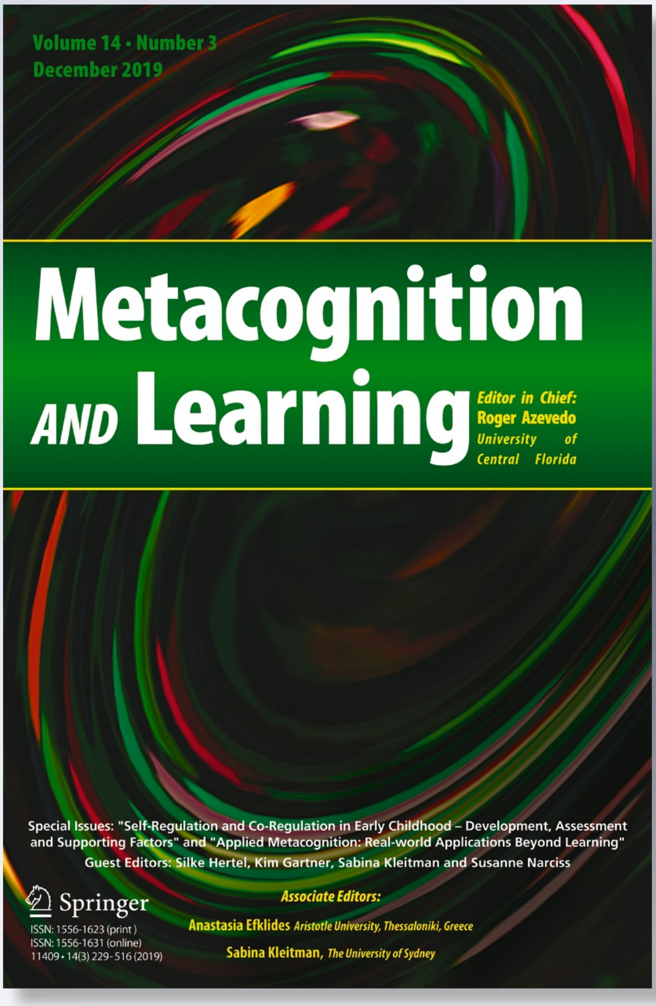
Confidence judgments: The monitoring of object-level and same-level performance

Asher Koriat

Metacognition and Learning

ISSN 1556-1623
Volume 14
Number 3

Metacognition Learning (2019)
14:463-478
DOI 10.1007/s11409-019-09195-7


Volume 14 • Number 3
December 2019
Metacognition
AND Learning
Editor in Chief:
Roger Azevedo
University of Central Florida
Special Issues: "Self-Regulation and Co-Regulation in Early Childhood – Development, Assessment and Supporting Factors" and "Applied Metacognition: Real-world Applications Beyond Learning"
Guest Editors: Silke Hertel, Kim Gartner, Sabina Kleitman and Susanne Narciss
Springer
ISSN: 1556-1623 (print)
ISSN: 1556-1631 (online)
11409 • 14(3) 229-516 (2019)
Associate Editors:
Anastasia Efklides Aristotle University, Thessaloniki, Greece
Sabina Kleitman, The University of Sydney

Springer

Metacognition and Learning (2019) 14:463–478
https://doi.org/10.1007/s11409-019-09195-7

Confidence judgments: The monitoring of object-level and same-level performance

Asher Koriat[1]

Received: 23 October 2018 / Accepted: 26 June 2019 / Published online: 27 July 2019

Abstract

The influential metacognitive framework of Nelson and Narens (1990) distinguishes between *object-level* and *meta-level*, with two metacognitive processes, monitoring and control, governing the interplay between them. Monitoring refers to the process by which the meta-level tracks the accuracy of object level-performance, whereas control refers to the processes by which the meta-level regulates object-level processes. In this study, I examine the prediction derived from Koriat's (*Psychological Review, 119*, 80–113 2012a) self-consistency model (SCM) that when people indicate their confidence in the accuracy of their choice, their confidence actually monitors the likelihood that others will make the same choice better than the accuracy of that choice. This was shown to be the case for three levels of processing: choosing the correct option, predicting the choice made by others, and predicting the predictions made by others about the majority choice. The conditions under which object-level correspondence and same-level correspondence are aligned or diverge are discussed.

Keywords Confidence judgments · Monitoring · Predicting others · Self-consistency model · Meta-object correspondence · Same-level correspondence

In their influential conceptual framework of the relationship between metacognition and cognition, Nelson and Narens (Nelson and Narens 1990) distinguished between two interrelated levels that they called *object-level* and *meta-level* (see Fig. 1). The object-level includes basic information processing operations that are involved in encoding, learning, and remembering. The meta-level, in turn, contains a model that the person has of the task and of the cognitive operations involved in performing it. The interplay between the object-level and the meta-level was assumed to involve two higher-order, metacognitive processes: monitoring and

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11409-019-09195-7) contains supplementary material, which is available to authorized users.

