ANNALS OF PSYCHOLOGY/ROCZNIKI PSYCHOLOGICZNE
2024, XXVII, 4, 299–329
DOI: https://doi.org/10.18290/rpsych2024.0017

# IS METAMEMORY JUDGEMENTS REACTIVITY A TOPIC WORTH INVESTIGATING? AN OVERVIEW OF LITERATURE AND FUTURE DIRECTIONS

Ngoc Diep Le

Institute of Applied Psychology, Jagiellonian University

Judgements of learning (JOLs) are commonly used in metacognitive research to assess a person's ability to monitor their learning. However, despite the widespread use of this type of self-reported measure, only recently have metacognitive researchers become interested in investigating its potential reactive effects in a proper empirical manner, via directly addressing the question whether the sole act of systematically monitoring memory influences memory performance. The issue of possible reactive JOL effects has been raised since the early days of metamemory monitoring research, but has taken a completely new direction in the lately emerging studies—from issuing warnings about JOL reactivity as a potential limitation, to dedicating entire series of experiments to uncover the principles behind certain reactivity patterns, and its potential moderators. More research is needed on educationally relevant materials to determine whether JOL reactivity can be utilised in educational contexts. Finally, this phenomenon, as shown in the performed literature overview, leads to important theoretical implications reflected in the four hypotheses explaining the reactivity effect.

**Keywords:** metamemory; learning; reactivity; judgements of learning; JOL reactivity

Since the dawn of psychological research, scholars have been using self-reported measures to assess humans' introspection. The underlying presumption was that such measures enable passive evaluation of their subjective processes, without altering the targeted processes themselves (Mitchum et al.,

NGOC DIEP LE, https://orcid.org/0009-0008-0838-9520. Correspondence concerning this article should be addressed to Ngoc Diep Le, Instytut Psychologii Stosowanej, ul. Prof. S. Łojasiewicza 4, 30-348 Kraków, Poland; e-mail: ngocdiep.le@student.uj.edu.pl. I would like to express my gratitude for the invaluable constructive criticism I have received from Aleksandra Krogulska.

Handling editors: WIKTOR RAZMUS, John Paul II Catholic University of Lublin, and MARIOLA ŁAGUNA, John Paul II Catholic University of Lublin. Received 18 April 2024. Received in revised form 30 Oct. 2024, 11 Feb. 2025, 20 March 2025. Accepted 24 March 2025. Published online 14 May 2025.

2016). However, it is currently acknowledged that self-reports may unintentionally modify what they were intended to assess. This phenomenon can be classified as pertaining to a broader issue in psychological research, named reactivity—which occurs any time an individual modifies their behaviour in response to being either observed or measured (Double & Birney, 2019). Although reactivity can be problematic for the accuracy of measurement and drawing theoretical conclusions from such measurements, it does not always inherently have negative effects. In case of cognitive tasks, reactivity can sometimes have positive consequences, improving performance on a given test. This paper aims to address the potential reactive effects of one such self-reported measure used to evaluate one's own learning process, whose reactivity has only recently been properly demonstrated. I will discuss, through a critical lens, the most widely used self-reported measure in metamemory research—judgements of learning (JOLs), in the context of its potential reactive effects on subsequent memory performance.

The following article serves as a non-systematic critical literature review, according to Kraus et al.'s (2022) classification by which such reviews are conducted "without any systematic procedure or protocol; instead, they weave together relevant literature based on the critical evaluations and (subjective) choices of the author(s) through a process of discovery and critique" (p. 2581). First, I will explain the phenomenon of JOL reactivity and then provide a brief overview of literature on the matter, focusing on the key findings. Next, I will outline four notable theoretical frameworks used to explain the JOL reactivity effects along with their empirical evidence and new research avenues to test them, after which I will highlight the practical and theoretical implications of these studies. As a result, some general possible directions for future research will be proposed.

Lastly, some acknowledgements are necessary with respect to the following literature review. While the literature coverage may not be entirely exhaustive, the works selected for inclusion in this paper are the most relevant and important discoveries to date in context of JOL reactivity research. Given the subjective nature of non-systematic literature reviews, the current article is a reflection of the authors' own informed opinion on the topic, and is therefore, inevitably subject to potential biases.

## Definition of Judgements of Learning Reactivity

Metacognition is widely understood as the knowledge and regulation of one's own cognition (Nelson & Narens, 1990; Veenman et al., 2006). Specifically, metamemory is a subsection of metacognition that involves one's knowledge, awareness and understanding of one's own memory capabilities, strategies, strengths, and weaknesses. Metamemory is also present at all stages of the memory processes: encoding, retention and retrieving information. In the classical model of metamemory devised by Nelson and Narens (1990), a distinction is made between the interrelated processes of monitoring and control. Monitoring refers to subjectively assessing one's ongoing learning, whereas control refers to all behavioural decisions that learners must make during learning. While monitoring may inform metacognitive control, the control may, in turn, modify monitoring. For instance, when preparing for an exam, a learner may realise that they have low confidence in their memory for certain parts of the material (monitoring). As a result, they may decide to delegate more time to studying them (control). The time spent on those portions of the material may then be positively translated into assessing them as highly likely to be remembered later (control informs monitoring).

Judgements of learning (JOLs) are one of the most used self-reported measures of metamemory monitoring that occur at the stage of information encoding (Dunlosky & Tauber, 2016). They concern the prediction about one's future memory performance. Particularly, when providing a JOL, usually an estimate is made about the likelihood of being able to recall a given to-be-encoded information on a later test (Rhodes, 2016). For example, JOLs may answer such questions as "How much of the presented text will I remember?" when passages of textbooks are presented. Generally, in the studies employing JOLs, participants are presented with items to-be-learnt on which they are asked to formulate their judgements, typically on a scale of 0–100. Ultimately, they have their memory on those items assessed through a test, to then enable researchers to evaluate the accuracy of those predictions.

The assessment of JOLs' accuracy, when compared to actual learning performance, is widely believed to be indicative of a learner's effectiveness in monitoring their learning process (Thompson, 1999). Generally, there are two methods of evaluating metacognitive monitoring accuracy: calibration and resolution. Calibration (or absolute accuracy) refers to the extent to which predictive judgements reflect actual memory performance (Dunlosky & Tauber, 2016). For instance, if items that are given JOLs of 70% are actually recalled

at an average of 40%, the participant can be deemed overconfident. Resolution (or relative accuracy) is the extent to which a learner discriminates between items that will be remembered and those that will not (Dunlosky & Tauber, 2016). To illustrate, if the mean prediction value for the actually recalled items is 90%, and thus higher compared to 50% predicted on average for the unrecalled items, the resolution can be deemed positive.

The issue concerning both methods of evaluating JOL's accuracy is that they were not intended to measure the potential for JOLs to be reactive: whether the sole act of soliciting JOLs could alter memory representation itself. Historically, reactivity was not so much of a concern, given that JOLs have been utilised mainly for the purpose of assessing participants' subjective beliefs about their learning. However, the issue of JOLs' potential for reactivity has been raised ever since the beginning of JOL implementation in metamemory research. Specifically, Spellman and Bjork (1992) were one amongst first researchers to point out that the task of providing JOLs may alter later memory performance. Some authors even warned about these effects as a potential confound in interpreting the results (Kimball & Metcalfe, 2003). Despite this, there was still a tacit assumption that JOLs do not affect the underlying memory processes and thus, they continued to be used in metamemory studies without employing any appropriate measures to examine their potential for reactivity (e.g., Koriat & Bjork, 2005). Consequently, previous studies have yielded inconclusive evidence on whether JOL reactivity actually occurs, with the scientific debate on the matter being mainly informed by studies which did not address this concern directly. In the past, only a limited number of studies have undertaken JOL reactivity as a primary goal of a study. This would entail comparing two experimental conditions in regard to test performance: one where participants are asked to provide JOLs, and the other without these metamemory judgements. Only later did some vocal researchers stress the importance of adding an additional non-judgement group to a study as a standard practice to determine whether JOLs are reactive (Soderstrom et al., 2015). With this in mind, the studies selected for inclusion in this paper directly examine JOL reactivity by employing the aforementioned participant groups: the experimental JOL group and the no-JOL control group.

In general, three directions of JOL reactivity have been demonstrated in research literature—either positive reactivity (Double et al., 2018; Soderstrom et al., 2015) in the form of improved memory performance for judged items, negative reactivity (Mitchum et al., 2016), that is, impaired memory performance on judged items or lack of reactivity (Ariel et al., 2021; Benjamin

et al., 1998; Dougherty et al., 2018), a too small effect in terms of recall to be deemed significant.

