Bjork, 2006; Koriat et al., 2006; Rhodes & Castel, 2008; Yan et al., 2016). The current study demonstrates that hindsight bias on JOLs is resistant against warnings and incentives, as are hindsight biases on judgments about the external world.

The current study sheds light on the processes that may contribute to hindsight bias on JOLs. In particular, the findings that neither warnings nor incentives alleviate this bias underline the role of cognitive processes. For example, they are compatible with the idea of an *immediate assimilation* (Fischhoff, 1975, 1977) assuming that outcome knowledge is automatically integrated into people's knowledge base (see also Bernstein et al., 2012; Harley et al., 2004; Pohl & Hell, 1996). According to this account, hindsight bias is due to people accessing their updated knowledge base (e.g., Hoffrage et al., 2000). Notably, however, the immediate assimilation account is inconsistent with other facets of hindsight bias (see, e.g., Bernstein et al., 2016; Pohl & Hell, 1996).

Our results are also consistent with the *trace-strength hypothesis* (Hell et al., 1988) assuming a coexistence of memory traces of one's original judgment and outcome knowledge, with the size of hindsight bias depending on the relative strengths of these two memory traces. According to this view, correct recollections stem from strong memory traces of original judgments that leave little room for effects of outcome knowledge on recollected judgments, whereas weak memory traces of original judgments promote hindsight bias. This hypothesis can explain why, in the current study, people correctly recollected some of their JOLs (between 25% and 36% across experiments) but exhibited hindsight bias on JOLs for other items.

Finally, our results are also consistent with the *anchoring and adjustment approach* (Tversky & Kahneman, 1974). This approach assumes that people use outcome knowledge as an anchor to reconstruct their original judgments when they are unable to recollect them (see also Pohl, Eisenhauer, et al., 2003; Schwarz & Stahlberg, 2003; Stahlberg & Maass, 1997). Hindsight bias then results from insufficient adjustment from this anchor (cf. Epley & Gilovich, 2006). According to this account, learning that one has correctly recalled an item at test may serve as a high anchor for one's RJOLs, with insufficient downward adjustment resulting in higher RJOLs than OJOLs. Conversely, learning that one has failed to recall an item may serve as a low anchor, with insufficient upward adjustment leading to lower RJOLs than OJOLs. Consistent with this approach, several studies have shown that anchors can affect JOLs (e.g., England & Serra, 2012; Scheck & Nelson, 2005; Yang et al., 2018). Further research is needed to investigate which of these cognitive processes contribute to hindsight bias on JOLs.

In contrast, the current study does not provide support for accounts that attribute hindsight bias to motivational processes. If hindsight bias on JOLs was due to motivational processes, then it should have been reduced by warnings in Experiment 1 and by monetary incentives in Experiments 2 and 3. The results of Experiment 3 are most telling in this regard: Hindsight bias on JOLs was similar for high- and low-value items although participants reported that they had been more motivated to correctly recollect their JOLs for high- than for low-value items. This finding demonstrates that hindsight bias on JOLs is not due to a lack of motivation to correctly recollect original judgments.

Practical Implications

It is an open question whether debiasing hindsight bias on JOLs would have beneficial effects on learning. As described elsewhere (Zimdahl & Undorf, 2021), hindsight bias on JOLs could foster or, alternatively, hinder learning. On the one hand, hindsight bias on JOLs can be regarded as a by-product of learning (e.g., Hoffrage et al., 2000). From this perspective, updating one's knowledge by replacing erroneous memory contents (original judgments) with correct information (outcome knowledge) is an adaptive mechanism that allows people to make better-informed and more accurate JOLs. If so, debiasing hindsight bias would hinder learning and might impair the accuracy of future JOLs.

On the other hand, hindsight bias on JOLs could be detrimental for learning. Because of hindsight bias, people might fail to recognize that their original JOLs were incorrect. If so, hindsight bias would deprive learners of the opportunity to learn from their mistakes and encourage future misjudgments. In this case, debiasing hindsight bias would be critical for improving the accuracy of peoples' JOLs and, more generally, their metacognitive abilities (see Arkes, 2013, for discussion). Understanding which view is more appropriate for hindsight bias in general and for hindsight bias on JOLs in particular is an important avenue for future research (see also Ackerman et al., 2020; Bernstein et al., 2016).