✉ Asher Koriat
akoriat@univ.haifa.ac.il

[1] Department of Psychology and Institute of Information Processing and Decision Making, University of Haifa, 3498838 Haifa, Israel

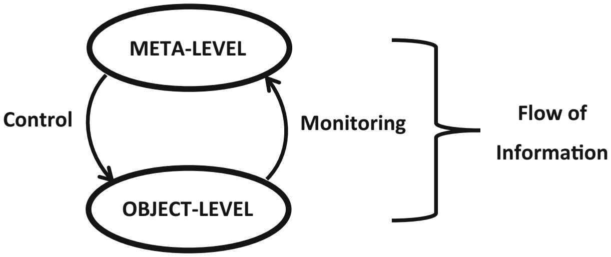


Fig. 1 Nelson and Narens' conceptual model of the distinction between meta-level and object-level, and the flow of information between them (Nelson and Narens 1990)

control. Monitoring refers to the process by which the metal-level is informed about what is occurring at the object-level, whereas control refers to the mechanisms by which the meta-level regulates the operation of object-level processes towards the achievement of different goals. This conceptual framework has proved very useful in driving research on the monitoring and self-regulation processes that occur during learning, remembering and deciding (Ariel et al. 2009; Benjamin 2008; Koriat and Goldsmith 1996; see Dunlosky and Metcalfe 2009).

In this study, I focus on resolution or relative accuracy (see Dunlosky and Metcalfe 2009; Koriat 2016; Lichtenstein et al. 1982). Resolution refers to the extent to which metacognitive judgments predict inter-item differences in performance. In particular, research has suggested that people are generally skilled at monitoring the accuracy of their knowledge across items, and that their reliance on the output of their monitoring in regulating their behavior is generally beneficial. For example, judgments of learning (JOL) elicited in the course of studying new material are relatively accurate in predicting recall performance (see Rhodes 2016). In turn, results suggest that learners rely on JOLs in choosing which items to restudy and in allocating study time differentially to different items in a list (Dunlosky and Hertzog 1998; Metcalfe 2009; Metcalfe and Finn 2008; Son and Metcalfe 2000; Thiede and Dunlosky 1999). Manipulations that enhance monitoring accuracy were found to improve the effectiveness of study time allocation and in turn, to improve overall recall performance (Thiede et al. 2003). Similarly, confidence in an answer is generally diagnostic of the accuracy of that answer (see Koriat 2012a). Participants were found to rely heavily on their confidence in the answer in deciding whether to volunteer it, thereby increasing the overall accuracy of the information that they do report (Benjamin 2008; Goldsmith and Koriat 2008).

In the present study, I focus on subjective confidence in the response to two-alternative forced-choice (2AFC) items. In a typical experiment, participants are presented with a 2AFC question, for example, "what is the capital of Australia, (a) Sydney, (b) Canberra?" They are asked to choose the correct answer ("first-order" judgment), and then to indicate their confidence in the correctness of that answer ("second-order" judgment). In the Nelson and Narens (1990) framework, monitoring represents the process by which the meta-level is informed about the object-level. Hence, monitoring is captured by the relationship between the second-order judgments and the accuracy of first-order judgments. In the present study, however, I review evidence indicating that confidence judgments actually predict better the first-order responses made by the majority of other participants who perform the same task. Specifically, they predict the likelihood that other participants will make the same response. This is so despite the fact that confidence judgments are targeted explicitly at the accuracy of the respective object-level response.

I shall refer to these two types of relationship as Meta-Object (M-O) correspondence and Same-Level (S-L) correspondence, respectively. Figure 2 depicts the difference between them.

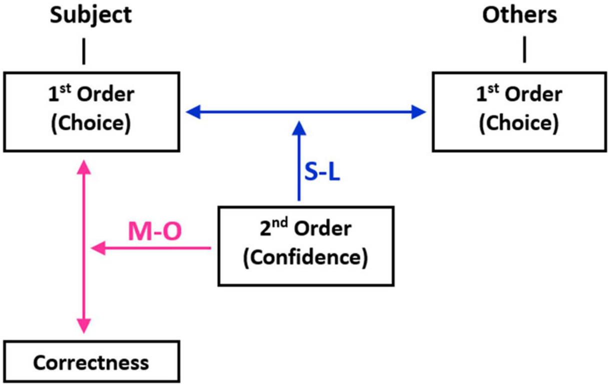


Fig. 2 Monitoring Object-Level (M-O) and Same-Level (S-L) Correspondence

M-O correspondence captures what Nelson and Narens (1990) refer to as monitoring: The correspondence between my second-order judgment (confidence) and the accuracy of my first-order judgment (choice). In other words, it refers to the extent to which confidence tracks the agreement between my first-order judgment and some criterion of correctness. S-L correspondence, in contrast, refers to the extent to which my second-order judgment (confidence) tracks the agreement between my first-order judgment and the first-order judgment made by the majority of others. Thus, the criterion here is the first-order judgment made by others who are presented with the same task.