Both the magnitude and direction of JOL reactivity may depend on the type of JOL, stimuli and test used to assess memory performance (Chang & Brainerd, 2023). This may explain the occurrences of positive, negative and no reactivity in up-to-date research. Specifically, JOL reactivity has been investigated across various study materials, test formats and populations, with even different varieties of JOLs, which enabled researchers to establish some of its moderating factors. Previously, the majority of research literature has been primarily focused on delayed JOLs, provided some time after initially studying an item, and those delayed judgements have been consistently shown to relatively accurately predict test performance (Nelson & Dunlosky, 1991) and cause positive reactivity under some conditions (Rhodes & Tauber, 2011). Over time, researchers' attention has shifted towards reactivity of immediate JOLs—made either immediately after or during each individual item presentation. In particular, scholars have raised a concern that immediate JOLs may be especially prone to reactivity (Double et al., 2018). They may alter the processes they intended to measure by affecting the ongoing memorisation process and subsequent memory performance (Mitchum et al., 2016).

Some studies have reported positive reactivity when utilising lists of single words (Senkova & Otani, 2021) and identical word pairs such as "rock—rock" (Halamish & Undorf, 2023). Additionally, current studies imply that positive JOL reactivity can be extended to memory for visual materials such as images of objects and scenes (Shi et al., 2023). Interestingly, JOLs' reactive effects may not be short-lived, as the study by Witherby and Tauber (2017) found that the reactivity was exhibited even after testing two days later.

However, the most widely contended dissociable effect over immediate JOL reactivity concerned related (e.g., loaf—bread) and unrelated (e.g., practice—tree) word pairs, with the discourse beginning with two influential studies. While Soderstrom et al. (2015) discovered positive JOL reactivity for related pairs but no reactivity for unrelated pairs, Mitchum et al. (2016) found otherwise—they discovered negative JOL reactivity for word pairs that were unrelated; however, their study failed to produce any reactivity effects for related word pairs. Janes et al. (2018; Experiment 1) tested the replicability of both studies' findings, employing the same word-pair learning paradigm. They accounted for the main disparity between both studies in the form of either experimenter versus self-paced learning by incorporating those two separate procedures within one study. It was found that making JOLs further increases

the tendency to recall more related pairs than without eliciting such judgements, an effect mainly driven by positive reactivity. This effect was more robust for the experimental-paced study than the self-paced study. Since then, numerous studies have also shown positive reactivity for semantically related word pairs, but no other significant effects were found for the other types of stimuli such as unrelated and mixed pairs (see Double et al., 2018, for review). This positive reactivity pattern has emerged not only for the traditionally used forward paired associates (e.g., peanut—butter), but also for pairs associated backwards (e.g., butter—peanut) and symmetrically (e.g., closed—open) as well (Maxwell & Huff, 2022). This implies that the mere presence of an association, regardless of its direction, is sufficient to induce such positive effects. Moreover, even when immediate JOLs were manipulated within participants and not between participants, as has been traditionally done, the positive reactive effects of JOLs for related pairs still emerged (Rivers et al., 2021). Of note, however, the way memory is tested may have an impact over whether or not the positive effect will occur. Using a test format such as cued-recall tests has allowed researchers to successfully reproduce the JOL memory enhancement for related pairs, but not on such tests as free recall (Myers et al., 2020).

Most recently, however, Undorf et al. (2024) have provided solid evidence that positive JOL reactivity observed for related pairs is just as robust and reliable as negative reactivity found for unrelated pairs. They conducted small-scale meta-analyses including 17 experiments ran by either author, testing multiple potential moderators of JOL reactivity. The experimental setting moderated negative reactivity for unrelated word pairs, with the memorial costs being more pronounced in tightly controlled experiments than in unsupervised online experiments. The language of study items moderated positive reactivity, as the memorial benefits were stronger for the UK or US participants who studied word pairs in their everyday English language, than for German and Israeli participants studying in their respective first languages. No other factors such as the presence of an additional pair type, study time or total number of study items, influenced JOL reactivity.

However, studies using educational texts as study materials have consistently shown that the reactive effects of JOLs translate rather poorly to more complex verbal materials (cf. Kubik et al., 2022). For instance, a study by Ariel et al. (2021) failed to find any reactive effects for JOLs when using educationally relevant texts in most of their experiments, except one: when overt retrieval preceded making JOLs. The researchers mainly used target-absent JOLs (provided in the absence of the targeted information), which are

assumed to afford covert retrieval practice opportunities by initiating the memory search for an answer during the judgement forming process (Nelson & Dunlosky, 1991). When the retrieval dynamics assumed to be involved in making such JOLs (i.e., covert retrieval practice) were compared to those that govern overt retrieval practice, contrary to latter task, solely making JOLs did not provide any learning benefits. However, when combined together, making JOLs led to larger learning gains than overt retrieval alone. This implied that making JOLs following the instruction to overtly recall targeted information may indirectly reactively enhance the benefits of testing for educationally relevant materials, thus jointly providing larger benefits than testing alone. Critically, however, Zhao, Xu, et al. (2023) consistently failed to replicate those incidental finding across various experiments, despite ensuring close resemblance to the original study. Similarly, Schäfer and Undorf (2024) also found no evidence of JOL reactivity when learning general knowledge facts.

However, Hausman and Kubik (2023) argued that in the previous studies, the conditions accompanying making JOLs were less than optimal for the potential reactive benefits to appear on learning educational materials. Similarly to Ariel et al. (2021), they used target-absent, cue-only JOLs. These judgements were provided at a delay, with the intent to elicit more complex covert retrieval than if their immediate counterparts were used. Contrary to other studies, they included a one-week long delay before administering the final test, to further boost the effectiveness of learning enhancement from retrieval practice. Some participants were also given the instructions to engage in covert retrieval of the read text prior to giving JOLs, for more enhanced text comprehension than JOLs alone. However, even after creating a more optimal setup, the study found no reactively enhanced comprehension of expository texts.

Lastly, there have also been findings implying that age differences may play a role in reactively enhancing recall via JOLs—namely, it was found that positive JOL reactivity occurs even amongst young children (Zhao et al., 2022) and young adults, but not older adults (Tauber & Witherby, 2019). This implies that JOL reactivity shows a trend of improving memory at the early stages of life, followed by a decline nearing its end (Chang & Brainerd, 2023).

## Theoretical Framework: The Four Main JOL Reactivity Hypotheses

The common feature of all theories explaining JOL reactivity is the assumption that the measurement itself alters the way people monitor their memory. Particularly, Ericsson and Simon (1980) showed that providing verbal reports during information encoding may affect memory performance, as the participants' attention is drawn to information that would otherwise not be available to them. Similarly, Double et al. (2018) note that JOL reactivity may suggest that judgements of such nature may not be generated spontaneously during monitoring, in the same manner as when JOLs are elicited (see also Mitchum et al., 2016). In the absence of JOLs, the corresponding judgements may either not be made or are made, but in a different way.

Four hypotheses have emerged in an attempt to explain the mechanism underlying memory alteration due to JOL reactivity: the cue strengthening hypothesis (Soderstrom et al., 2015), the changed-goal hypothesis (Mitchum et al., 2016), the item-specific processing hypothesis (Senkova & Otani, 2021) and the attention-orienting or enhanced engagement account (Shi et al., 2023; Tauber & Witherby, 2019). Their common feature is that they attempt to explain certain dissociable effects present in the studies mentioned in the previous section.

### The Cue-Strengthening Hypothesis

The most empirically supported theoretical hypothesis, which mainly explains positive reactivity, is the cue-strengthening hypothesis (Soderstrom et al., 2015), which draws upon the Koriat's cue-utilization approach to JOLs (Koriat, 1997) and de Winstanley et al.'s (1996) transfer-appropriate multifactor account of generation effects. The cue-utilization approach to JOLs postulates that JOLs are inferential in nature: when making JOLs, learners utilise the readily available heuristics and cues at the time of encoding. Koriat (1997) distinguishes between three types of cues: intrinsic—involving the item characteristics related to their perceived ease or difficulty of learning (cue-target relatedness); extrinsic—either concerned with learning conditions or applied encoding operations (presentation time, level of processing); and mnemonic—subjective cues that are phenomenal experiences accompanying one's learning process (encoding fluency, ease of processing). De Winstanley et al.'s (1996) transfer-appropriate multifactor account of generation effects suggests

that the act of generating, that is, taking an active part in producing information during encoding (e.g., po_at_) rather than receiving it via an external source, for example, when reading it intact (e.g., potato), strengthens the information used to complete a generative task. If this information is pertinent to the later criterion test, a generation effect will occur, providing a memory advantage. Precisely, it is the joint account of the two that led Soderstrom et al. (2015) to formulate the cue-strengthening hypothesis, whereby the act of eliciting JOLs strengthens the cues which are used as basis for arriving at these judgements. Moreover, the heightened performance will happen provided that the memory test is sensitive to the cues used to inform JOLs (Soderstrom et al., 2015).