Limitations and Future Directions

One possible concern could be that the incentive manipulations in the current experiments were too weak to increase people's motivation. We consider this possibility unlikely for three reasons. First, the fact that prior studies found profound differences in JOLs for high- and low-value items when incentives for high-value items were 3 and 5 times higher than those for low-value items (Koriat et al., 2014; Koriat et al., 2006) suggests that increasing incentives by factors 10 (Experiment 2) or 30 (Experiment 3) should be effective. Second, the maximum monetary bonus was considerably higher than people's regular compensation (Experiment 2: up to 188%, Experiment 3: up to 348%). Finally and most importantly, participants reported in Experiment 3 that they were more motivated to correctly recollect JOLs for high- than for low-value items. Overall, this suggests that high incentives increased participants' motivation in the current study.

The current study does not exclude the possibility that motivational processes contribute to hindsight bias on JOLs under different conditions. Roese and Vohs (2012) argue that motivational contributions to hindsight bias serve the function to protect or enhance one's self-esteem (see also Pezzo & Pezzo, 2007). Maybe, increasing the self-relevance of correctly recollecting JOLs boosts contributions of motivational processes to hindsight bias on JOLs. If so, having participants publicly announce a to-be-achieved level of metacognitive accuracy or making the test critical for a personalized promotion or demotion should increase motivational contributions to hindsight bias on JOLs.

Finally, despite the current findings, it may be possible to alleviate hindsight bias on JOLs. There is evidence to suggest that making participants think about other potential outcomes (*counterfactual reasoning* or *consider-the-opposite strategy*, Lord et al., 1984; Pohl & Hell, 1996) reduces hindsight biases, presumably because people appreciate why other outcomes might have been possible as well (see also Hoch, 1985; Koriat et al., 1980; Nario & Branscombe, 1995; Slovic & Fischhoff, 1977). Future research might explore whether encouraging participants to think about different outcomes in the memory test alleviates hindsight bias on JOLs.

Conclusion

The current study aimed at investigating the robustness of hindsight bias on JOLs. Results showed that warnings and incentives for correct JOL recollections did not alleviate hindsight bias on JOLs. These findings revealed that hindsight bias on JOLs is a robust phenomenon, as are hindsight biases on judgments about the external world.

References

Ackerman, R., Bernstein, D. M., & Kumar, R. (2020). Metacognitive hindsight bias. *Memory & Cognition*, *48*(5), 731–744. https://doi.org/10.3758/s13421-020-01012-w

Arkes, H. R. (2013). The consequences of the hindsight bias in medical decision making. *Current Directions in Psychological Science*, *22*(5), 356–360. https://doi.org/10.1177/0963721413489988

Arkes, H. R., Wortmann, R. L., Saville, P. D., & Harkness, A. R. (1981). Hindsight bias among physicians weighing the likelihood of diagnoses. *The Journal of Applied Psychology*, *66*(2), 252–254. https://doi.org/10.1037/0021-9010.66.2.252

Asendorpf, J. B., Conner, M., de Fruyt, F., de Houwer, J., Denissen, J. J. A., Fiedler, K., Fiedler, S., Funder, D. C., Kliegl, R., Nosek, B. A., Perugini, M., Roberts, B. W., Schmitt, M., van Aken, M. A. G., Weber, H., & Wicherts, J. M. (2013). Recommendations for increasing replicability in psychology. *European Journal of Personality*, *27*(2), 108–119. https://doi.org/10.1002/per.1919

Bayen, U. J., Pohl, R. F., Erdfelder, E., & Auer, T.-.-S. (2007). Hindsight bias across the life span. *Social Cognition*, *25*, 83–97. https://doi.org/10.1521/soco.2007.25.1.83

Bernstein, D. M., Aßfalg, A., Kumar, R., & Ackerman, R. (2016). Looking backward and forward on hindsight bias. In J. Dunlosky & S. U. K. Tauber (Eds.), *The Oxford handbook of metamemory* (pp. 289–304). Oxford University Press.