As cue-target relatedness is one such cue that can be salient during encoding, it can be used as a basis to determine what will be remembered, which will then be reflected in JOLs of varying magnitude. Therefore, the sole act of making JOLs can strengthen the association between the related word pairs, which in practical terms will result in improved memory performance. In consequence, positive reactivity will occur for related pairs to a greater extent than for unrelated pairs, due to the pre-existing association between the words. In the absence of such cues, there will be no meaningful relationship to strengthen and thus no if any reactivity should occur. For example, when encountering a semantically related pair such as "banana—monkey" and assessing its chances of being remembered in the future, the formed judgements will be informed by the a priori semantic association of the cue-target, therefore benefitting the material being learnt. By forming JOLs in such context, learners reap an added benefit of additional processing. If the word pair is unrelated, for instance "banana—drill", giving JOLs cannot per se strengthen the association due to the lack thereof. However, we will not observe reactivity if we simply ask to recall all word pairs in a free-recall test. Instead, if we ask individuals to recall the target words after presenting their corresponding cues (e.g., banana—?) in a cued-recall test, then reactivity will be observed, as such test is sensitive to the cue of cue-target relatedness.

The seminal work by Soderstrom et al. (2015) not only provided sound evidence in favour of the cue-strengthening hypothesis, but it also pioneered in systematically controlling for the potential confounding variable in the previous research, which was the different duration of item exposure between the two participants groups. They used pairs which differed in associative relatedness in order to increase the salience of this intrinsic cue, so that it would be more likely used as a basis to inform participants' JOLs. The JOL group

gave their JOLs on a scale of 0–100% (denoting the likelihood of successful recall) halfway through the exposure of each pair, while the control group made no JOLs during the entire duration of each pair exposure. Enhanced cued-recall memory performance was observed for strongly related pairs. However, positive reactivity did not occur for the weakly and unrelated pairs as a result of making JOLs. Their experiments jointly demonstrate that when the intrinsic cue of relatedness was salient for participants, it was used as a basis for formulating their JOLs, which in turn strengthened the existing cue-target association.

Nonetheless, the study by Mitchum et al. (2016; Experiment 5) using an experimenter-paced procedure, in which the participants had to study each word pair within the time constraints provided externally, directly contradicts Soderstrom et al.'s (2015) finding on positive reactivity for related word pairs. Despite using the same paired associates learning paradigm, the study found that making JOLs impairs memory performance for unrelated pairs and shows no effect on the related pairs. Therefore, there was no overall memory enhancement that could be attributed to cue-strengthening as a result of eliciting JOLs, which is typically indicative of positive JOL reactivity.[1]

Despite these disparities in results, many studies after Soderstrom et al.'s (2015) publication still managed to replicate and even extend their findings (e.g., Chang & Brainerd, 2023; Myers et al., 2020; Rivers et al., 2021). For related pairs specifically, it was shown that positive reactivity only occurs when memory tests are sensitive to cue-target relatedness—in other words, improved memory will only occur if the relatedness information is useful on that test, such as on cued-recall tests and item-recognition tests (Myers et al., 2020). Conversely, the reactive effect is absent when using less sensitive tests which are devoid of cues—formally established during the study phase as the first word of each pair that later serves as a hint for the paired second word at test—as the benefit of cue-strengthening is rendered useless. To demonstrate, no reactivity was observed for related pairs on free recall tests (e.g., Myers et al., 2020). Therefore, a crucial premise of the cue-strengthening hypothesis is that

[1] Of note, despite coming from different theoretical standpoints and arriving at divergent reactivity results, both Soderstrom et al. (2015) and Mitchum et al. (2016) demonstrated the enhanced relatedness effect. The relatedness effect is the natural tendency to recall more related pairs than unrelated pairs. However, JOL solicitation further increases this effect, leading to even bigger differences in recall between the related and unrelated pairs in the JOL group compared to the control (Janes et al., 2018). Janes et al. (2018) examined the main mechanism driving this effect (whether it is positive reactivity for related pairs, negative reactivity for unrelated pairs or both) and found that this increased effect is mostly driven by positive reactivity.

JOL reactivity relies on the interaction between study materials, JOLs and memory tests (Chang & Brainerd, 2023).

With that in mind, Chang and Brainerd (2023; Experiment 3) also examined the interaction between the study material type (target-target related and unrelated pairs), JOL type (item-, list-level and no- JOL) and test format (associative recall and free recall). Normally, if participants are given cue-target word pairs to learn and a corresponding cued-recall or associative test is later administered, making JOLs individually for each pair (i.e., item-level JOLs) will strengthen the association within the related pairs on that test, causing positive reactivity. However, in this study, participants were given target–target word pairs, for which only making JOLs on the entire list (i.e., list-level JOLs) of such related pairs strengthened the categorical relatedness between the pairs on a free-recall test, leading to positive reactivity. In any other combination than the specified, list-level JOLs led to no reactivity. No interplay of factors including item-level JOLs led to any reactivity on either type of pair or test, as the study materials (i.e., target-target related pairs) were designed to favour relatedness processing between pairs and not within them.

However, the classic cue-strengthening account may be insufficient to explain positive reactivity observed on related pairs with different associative direction (e.g., backward associates) or type of association (direct vs. indirect) than the typically used direct forward associates (e.g., peanut—butter). Even when such pairs are somewhat related, the relatedness cues may not be readily available (i.e., salient enough) to be strengthened at encoding or fail to be useful at a later cued-recall test if the target is not a typical response to the cue, yet positive reactivity may still be observed for these types of pairs. Such is the case for mediated associates like "stripes—lion", which are indirectly related through an absent mediator like "tiger" (Maxwell & Huff, 2024), and backward associates such as "butter—peanut" (Maxwell & Huff, 2022, 2023; cf. Mitchum et al., 2016), respectively. It is therefore likely that cue-strengthening also entails relational encoding (Halamish & Undorf, 2023; Maxwell & Huff, 2024). This means that making JOLs not only strengthens the observable relatedness cues et encoding, but it also induces relational encoding processing by additionally reinforcing the underlying cue–target relations (even through indirect associations, as was seen for mediated pairs). In short, the forces of cue-strengthening and relational encoding seem to operate in unison, with the dominance of either one being a function of the type of pre-existing association.

Positive reactivity, which likely reflects cue-strengthening via relational encoding, is therefore less concerned with the extent to which the relatedness cues are apparent at encoding or testing, and more with the presence of underlying cue-target associations. Specifically, Witherby et al. (2023) found that for such benefits to occur, the said relationships cannot be created through simply forming mental images of cue-target words interacting with one another, but they have to be pre-existing and semantic by nature. Per the cue-strengthening account, the researchers predicted that the relationship established through forming interactive imagery would be strengthened by making JOLs, thus benefitting unrelated pairs. However, the requirement to use interactive imagery during learning did not affect the reactive effects of JOLs for unrelated pairs, since negative reactivity was observed regardless of this prompt. Similarly, Rivers, Dunlosky, et al. (2023) found that making JOLs led to no reactivity for lexically associated letter-cued pairs (e.g., ja—jade) but produced positive reactivity for semantically related category-cued pairs (e.g., a type of gem—jade). These findings provide converging evidence that the benefits of cue-strengthening may solely be restricted to pre-existing semantic associations, therefore setting probable bounds of its explanatory power.

The most notable criticism towards the cue-strengthening hypothesis is that while it explains positive JOL reactivity, it does not precisely account for negative reactivity (Janes et al., 2018; Mitchum et al., 2016; Rivers et al., 2021), as its current assumptions are that in the absence of cues, little to no reactivity should occur. Some scholars such as Janes et al. (2018; see also Mitchum et al., 2016) have suggested to extend upon this account to include dual-task costs, attributing negative JOL reactivity to an interference in learning, resulting from having to perform two tasks concurrently: monitoring one's learning by making JOLs and the primary learning task. This particularly poses an issue when the learning task is difficult and resource-demanding, as in case of unrelated pairs.

Alternatively, Double and Birney (2019) addressed the said limitation by introducing a related theoretical framework named the cue-processing account, which draws directly from cue-strengthening. It rests on the premise that when making JOLs, cognitive resources are directed to processing the cues that are salient in the learning environment, with only some of them being relevant to future recall. These are then considered for their potential to impact future recall. However, even if a learner ultimately decides not to utilise some of the cues to inform their JOLs, their salience is sufficient enough to generate the additional processing of these cues. As a result, if the later criterion test is

sensitive to these cues, the additional processing will lead to memorial benefits. Conversely, if the additional processing afforded by JOLs is expended on the more salient uninformative cues instead of the less salient informative ones (e.g., relatedness cue), negative reactivity will occur. As evidenced by Double (2023), when a salient yet uninformative cue, that is, font size, was made present during an experimenter-paced study, making JOLs led to memory impairment —an assumption that fits the cue-processing account.

Lastly, Janes et al. (2018) raised an important point that the cue-strengthening hypothesis also fails to explain a lack of reactivity shown in some studies using pure lists of either related or unrelated word pairs as observed in their study (see also Mitchum et al., 2016). However, other experiments did manage to show this finding (Tauber & Witherby, 2019; Witherby & Tauber, 2017).

The cue-strengthening hypothesis (with dual-task costs) still remains the most comprehensive explanation for JOL reactivity to date, having received substantial empirical support through various testing. If revised to include the relatedness processing assumption, the cue-strengthening hypothesis accounts for patterns of reactivity observed on semantic, pre-existing associations, regardless of the direction or type of association.

Thus far, only the relatedness cues have been shown to systematically affect JOL reactivity. This is justified, given the extensive focus on the paired associates. However, other cues have been found to influence reactivity as well (e.g., font size; Double, 2023). The question that remains unanswered is what their relative contribution is to producing the final reactive outcome. The relatedness cues likely override the influence of any other salient cues, as they are perceived to be highly diagnostic of one's future performance (e.g., Rivers, Dunlosky, et al., 2023). As such, other cues may be strengthened to a lesser degree, which may even be insufficient to be detected through testing.

An avenue for future research will be to establish the degree to which specific cues have to be salient in order to produce the reactive JOL effects. Future studies can manipulate the salience of single or multiples cues, more or less informative ones, paired with other study materials and a test that could be relevant to these cues. To demonstrate, JOL reactivity could be examined in context of a list of single words differing in the intrinsic cue that is concreteness (vs. abstractness) with a recognition test administered later.

### The Changed-Goal Hypothesis

An alternative hypothesis, which mainly explains negative JOL reactivity, is the changed-goal hypothesis (Mitchum et al., 2016). It posits that requiring participants to make JOLs will incline them to notice that the to-be-learned material varies in difficulty, that is, only some items will be remembered. By default, learners adopt a mastery approach, wherein more time is expended on difficult items, as opposed to easy ones (Dunlosky & Hertzog, 1998). However, the mentioned distinction in difficulty will lead to learners shifting their standard study goal from mastering all items to only memorising the easy and moderately difficult ones at the expense of harder items, in order to optimise their performance (Metcalfe & Kornell, 2005). For instance, the shift caused by making JOLs would occur towards focusing on related pairs (i.e., perceived as easy) instead of unrelated ones (i.e., perceived as difficult), thus increasing the discrepancy in recall—an effect mainly driven by negative reactivity for unrelated pairs. This tendency is especially likely to occur under time constraints, such as the one imposed by an experimenter (see Mitchum et al., 2016). This hypothesis initially served as an explanation for negative JOL reactivity, as observed in the Mitchum et al.'s (2016) study mentioned earlier. However, other researchers have extended upon this theory to account for positive reactivity via a logical assumption: inversely to the costs pertaining to difficult items, the prioritisation of easy items should result in positive reactive effects (e.g., Myers et al., 2020).

The changed-goal hypothesis has yielded contradictory findings, with some studies supporting the notion of goal change in regard to JOLs (Mitchum et al., 2016; see also: Janes et al., 2018; Li et al., 2024) and others challenging it (Halamish & Undorf, 2023; Tekin & Roediger, 2020). Specifically, in three experiments by Mitchum et al. (2016) the researchers utilised a mixed list of related and unrelated word pairs, ensuring that the sets of cue-target pairs varied in memorability within the list, across all of them. In all three, the correlation between the objective item difficulty (varying degree of cue-target relatedness) and the self-paced study time was weaker when JOLs were solicited, relative to when not, irrespective of their form. In other words, the tendency to master all items, which can be translated into studying unrelated items longer, was weaker as a result of the JOL probe. The differences in memory performance between the items were larger in the JOL condition, with a main effect of negative JOL reactivity occurring for unrelated pairs. Additionally, when there was no salient cue for difficulty to assess the memorability

between items, that is, all word pairs were unrelated, the researchers observed no difference in recall performance between the JOL and non-JOL condition. A similar study by Janes et al.'s (2018; Experiment 2) showed a similar finding, namely that positive JOL reactivity occurs when using a mixed list of related and unrelated pairs, but the reactive effects are reduced when using a pure list, only consisting either entirely of related or unrelated pairs (cf. Tauber & Witherby, 2019). This stands in line with the changed-goal hypothesis, as in pure lists as opposed to mixed lists, there are no pairs from which one could easily distinguish items that are difficult or easy to remember.

When the study time was experimenter-paced (Mitchum et al., 2016; Experiment 5), the discrepancy in performance between related and unrelated items was even greater in the JOL group as opposed to the control. These results were also replicated by Li et al. (2024) and Janes et al. (2018). However, when those researchers replicated the self-paced study procedure by Mitchum et al. (2016) mentioned earlier, the results either showed no enhanced relatedness effect or found no relatedness effect at all, respectively. Furthermore, both of the studies failed to find a weaker correlation between cue-target relatedness and study time for the JOL group when compared to the non-JOL group, showing that the time spent on studying items is not relative to their difficulty and therefore challenging the changed-goal hypothesis.

On the other hand, Li et al. (2024) also included prestudy JOLs (i.e., made before studying each item), hypothesising based on the theory that soliciting JOLs before even seeing each study item should bring awareness to differences in memorability. Prestudy JOLs did not, in fact, reactively induce enhanced relatedness effects at all in their study. Chang and Brainerd (2023; Experiment 2) also examined the reactivity of prestudy JOLs, comparing them to immediate JOLs and no probe for JOLs, when using a list of related and unrelated word pairs. The study found that making prestudy JOLs led to no reactivity, regardless of the pair type. These results contradict the notion of a goal change, as similarly to immediate JOLs, prestudy JOLs should induce negative reactivity for unrelated pairs and positive reactivity for related pairs, given that they both highlight differences in item difficulty. However, the cue-strengthening hypothesis is a more suitable account in this instance, as it predicts weaker reactivity for prestudy JOLs compared to immediate JOLs, due to the comparably fewer cues being present when forming prestudy JOLs and thus reduced usefulness on the associative test. Especially that immediate JOLs produced positive reactivity for related pairs further corroborated this hypothesis.

Furthermore, Tekin and Roediger's (2020) findings also contradict the changed-goal hypothesis. The study found greater positive reactivity in the form of improved recognition for items processed shallowly more so than deeply, in a levels-of-processing (LOP) paradigm. According to the LOP effect, it would follow that a deeply processed item will be remembered better than an item processed shallowly—relating to the changed-goal hypothesis in that the former will be perceived as easy, and the latter as difficult. However, the study found otherwise: JOL solicitation was more beneficial for the shallow tasks rather than the deep ones, thus attenuating the LOP effect. The authors further suggest that their results may instead serve as an extension of the cue-strengthening hypothesis, as making JOLs may strengthen the information that normally would not be strengthened and thus, JOLs enhanced performance by promoting semantic processing in shallow orienting tasks while the test recognition was sensitive to this information. In other words, there was still room for further strengthening for items processed shallowly, but not deeply.

Although Halamish and Undorf's (2023) study did not test the changed-goal hypothesis directly, its results also run counter to it. It was the first study to examine whether JOLs affect recall for identical pairs, additionally using related and unrelated pairs for comparison. It revealed a medium-sized positive reactivity for related pairs, no reactivity for unrelated pairs, and a small-sized positive reactivity for identical pairs. These results challenge the predictions of a changed goal, as the positive reactivity for related pairs was larger than for identical pairs, even though it would follow that identical pairs should receive even greater positive reactivity due to their perceived ease of learning. According to the authors, these results instead fitted the predictions of the cue-strengthening hypothesis, in which relatedness processing assumption is tacitly embedded (see also Maxwell & Huff, 2024), showing that the enhanced relatedness processing that occurs when making JOLs can be generalised to a different type of relatedness than related pairs—that is, identical pairs.