Bernstein, D. M., Erdfelder, E., Meltzoff, A. N., Peria, W., & Loftus, G. R. (2011). Hindsight bias from 3 to 95 years of age. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 378–391. https://doi.org/10.1037/a0021971

Bernstein, D. M., Wilson, A. M., Pernat, N. L. M., & Meilleur, L. R. (2012). Auditory hindsight bias. *Psychonomic Bulletin & Review*, *19*, 588–593. https://doi.org/10.3758/s13423-012-0268-0

Blank, H., Musch, J., & Pohl, R. F. (2007). Hindsight bias: On being wise after the event. *Social Cognition*, *25*, 1–9. https://doi.org/10.1521/soco.2007.25.1.1

Blank, H., Nestler, S., von Collani, G., & Fischer, V. (2008). How many hindsight biases are there? *Cognition*, *106*, 1408–1440. https://doi.org/10.1016/j.cognition.2007.07.007

Camerer, C., Loewenstein, G., & Weber, M. (1989). The curse of knowledge in economic settings: An experimental analysis. *Journal of Political Economy*, *97*(5), 1232–1254. https://doi.org/10.1086/261651

Carpenter, S. K., & Olson, K. M. (2012). Are pictures good for learning new vocabulary in a foreign language? Only if you think they are not. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(1), 92–101. https://doi.org/10.1037/a0024828

England, B. D., & Serra, M. J. (2012). The contributions of anchoring and past-test performance to the underconfidence-with-practice effect. *Psychonomic Bulletin & Review*, *19*, 715–722. https://doi.org/10.3758/s13423-012-0237-7

Epley, N., & Gilovich, T. (2006). The anchoring-and-adjustment heuristic: Why the adjustments are insufficient. *Psychological Science*, *17*(4), 311–318. https://doi.org/10.1111/j.1467-9280.2006.01704.x

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/bf03193146

Fischhoff, B. (1975). Hindsight is not equal to foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, *1*, 288–299. https://doi.org/10.1037/0096-1523.1.3.288

Fischhoff, B. (1977). Perceived informativeness of facts. *Journal of Experimental Psychology: Human Perception and Performance*, *3*, 349–358. https://doi.org/10.1037/0096-1523.3.2.349

Formann, A. K., & Piswanger, K. (1979). *Wiener Matrizen-Test [Vienna Matrices Test]: Ein Rasch-skalierter sprachfreier Intelligenztest*. Beltz.

Giroux, M. E., Coburn, P. I., Harley, E. M., Connolly, D. A., & Bernstein, D. M. (2016). Hindsight bias and law. *Zeitschrift für Psychologie*, *224*, 190–203. https://doi.org/10.1027/2151-2604/a000253

Goodman, L. A., & Kruskal, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association*, *49*(268), 732. https://doi.org/10.2307/2281536

Harley, E. M., Carlsen, K. A., & Loftus, G. R. (2004). The "saw-it-all-along" effect: Demonstrations of visual hindsight bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 960–968. https://doi.org/10.1037/0278-7393.30.5.960

Hastie, R., Schkade, D. A., & Payne, J. W. (1999). Juror judgments in civil cases: Hindsight effects on judgments of liability for punitive damages. *Law and Human Behavior*, *23*(5), 597–614. https://doi.org/10.1023/A:1022352330466

Hawkins, S. A., & Hastie, R. (1990). Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin*, *107*, 311–327. https://doi.org/10.1037/0033-2909.107.3.311

Hell, W., Gigerenzer, G., Gauggel, S., Mall, M., & Müller, M. (1988). Hindsight bias: An interaction of automatic and motivational factors? *Memory & Cognition*, *16*, 533–538. https://doi.org/10.3758/BF03197054

Hoch, S. J. (1985). Counterfactual reasoning and accuracy in predicting personal events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(4), 719–731. https://doi.org/10.1037/0278-7393.11.1-4.719

Hoffrage, U., Hertwig, R., & Gigerenzer, G. (2000). Hindsight bias: A by-product of knowledge updating? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 566–581. https://doi.org/10.1037/0278-7393.26.3.566

Koriat, A., Ackerman, R., Adiv, S., Lockl, K., & Schneider, W. (2014). The effects of goal-driven and data-driven regulation on metacognitive monitoring during learning: A developmental perspective. *Journal of Experimental Psychology: General*, *143*(1), 386–403. https://doi.org/10.1037/a0031768

Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one's knowledge during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(2), 187–194. https://doi.org/10.1037/0278-7393.31.2.187

Koriat, A., & Bjork, R. A. (2006). Mending metacognitive illusions: A comparison of mnemonic-based and theory-based procedures. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(5), 1133–1145. https://doi.org/10.1037/0278-7393.32.5.1133