Likewise, Chang and Brainerd (2023) tested the two major JOL reactivity frameworks against one another, using identical pairs in one of their experiments. They examined JOL reactivity across three types of word pairs with varying difficulty: identical, strongly and weakly related pairs, on an associative-recall test. In all three cases, making JOLs resulted in enhanced recall, which directly contradicts the changed-goal hypothesis. According to that hypothesis, it would follow that making JOLs should provide discrepant results for items with varying difficulty, with the goal change being reflected

via positive reactivity for easy and moderately challenging items (i.e., identical and strongly related pairs) and negative reactivity for difficult items (i.e., weakly related pairs). However, the results are more in line with the assumptions of the cue-strengthening hypothesis, as for all three types of word pairs the intrinsic cues (i.e., cue–target identity, cue–target relatedness) were strengthened through making JOLs and were diagnostic in the later administered associative test, resulting in positive reactivity across all of them.

Overall, when compared to the cue-strengthening hypothesis, the changed-goal hypothesis lacks the explanatory power to account for the various contradictory findings, for which cue-strengthening may instead serve as a more reliable explanation. Specifically, the changed-goal account has difficulty accounting for instances when the distinction between item difficulty (i.e., degree of pair relatedness) is not overtly clear. As such, this explanation may apply only to simple cases dealing with a mixed list of related and unrelated pairs, but not when the study items display varying degrees of difficulty.

Additionally, this hypothesis has premises which are objectively more difficult to confirm in order to corroborate the entire hypothesis, while the cue-strengthening hypothesis provides a more concise and simplistic explanation. However, the findings on no reactivity on pure lists, found in some studies, may be explained only in terms of the changed-goal hypothesis (Janes et al., 2018; Mitchum et al., 2016).

Critically, Undorf et al.'s (2024) publication mentioned earlier provides rather convincing evidence in support of the cue-strengthening hypothesis with dual-task costs, against the changed-goal hypothesis, as experiments which found positive reactivity tended not to reveal negative reactivity and vice versa. This heavily contradicts the changed goal, as it would be expected that the experiments show both strong positive and negative reactivity relative to major shifts to related pairs. In addition, the failure to reproduce Mitchum et al.'s (2016) self-paced study findings, as was seen in Li et al. (2024) and Janes et al. (2018), questions the replicability of certain results.

Future research testing the changed-goal account of JOL reactivity could focus on the subjectively perceived item difficulty, involving mnemonic cues—such as encoding fluency or ease of processing—in terms of manipulating the subjective experience of an item difficulty. This proposed avenue of research seems rather promising, given the earlier findings on negative reactivity found on pure lists of related and unrelated pairs, when the font sizes were inconsistent rather than consistent (Double, 2023), a cue known to affect the perception of fluency (Chang & Brainerd, 2022). Alternatively, one could

use a list of comparably memorable words, which differ in font sizes, and examine reactivity as a function of ease of processing the words.

### The Item-Specific Processing Hypothesis

Besides the two major theoretical accounts described earlier, there is a lesser-known hypothesis named the item-specific processing hypothesis (Senkova & Otani, 2021; Zhao, Li, et al., 2023) that mainly explains JOL reactivity for lists of words. This hypothesis arose from the item-specific and relational processing framework (Einstein & Hunt, 1980; Hunt & Einstein, 1981) wherein encoding is broken down into two distinctive processes—item-specific encoding and relational encoding. Item-specific encoding concentrates on unique characteristics that differentiate an item from another, whereas relational encoding focuses on shared characteristics among the items. It is the combination of the two that is deemed optimal for memory (Hunt, 2006). The rationale behind this hypothesis is that JOLs facilitate memory performance by inducing item-specific processing. In other words, the act of eliciting JOLs draws attention towards a specific item, and thus enhances its distinctiveness in memory.

Senkova and Otani (2021) theorised that when a task such as JOL-making induces item-specific processing, the list (or rather its structure) should complementarily promote relational processing (e.g., categorised vs. uncategorised lists), and it is the joint forces of both factors that ultimately lead to memory enhancement. When compared to two other tasks which similarly to JOLs induce item-specific processing, that is, pleasantness rating and single imagery tasks, all three resulted in greater recall than the control when the list was categorised. Moreover, the improvement of recall occurred to a similar degree across all three conditions. Zhao, Li, et al. (2023) conceptualised this hypothesis in a similar manner, in the context of temporal order memory for which JOL reactivity was assessed in their study, naming it the item-order account. They theorised that while JOLs enhance item-specific processing, the focus on individual items interferes with relational processing. Specifically, Zhao, Li, et al. found that making JOLs concurrently with each word presentation impairs the reconstruction of the studied order of words, in a test which relies heavily on relational processing. However, when using forced-choice recognition tests, which instead rely on item-specific processing, concurrent JOL solicitation bolstered memory performance.

Nonetheless, there is evidence to suggest that the item-specific processing hypothesis may not be the driving mechanism behind JOL reactivity after all. Stevens and Pierce (2019) failed to find any reactive effects for item-level JOLs when using categorised lists that are meant to encourage relational processing. However, they managed to show positive reactivity for list-level JOLs, which are solicited after studying an entire block presentation of a categorised list. The researchers argued that item-level JOLs do indeed encourage item specific processing, but lists that are structured categorically instead benefit from relational processing, which is induced by list-level JOLs. This runs contradictorily to Senkova and Otani's (2021) finding on JOL memory enhancement for categorised lists for concurrent JOLs.

Chang and Brainerd (2024) pointed out that the two studies mentioned earlier differ in the blocked versus random presentation of the words as well as the implementation of item- versus list-level JOLs. In an attempt to resolve those discrepancies, they implemented appropriate manipulations of the structural organisation of categorised lists (randomised vs. blocked) and the JOL conditions (item- vs. list-level JOLs vs. no-JOL) in their study. The researchers found that item-level JOLs led to positive reactivity when the words were randomised within the list, but produced no reactivity when they were blocked according to the category. This finding contradicts the item-specific processing hypothesis, since regardless of the structural form, that is, whether blocked or randomised, positive JOL reactivity should occur for categorised lists by enhancing item-specific processing complementarily to relational processing, which category lists mainly promote. Furthermore, an even greater memory enhancement should occur for blocked lists, in which relational processing predominates (Ackerman, 1986). The study somewhat failed to replicate Stevens and Pierce's (2019) findings, as neither the item- nor the list-levels JOLs improved memory for blocked lists, although this specifically could be attributed to differences in the test format (i.e., the use of free recall instead of cued recall). Interestingly, the researchers speculate that positive item-level JOL reactivity for categorised lists may instead be driven by relational processing, which would complement randomised lists for which relational processing is hindered.

On a related note, Maxwell and Huff (2022) conducted a series of experiments which led them to conclude that it is the selective use of relational encoding that drives positive JOL reactivity—a strategy that is encouraged both implicitly and explicitly. For example, in one of their experiments, the task of making JOLs was compared with two other judgements tasks that resembled

JOLs in terms of the rating process, both of which only implicitly directed attention towards the relatedness of items and excluded its typical component of memorial forecasting. Across all three tasks, there were equivalent patterns of positive reactivity on related pairs and no reactivity on unrelated pairs, which suggests that all of them likely depend on the selective relational encoding for related pairs. In another study by Maxwell and Huff (2024) mentioned earlier, it was suggested the positive reactivity in case of mediated pairs likely indicates the contribution of increased relational encoding (i.e., strengthening of underlying cue-target relations) rather than the strengthening of cues, as the relatedness cues were virtually unavailable at encoding for this type of pairs. All these findings strongly point to relational encoding being the driving mechanism for JOL reactivity.

While it may still be argued that the item-specific processing hypothesis serves as a plausible explanation for JOL reactivity found on single words in some studies (Senkova & Otani, 2021; Zhao, Li, et al., 2023), others have arrived at disparate results (e.g., Chang & Brainerd, 2024). It may be concluded that the determining factor of whether JOL reactivity will occur is the organisational structure of the categorised list (Chang & Brainerd, 2024), which greatly undermines the explanatory power of the item-specific hypothesis for the reasons mentioned earlier. Inversely, in the paired associates paradigm, it may instead be the selective use of relational, not item-specific, encoding that is responsible for the positive reactivity on related pairs (e.g., Maxwell & Huff, 2022). This notion also seems to align with the cue-strengthening hypothesis, with the selective relational encoding being a function of pair relatedness.

Nonetheless, future research could benefit from comparing the task of making JOLs to a task that induces relational processing, and another one that encourages item-specific processing within one study, in order to see whether making JOLs resembles either type of task in their respective outcomes. Akin to other studies with a similar setup, this would only serve as an indirect way to infer whether making JOLs is driven by either type of processing. As such, it would be more desirable to directly address whether item-specific processing and relational processing occur within one study, for example, by having participants report the type of strategy used at the study phase as a function of a pair type—an idea proposed in a similar context by Maxwell and Huff (2022).

### The Attention-Orienting or Enhanced Engagement Account

The last hypothesis worth mentioning is the attention-orienting or enhanced engagement account (Shi et al., 2023; Tauber & Witherby, 2019), whereby the solicitation of JOLs enhances memory performance due to enhanced learning engagement, for example, study time or effort. Normally, during prolonged learning our mind tends to wander, which is detrimental to the learning process (Seli et al., 2018). However, forming judgements of learning concurrently, on an item-by-item basis, requires participants to continuously sustain their attention across the duration of a learning task, as effort is needed to analyse items to then be able to formulate appropriate judgements (Shi et al., 2023). In their study, Shi et al. (2023) implemented mind wandering (MW) probes as a measure of learners' engagement throughout the learning task, which required providing JOLs for images of scenes. The researchers found that prompting JOLs reduces mind wandering and leads to positive reactivity for images of scenes, an effect which was partially mediated by enhanced learning engagement.

Tauber and Witherby (2019) were the first to propose the "attentional reorienting" hypothesis as a possible mechanism behind the reactive JOL effects. They suggested that without the JOL probe, learners can expend the entire study duration to learning each word pair, and as the presentation progresses, they may be increasingly more likely to engage in mind wandering. Conversely, the prompt to make JOLs requires learners to consistently reengage with each word pair, making it less likely for their attention to wane. The researchers identified probable boundaries to which this theory may be applicable—while associations may be easily retained for related items, more complex study materials such as unrelated pairs or long texts may be out of reach for the beneficial effects of additional attention. They speculated that making JOLs requires divided attention and due to those dual task costs, older adults may be more resistant to the positive JOL effect than the younger adults. Tauber and Witherby also implied that older adults are more motivated to perform well in experimental tasks. Reorienting attention may therefore be redundant when one is internally motivated to sustain attention. Shi et al. (2023) further tested this assumption by manipulating the motivational aspect of learning amongst the participants. They predicted that enhancing learning motivation should paradoxically decrease the positive reactivity, as innate motivation to attend to the task would leave no room for JOLs to further enhance the engagement. The study found that enhancing learning motivation exter-

nally via an experimental manipulation reduced the magnitude of positive reactivity effect, as well as decreased the number of participants who benefited from JOL solicitation. Though the positive effect was reduced, it still persisted but to a lesser degree, leaving room for future research to investigate why this effect lasted.

Rivers, Janes, et al. (2023) have tested the hypothesis in mention on moderately related word pairs during their encoding, showing that this theoretical framework may not be reliably transferred to such study materials as related pairs. The researchers hypothesised that positive JOL reactivity may occur when a JOL prompt that is typically made halfway through each pair presentation encourages learners to become more attentive or "reorient" to that pair for the remaining half of the encoding period. They designed an experimental procedure in which they could simulate the said "reorientation" using a fixation point. Their initial study included four variants of the groups, in which JOLs were either made halfway through presentation or were not provided at all, and learners were presented with a fixation point instead of the JOL or did not receive one. Although large positive reactivity was found as a result of soliciting JOLs, the fixation point itself did not enhance memory performance.

The attention-orienting or enhanced engagement account is still in its infancy and requires much needed empirical testing. Its key limitation is that it only accounts for positive reactivity, its reduced effects or null effects. It makes no mention of why negative reactivity is sometimes observed in studies. However, its strength lies in including the learner's role in producing JOL reactivity. Future studies can employ different experimental manipulations for enhancing learning engagement, especially within the studied material itself, such as differentially incentivising learning by rewarding more points for certain items (i.e., easy vs. difficult items). In general, framing the task as an opportunity to receive rewards may produce different reactivity results compared to the study by Shi et al. (2023), which motivated participants through a potential penalty, deterring them from performing below a certain standard. Thus far, this theory has been shown to reliably explain JOL reactivity on visual memory. However, directions for future research include further testing on verbal materials such as lists of words and paired associates, as there is reasonable concern that it may not transfer to such contexts as was seen in Rivers, Janes, et al. (2023).

### Theoretical Frameworks Summary

In the previous sections, the studies on JOL reactivity were critically evaluated in context of the four main theoretical frameworks, underlining both their strengths and limitations. Based on the previous evaluation, some general conclusions can be drawn about each theory. Table 1 provides a summary of these accounts, along with their key empirical evidence and critical insights.

It is worth noting that these theoretical accounts are not mutually exclusive and may all to some degree be responsible for reactive JOL effects, depending on the context involved (Chang & Brainerd, 2023; Rivers, Janes, et al., 2023). Specifically, the item-specific processing hypothesis may be similar to the cue-strengthening hypothesis in a way that relates to strengthening item-specific cues via JOLs, which leads to improved memory performance (Senkova & Otani, 2021). From another perspective, Halamish and Undorf (2023) argue that the cue-strengthening account has an underlying tacit assumption about relatedness processing. Similarly, Maxwell and Huff (2024) suggest that the memorial benefits of JOLs on related pairs reflect relational encoding in combination with cue-strengthening, with the dominance of either one being a function of the type of association.

As it stands, the cue-strengthening hypothesis remains the most empirically supported theoretical framework out of all, as it serves as a rather parsimonious explanation for reactivity observed amongst word pairs under various conditions, especially if revised to include the assumption about relatedness processing. However, similarly to the changed-goal hypothesis, it does not account for the dissociable effects on word lists such as the item-specific processing hypothesis (Zhao, Li, et al., 2023). Nevertheless, the last hypothesis likely pertains only to single words lists and due to disparities in results, requires some additional testing. Lastly, none of the hypotheses, except for the attention-orienting account, include the learner as an autonomous agent, directly involved in producing JOL reactivity. However, the changed-goal hypothesis may provide valuable insight on how making JOLs affects study decisions under time constraints. Therefore, it could be beneficial to examine whether learners' motivation is somewhat connected to the change of a study goal when making JOLs, as both mechanisms have been independently provided as an explanation for reduced positive reactivity observed on different occasions.

**Table 1**

*Summary of Four Main Theoretical Frameworks of JOL Reactivity*

| Theoretical framework | Key studies | | Key strengths and limitations |
|---|---|---|---|
| | Supporting evidence | Contradictory evidence | |
| Cue-strengthening hypothesis | Soderstrom et al. (2015)<br>Myers et al. (2020)<br>Tekin and Roediger (2020)<br>Rivers et al. (2021)<br>Chang and Brainerd (2023)<br>Halamish and Undorf (2023)<br>Undorf et al. (2024) | Mitchum et al. (2016) | Well-empirically supported. Focus mainly on relatedness cues. No specific explanation for negative reactivity. Fails to explain no reactivity found on pure lists. |
| Changed-goal hypothesis | Mitchum et al. (2016)<br>[a] Janes et al. (2018) and Li et al. (2024) | Tekin and Roediger (2020)<br>Chang and Brainerd (2023)<br>Undorf et al. (2024) | Irrelevant when study items display varying degrees of difficulty. Explains lack of reactivity on pure lists. |
| Item-specific processing hypothesis | Senkova and Otani (2021)<br>Zhao, Li, et al. (2023) | Stevens and Pierce (2019)<br>Chang and Brainerd (2024) | Explains reactivity on single words. Inconsistent results. Possibly limited to organisation of a categorised list. |
| Attention-orienting or enhanced engagement account | Shi et al. (2023) | Rivers, Janes, et al. (2023) | Explains reactivity on visual memory. Includes learners' role. Needs more testing on verbal materials. No explanation for negative reactivity. |

*Note.* This table only includes the original theoretical frameworks. Some studies were not presented due to space limitations.

[a] These studies both support and contradict the changed-goal hypothesis in a series of different experiments with disparate results.

This leads to a conclusion that the mechanisms behind JOL reactivity may be specific to the type of information on which memory is being assessed (e.g., word pairs, single words), which makes all hypotheses worth investigating under varying circumstances. Therefore, while one hypothesis may suit one context, it might not be useful in another. Notably, the changed-goal hypothesis and the cue-strengthening hypothesis especially offset each other's limitations in the boundaries of their explanatory power. In that sense, there is room for JOL reactivity research to find the boundaries to which all four hypotheses specifically may be applicable.