Koriat, A., Bjork, R. A., Sheffer, L., & Bar, S. K. (2004). Predicting one's own forgetting: The role of experience-based and theory-based processes. *Journal of Experimental Psychology: General*, *133*(4), 643–656. https://doi.org/10.1037/0096-3445.133.4.643

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 107–118. https://doi.org/10.1037/0278-7393.6.2.107

Koriat, A., Ma'ayan, H., & Nussinson, R. (2006). The intricate relationships between monitoring and control in metacognition: Lessons for the cause-and-effect relation between subjective experience and behavior. *Journal of Experimental Psychology: General*, *135*(1), 36–69. https://doi.org/10.1037/0096-3445.135.1.36

Lahl, O., Göritz, A. S., Pietrowsky, R., & Rosenberg, J. (2009). Using the world-wide web to obtain large-scale word norms: 190,212 ratings on a set of 2,654 german nouns. *Behavior Research Methods*, *41*(1), 13–19. https://doi.org/10.3758/BRM.41.1.13

Lord, C. G., Lepper, M. R., & Preston, E. (1984). Considering the opposite: A corrective strategy for social judgment. *Journal of Personality and Social Psychology*, *47*(6), 1231–1243. https://doi.org/10.1037/0022-3514.47.6.1231

Louie, T. A. (1999). Decision makers' hindsight bias after receiving favorable and unfavorable feedback. *The Journal of Applied Psychology*, *84*(1), 29–41. https://doi.org/10.1037/0021-9010.84.1.29

Louie, T. A., Rajan, M. N., & Sibley, R. E. (2007). Tackling the monday-morning quarterback: Applications of hindsight bias in decision-making settings. *Social Cognition*, *25*(1), 32–47. https://doi.org/10.1521/soco.2007.25.1.32

Melinger, A., & Weber, A. (2006). Database of noun associations for german. http://www.coli.uni-saarland.de/projects/nag/

Morey, R. D., & Rouder, J. N. (2015). BayesFactor: Computation of Bayes factors for common designs. https://CRAN.R-project.org/package=BayesFactor

Nario, M. R., & Branscombe, N. R. (1995). Comparison processes in hindsight and causal attribution. *Personality and Social Psychology Bulletin*, *21*(12), 1244–1255. https://doi.org/10.1177/01461672952112001

Pezzo, M. V. (2011). Hindsight bias: A primer for motivational researchers. *Social and Personality Psychology Compass*, *5*(9), 665–678. https://doi.org/10.1111/j.1751-9004.2011.00381.x

Pezzo, M. V., & Pezzo, S. P. (2007). Making sense of failure: A motivated model of hindsight bias. *Social Cognition*, *25*(1), 147–164. https://doi.org/10.1521/soco.2007.25.1.147

Pohl, R. F. (1992). Der Rückschau-Fehler: Systematische Verfälschung der Erinnerung bei Experten und Novizen. *Kognitionswissenschaft*, *3*, 38–44.

Pohl, R. F. (2007). Ways to assess hindsight bias. *Social Cognition*, *25*, 14–31. https://doi.org/10.1521/soco.2007.25.1.14

Pohl, R. F., Bender, M., & Lachmann, G. (2002). Hindsight bias around the world. *Experimental Psychology (formerly Zeitschrift für Experimentelle Psychologie)*, *49*, 270–282. https://doi.org/10.1026/1618-3169.49.4.270

Pohl, R. F., Eisenhauer, M., & Hardt, O. (2003). SARA: A cognitive process model to simulate the anchoring effect and hindsight bias. *Memory*, *11*, 337–356. https://doi.org/10.1080/09658210244000487

Pohl, R. F., & Erdfelder, E. (2022). Hindsight bias. In R. F. Pohl (Ed.), *Cognitive illusions: Intriguing phenomena in thinking, judgment and memory (Third Edition)* (pp. 436–454). Routledge. https://doi.org/10.4324/9781003154730-31

Pohl, R. F., & Hell, W. (1996). No reduction in hindsight bias after complete information and repeated testing. *Organizational Behavior and Human Decision Processes*, *67*, 49–58. https://doi.org/10.1006/obhd.1996.0064

Pohl, R. F., Schwarz, S., Sczesny, S., & Stahlberg, D. (2003). Hindsight bias in gustatory judgments. *Experimental Psychology (formerly Zeitschrift für Experimentelle Psychologie)*, *50*, 107–115. https://doi.org/10.1026//1618-3169.50.2.107

Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology: General*, *137*(4), 615–625. https://doi.org/10.1037/a0013684

Roese, N. J., & Vohs, K. D. (2012). Hindsight bias. *Perspectives on Psychological Science*, *7*(5), 411–426. https://doi.org/10.1177/1745691612454303

Rouder, J. N., Morey, R. D., Verhagen, J., Swagman, A. R., & Wagenmakers, E.-J. (2017). Bayesian analysis of factorial designs. *Psychological Methods*, *22*(2), 304–321. https://doi.org/10.1037/met0000057

Scheck, P., & Nelson, T. O. (2005). Lack of pervasiveness of the underconfidence-with-practice effect: boundary conditions and an explanation via anchoring. *Journal of Experimental Psychology: General*, *134*(1), 124–128. https://doi.org/10.1037/0096-3445.134.1.124

Schwarz, S., & Stahlberg, D. (2003). Strength of hindsight bias as a consequence of meta-cognitions. *Memory*, *11*, 395–410. https://doi.org/10.1080/09658210244000496

Slovic, P., & Fischhoff, B. (1977). On the psychology of experimental surprises. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(4), 544–551. https://doi.org/10.1037/0096-1523.3.4.544

Stahlberg, D., & Maass, A. (1997). Hindsight bias: Impaired memory or biased reconstruction? *European Review of Social Psychology*, *8*, 105–132. https://doi.org/10.1080/14792779643000092

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

Undorf, M., Navarro-Báez, S., & Zimdahl, M. F. (2022). Metacognitive illusions. In R. F. Pohl (Ed.), *Cognitive illusions: Intriguing phenomena in thinking, judgment and memory (Third Edition)* (pp. 307–323). Routledge. https://doi.org/10.4324/9781003154730-22

Võ, M. L.-H., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M. J., & Jacobs, A. M. (2009). The Berlin Affective Word List Reloaded (BAWL-R). *Behavior Research Methods*, *41*(2), 534–538. https://doi.org/10.3758/BRM.41.2.534

Wood, G. (1978). The knew-it-all-along effect. *Journal of Experimental Psychology: Human Perception and Performance*, *4*, 345–353. https://doi.org/10.1037/0096-1523.4.2.345

Yan, V. X., Bjork, E. L., & Bjork, R. A. (2016). On the difficulty of mending metacognitive illusions: A priori theories, fluency effects, and misattributions of the interleaving benefit. *Journal of Experimental Psychology: General*, *145*(7), 918–933. https://doi.org/10.1037/xge0000177

Yang, C., Sun, B., & Shanks, D. R. (2018). The anchoring effect in metamemory monitoring. *Memory & Cognition*, *46*(3), 384–397. https://doi.org/10.3758/s13421-017-0772-6

Zimdahl, M. F., & Undorf, M. (2021). Hindsight bias in metamemory: Outcome knowledge influences the recollection of judgments of learning. *Memory*, *29*(5), 559–572. https://doi.org/10.1080/09658211.2021.1919144

Appendix

Participants in the warning group of Experiment 1 received a detailed description of hindsight bias on JOLs before recollecting their JOLs. Instructions read (translated from German):

> WARNING!
>
> In a similar experiment, many participants made errors when recollecting their original judgments. Those participants' recollections were unduly influenced by the outcome of the memory test. This means that they
>
> - recollected higher judgments for the items they had recalled at test
> - recollected lower judgments for the items that they had not recalled at test
>
> This bias is known as *hindsight bias*. When recollecting your judgments, please make sure that the memory test's outcome does NOT bias your recollections and that you do NOT exhibit hindsight bias.

The following questions tested participants' understanding of the warning:

1. People who show hindsight bias are influenced by
   - whether they recalled the item in the memory test
   - whether they recollected their initial probability of remembering
2. For items that were recalled at test, people who show hindsight bias usually recollect their initial judgments
   - as higher than they were
   - as lower than they were
3. For items that were NOT recalled at test, people who show hindsight bias usually recollect their initial judgments
   - as higher than they were
   - as lower than they were

This dissertation thesis was supported by the *Graduate School of Economic and Social Sciences* (GESS) at the University of Mannheim.

Cover image designed by jackie_niam / Freepik.