## Practical and Theoretical Implications for JOL Reactivity Research

The previous sections suggest that the inconclusive evidence on JOL reactivity in the earlier studies was mostly driven by the lack of proper standardisation, for example, not controlling the time for task as a confounding variable or not including a judgement-free condition in the study. When using a well-established paired associates learning paradigm, JOL reactivity has been reliably found in majority of the studies, with robust positive reactivity being found for related pairs and either negative or no reactivity shown for unrelated pairs. Overall, whether JOL reactivity will occur is likely to depend on an interplay of various factors including, but not restricted to: type of study material (e.g., lists of single words), test format (e.g., free recall, cued recall), JOL type (e.g., item-level, list-level), participant population (e.g., young children, older adults) and their characteristics (e.g., one's motivation). Due to the complexity of JOL reactivity, further development is needed in regard to the hypotheses explaining its mechanism across various contexts. Future research should focus on developing methods to estimate their relative contribution in such settings (Rivers, Janes, et al., 2023). Specifically, Myers et al. (2020) suggested to use the tetrahedral model of memory by Jenkins (1979) to gain a better understanding about various moderators that contribute to JOL reactivity. This model includes four main factors that contribute to memory: stimuli (e.g., paired associates), participant (e.g., age), encoding (e.g., given instructions) and retrieval (e.g., test type).

Given the widespread use of JOLs in metacognitive research, the potential of JOLs to be reactive bears great significance to how researchers should approach studying the field. There is sufficient evidence to infer that JOL reactivity is undesirable when its effects should instead be controlled for: when

investigating JOL reactivity was not the primary aim of the study. The reactive effects of JOLs may unintentionally influence memory performance, serving as a potential confound for interpreting the causality of results to be strictly a consequence of an experimenters' manipulation and not the reactivity of the measure. To address the limitation of unforeseen effects of JOL reactivity, studies may need to employ an additional control condition to assess whether the measure alters the behaviour (Mitchum et al., 2016) or decide on other less reactive measure such as think-aloud-protocols (Fox et al., 2011). As concerns the JOL reactivity research itself, Zhao, Li, et al. (2023) proposed including an additional slider-rating task, unrelated to metamemory, in the non-JOL condition to better match the experimental JOL condition.

In light of the growing evidence supporting positive JOL reactivity, it may be appealing to assume that learners could reap the benefit of additional processing when monitoring their learning process using such metacognitive judgements as JOLs—at the very least, when learning simple verbal materials. However, given the recent influential findings that providing JOLs may well impair their memory performance (Undorf et al., 2024), this notion should be exercised with caution. This further highlights the need to examine JOLs in various contexts to uncover in which situations JOLs may boost versus hinder memory, and prevent learners from unintentionally jeopardising their test performance. Undoubtedly, more research is needed to be able to specify to which, if any, more complex and practical contexts positive JOL reactivity may be actually generalised, in order to utilise its potential in educational settings. Thus far, the vast majority of studies examining JOL reactivity on educational materials seem to imply that any statement regarding the possibilities of using JOLs as a viable learning intervention will be rather premature. However, it is worth noting that the influential studies such as Ariel et al. (2021) and its failed replications have not utilised JOLs in their standard form, but one that rates one's understanding rather than their actual prediction of recall. In particular, Hausman and Kubik (2023) argue that such judgements of comprehension are more based on familiarity and thus, lead to early memory search termination after potentially eliciting retrieval. Directions for future research may therefore include using a rating that more closely resembles JOLs on other educationally relevant materials, such as words in a foreign language, anatomical concepts, titled diagrams, reading assignments or actual exam materials. An investigation into JOL reactivity for different types of memory, for example, auditory memory, could also prove useful to determine the applicability of its effects.

As for the potential educational use of JOLs, Double et al. (2018) proposed that eliciting JOLs may facilitate learners' monitoring and decrease their chances of engaging in ineffective learning practices. The study by Shi et al. (2023) provides further evidence in support of this notion, as the researchers discovered that during prolonged periods of learning, making JOLs may aid in maintaining learning engagement throughout the learning period. This may especially be of use when considering the populations of low-performing learners who tend to misjudge their learning at monitoring (Dunning et al., 2003). Future directions for investigating JOL reactivity may therefore include learners with monitoring deficits (Double et al., 2018), which especially includes further testing on low-functioning older adults, who were found to use less successful learning strategies overall, and experience age-related deficit in processing (Tauber & Witherby, 2019). For example, an intervention concentrated on forming meaningful cue-target associations at both encoding and retrieval may improve memory for older adults and reduce their age-related associative deficits (Naveh-Benjamin et al., 2007).

To conclude, JOL reactivity research is still at the beginning of its journey. Due to its significant theoretical, but more so practical, implications for metacognitive research, it deserves further investigating.

## REFERENCES

Ackerman, B. P. (1986). The use of item-specific and relational episodic information in the recall of children and adults. *Journal of Experimental Child Psychology*, *42*(1), 115–143. https://doi.org/10.1016/0022-0965(86)90019-6

Ariel, R., Karpicke, J. D., Witherby, A. E., & Tauber, S. K. (2021). Do judgments of learning directly enhance learning of educational materials? *Educational Psychology Review*, *33*(2), 693–712. https://doi.org/10.1007/s10648-020-09556-8

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

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*, *17*(2), 443–476. https://doi.org/10.1007/s11409-021-09287-3

Chang, M., & Brainerd, C. J. (2023). Changed-goal or cue-strengthening? Examining the reactivity of judgments of learning with the dual-retrieval model. *Metacognition and Learning*, *18*, 183–217. https://doi.org/10.1007/s11409-022-09321-y

Chang, M., & Brainerd, C. J. (2024). Judgments of learning reactivity on item-specific and relational processing. *Journal of Intelligence*, *12*(1), Article 4. https://doi.org/10.3390/jintelligence12010004

de Winstanley, P. A., Bjork, E. L., & Bjork, R. A. (1996). Generation effects and the lack thereof: The role of transfer-appropriate processing. *Memory*, *4*(1), 31–48. https://doi.org/10.1080/741940667

Double, K. S. (2023). Do judgments of learning impair recall when uninformative cues are salient? *Journal of Intelligence*, *11*(10), Article 203. https://doi.org/10.3390/jintelligence11100203

Double, K. S., & Birney, D. P. (2019). Reactivity to measures of metacognition. *Frontiers in Psychology*, *10*, Article 2755. https://doi.org/10.3389/fpsyg.2019.02755

Double, K. S., Birney, D. P., & Walker, S. A. (2018). A meta-analysis and systematic review of reactivity to judgements of learning. *Memory*, *26*(6), 741–750. https://doi.org/10.1080/09658211.2017.1404111

Dougherty, M. R., Robey, A. M., & Buttaccio, D. (2018). Do metacognitive judgments alter memory performance beyond the benefits of retrieval practice? A comment on and replication attempt of Dougherty, Scheck, Nelson, and Narens (2005). *Memory & Cognition*, *46*(4), 558–565. https://doi.org/10.3758/s13421-018-0791-y

Dunlosky, J., & Hertzog, C. (1998). Training programs to improve learning in later adulthood: Helping older adults educate themselves. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 249–275). Lawrence Erlbaum Associates Publishers.

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

Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science*, *12*(3), 83–87. https://doi.org/10.1111/1467-8721.01235

Einstein, G. O., & Hunt, R. R. (1980). Levels of processing and organization: Additive effects of individual-item and relational processing. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(5), 588–598. https://doi.org/10.1037/0278-7393.6.5.588

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, *87*(3), 215–251. https://doi.org/10.1037/0033-295x.87.3.215

Fox, M. C., Ericsson, K. A., & Best, R. (2011). Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. *Psychological Bulletin*, *137*(2), 316–344. https://doi.org/10.1037/a0021663

Halamish, V., & Undorf, M. (2023). Why do judgments of learning modify memory? Evidence from identical pairs and relatedness judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *49*(4), 547–556. https://doi.org/10.1037/xlm0001174

Hausman, H., & Kubik, V. (2023). Delayed metacomprehension judgments do not directly improve learning from texts. *Journal of Intelligence*, *11*(7), Article 150. https://doi.org/10.3390/jintelligence11070150

Hunt, R. R. (2006). The concept of distinctiveness in memory research. In R. R. Hunt & J. B. Worthen (Eds.), *Distinctiveness and memory* (pp. 3–25). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195169669.003.0001

Hunt, R. R., & Einstein, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning & Verbal Behavior*, *20*(5), 497–514. https://doi.org/10.1016/s0022-5371(81)90138-9

Janes, J. L., Rivers, M. L., & Dunlosky, J. (2018). The influence of making judgments of learning on memory performance: Positive, negative, or both? *Psychonomic Bulletin & Review*, *25*(6), 2356–2364. https://doi.org/10.3758/s13423-018-1463-4

Jenkins, J. J. (1979). Four points to remember: A tetrahedral model of memory experiments. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory* (pp. 429–446). Lawrence Erlbaum Associates.

Kimball, D. R., & Metcalfe, J. (2003). Delaying judgments of learning affects memory, not metamemory. *Memory & Cognition*, *31*(6), 918–929. https://doi.org/10.3758/bf03196445

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., & 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

Kraus, S., Breier, M., Lim, W. M., Dabić, M., Kumar, S., Kanbach, D. K., Mukherjee, D., Corvello, V., Piñeiro-Chousa, J., Liguori, E., Marqués, D. P., Schiavone, F., Ferraris, A., Fernandes, C. I., & Ferreira, J. J. (2022). Literature reviews as independent studies: Guidelines for academic practice. *Review of Managerial Science*, *16*(8), 2577–2595. https://doi.org/10.1007/s11846-022-00588-8

Kubik, V., Koslowski, K., Schubert, T., & Aslan, A. (2022). Metacognitive judgments can potentiate new learning: The role of covert retrieval. *Metacognition and Learning*, *17*(3), 1057–1077. https://doi.org/10.1007/s11409-022-09307-w

Li, B., Shanks, D. R., Zhao, W., Hu, X., Luo, L., & Yang, C. (2024). Do changed learning goals explain why metamemory judgments reactively affect memory? *Journal of Memory and Language*, *136*, Article 104506. https://doi.org/10.1016/j.jml.2024.104506

Maxwell, N. P., & Huff, M. J. (2022). Reactivity from judgments of learning is not only due to memory forecasting: Evidence from associative memory and frequency judgments. *Metacognition and Learning*, *17*(2), 589–625. https://doi.org/10.1007/s11409-022-09301-2

Maxwell, N. P., & Huff, M. J. (2023). Is discriminability a requirement for reactivity? Comparing the effects of mixed vs. pure list presentations on judgment of learning reactivity. *Memory & Cognition*, *51*(5), 1198–1213. https://doi.org/10.3758/s13421-022-01381-4

Maxwell, N. P., & Huff, M. J. (2024). Judgment of learning reactivity reflects enhanced relational encoding on cued-recall but not recognition tests. *Metacognition and Learning*, *19*(1), 189–213. https://doi.org/10.1007/s11409-023-09369-4

Metcalfe, J., & Kornell, N. (2005). A region of proximal learning model of study time allocation. *Journal of Memory and Language*, *52*(4), 463–477. https://doi.org/10.1016/j.jml.2004.12.001

Mitchum, A. L., Kelley, C. M., & Fox, M. C. (2016). When asking the question changes the ultimate answer: Metamemory judgments change memory. *Journal of Experimental Psychology: General*, *145*(2), 200–219. https://doi.org/10.1037/a0039923

Myers, S. J., Rhodes, M. G., & Hausman, H. E. (2020). Judgments of learning (JOLs) selectively improve memory depending on the type of test. *Memory & Cognition*, *48*(5), 745–758. https://doi.org/10.3758/s13421-020-01025-5

Naveh-Benjamin, M., Brav, T. K., & Levy, O. (2007). The associative memory deficit of older adults: The role of strategy utilization. *Psychology and Aging*, *22*(1), 202–208. https://doi.org/10.1037/0882-7974.22.1.202

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), 267–270. https://doi.org/10.1111/j.1467-9280.1991.tb00147.x

Nelson, T. O., & Narens, O. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 125–173). Academic Press. https://doi.org/10.1016/S0079-7421(08)60053-5

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., & 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

Rivers, M. L., Dunlosky, J., Janes, J. L., Witherby, A. E., & Tauber, S. K. (2023). Judgments of learning enhance recall for category-cued but not letter-cued items. *Memory & Cognition*, *51*(7), 1547–1561. https://doi.org/10.3758/s13421-023-01417-3

Rivers, M. L., Janes, J. L., & Dunlosky, J. (2021). Investigating memory reactivity with a within-participant manipulation of judgments of learning: Support for the cue-strengthening hypothesis. *Memory*, *29*(10), 1342–1353. https://doi.org/10.1080/09658211.2021.1985143

Rivers, M. L., Janes, J. L., Dunlosky, J., Witherby, A. E., & Tauber, S. K. (2023). Exploring the role of attentional reorienting in the reactive effects of judgments of learning on memory performance. *Journal of Intelligence*, *11*(8), Article 164. https://doi.org/10.3390/jintelligence11080164

Schäfer, F., & Undorf, M. (2024). On the educational relevance of immediate judgment of learning reactivity: No effects of predicting one's memory for general knowledge facts. *Journal of Applied Research in Memory and Cognition*, *13*(1), 113–123. https://doi.org/10.1037/mac0000113

Seli, P., Beaty, R. E., Cheyne, J. A., Smilek, D., Oakman, J., & Schacter, D. L. (2018). How pervasive is mind wandering, really? *Consciousness and Cognition*, *66*, 74–78. https://doi.org/10.1016/j.concog.2018.10.002

Senkova, O., & Otani, H. (2021). Making judgments of learning enhances memory by inducing item-specific processing. *Memory & Cognition*, *49*(5), 955–967. https://doi.org/10.3758/s13421-020-01133-2

Shi, A., Xu, C., Zhao, W., Shanks, D. R., Hu, X., Luo, L., & Yang, C. (2023). Judgments of learning reactively facilitate visual memory by enhancing learning engagement. *Psychonomic Bulletin & Review*, *30*(2), 676–687. https://doi.org/10.3758/s13423-022-02174-1

Soderstrom, N. C., Clark, C. T., Halamish, V., & Bjork, E. L. (2015). Judgments of learning as memory modifiers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(2), 553–558. https://doi.org/10.1037/a0038388

Spellman, B. A., & Bjork, R. A. (1992). When predictions create reality: Judgments of learning may alter what they are intended to assess. *Psychological Science*, *3*(5), 315–316. https://doi.org/10.1111/j.1467-9280.1992.tb00680.x

Stevens, A. R., & Pierce, B. H. (2019, November 14–17). *Do reactive effects of judgments of learning extend to words lists?* [Poster session]. 2019 Annual Meeting of the Psychonomic Society, Montreal, QC, Canada.

Tauber, S. K., & Witherby, A. E. (2019). Do judgments of learning modify older adults' actual learning? *Psychology and Aging*, *34*(6), 836–847. https://doi.org/10.1037/pag0000376

Tekin, E., & Roediger, H. L. III. (2020). Reactivity of judgments of learning in a levels-of-processing paradigm. *Zeitschrift Für Psychologie*, *228*(4), 278–290. https://doi.org/10.1027/2151-2604/a000425

Thompson, W. B. (1999). Individual differences in memory-monitoring accuracy. *Learning and Individual Differences*, *11*(4), 365–376. https://doi.org/10.1016/s1041-6080(99)80009-0

Undorf, M., Schäfer, F., & Halamish, V. (2024). Making judgments of learning either enhances or impairs memory: Evidence from 17 experiments with related and unrelated word pairs. *Collabra: Psychology*, *10*(1), Article 117108. https://doi.org/10.1525/collabra.117108

Veenman, M. V., Van Hout-Wolters, B. H., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning*, *1*, 3–14. https://doi.org/10.1007/s11409-006-6893-0

Witherby, A. E., Babineau, A. L., & Tauber, S. K. (2023). Does interactive imagery influence the reactive effect of judgments of learning on memory? *Journal of Intelligence*, *11*(7), Article 139. https://doi.org/10.3390/jintelligence11070139

Witherby, A. E., & Tauber, S. K. (2017). The influence of judgments of learning on long-term learning and short-term performance. *Journal of Applied Research in Memory and Cognition*, *6*(4), 496–503. https://doi.org/10.1016/j.jarmac.2017.08.004

Zhao, W., Li, B., Shanks, D. R., Zhao, W., Zheng, J., Hu, X., Su, N., Fan, T., Yin, Y., Luo, L., & Yang, C. (2022). When judging what you know changes what you really know: Soliciting metamemory judgments reactively enhances children's learning. *Child Development*, *93*(2), 405–417. https://doi.org/10.1111/cdev.13689

Zhao, W., Li, J., Shanks, D. R., Li, B., Hu, X., Yang, C., & Luo, L. (2023). Metamemory judgments have dissociable reactivity effects on item and interitem relational memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *49*(4), 557–574. https://doi.org/10.1037/xlm0001160

Zhao, W., Xu, M., Xu, C., Li, B., Hu, X., Yang, C., & Luo, L. (2023). Judgments of learning following retrieval practice produce minimal reactivity effect on learning of education-related materials. *Journal of Intelligence*, *11*(10), Article 190. https://doi.org/10.3390/jintelligence11100190