Note that although Nelson and Narens (1990) distinguished only between an object-level and a meta-level, Nelson and Narens (1994) expanded their scheme to include a multi-level organization. According to them, what is critical is the dominance relation whereby higher-order processes dominate lower-order processes so that monitoring always involves obtaining information about the processes that occur at a lower level.

Three levels of processing

In this article, I review results pertaining to three levels of processing, examining the possibility that regardless of the level of processing involved, confidence judgments predict S-L correspondence better than M-O correspondence.

Let us consider the following situation: There are three groups of participants, each performing a task that represents one of three levels of processing. Participants in Level A are presented with 2AFC items for which the answer can be scored as correct or wrong. They are instructed to choose the correct answer to each item and to indicate their confidence in the correctness of their choice.

In Level B, participants are required to predict the responses of Level-A participants: They are asked to predict for each item which of the two response options is the more likely to be endorsed by Level-A participants, and to indicate their confidence in the correctness of their prediction.

Participants in Level C, in turn, are required to predict the predictions of Level-B participants. For each item, they predict which of the two predictions would be made by the majority of Level-B participants. They also indicate their confidence in the accuracy of their predictions of the majority prediction made by Level-B participants.

I examine the proposition that confidence judgments exhibit the same pattern of relationships for the three levels of processing: In each level, people's confidence judgments predict the likelihood that other people will make the same judgment/prediction better than they predict the correctness of that judgment/prediction (see Fig. 3). Level B processing is important in many domains in which people need to predict the views, and attitudes of others (Barr and Keysar 2005). The prediction of others' predictions (Level C processing), in turn, has received particular interest in economics, in connection with Keynes' (1936) beauty contest analogy of equity markets. Investment decisions are said to depend not only on one's own predictions of market developments but also on what one thinks other people predict, because their predictions may influence their own investment policy. Predictions in general have been found to exhibit many biases and errors (see Dunning 2007), and it is of interest to show that they also yield stronger S-L correspondence than M-O correspondence.

What is the rationale for our proposal? Koriat's (2012a) Self-Consistency Model (SCM) assumes that confidence judgments are based on the reliability of a choice as a proxy for its validity: When people are required to choose between two response options, for example, between two answers to a general-knowledge question, they retrieve a small sample of cues sequentially from a population of cues associated with the item, draw the implications of each cue, and choose the answer that is supported by the largest number of cues (see Baranski and Petrusic 1998). Confidence in the answer is based on the consistency with which that answer is supported across the retrieved cues (see Alba and Marmorstein 1987; Armelius 1979; Brewer and Sampaio 2012; Slovic 1966). In addition, because the sampling of cues is terminated when several cues in a row support the same answer, responses become faster as self-consistency increases.

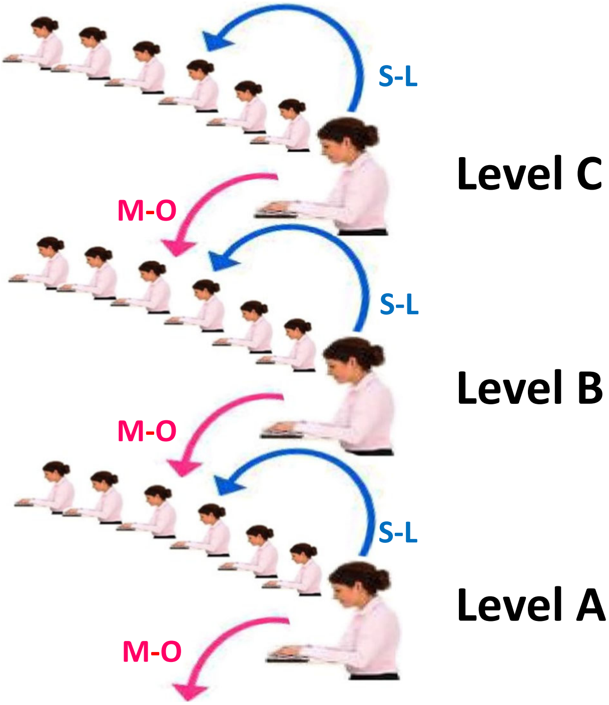


Fig. 3 Monitoring Object-Level (M-O) and Same-Level (S-L) Correspondence for Three Processing Levels

Koriat (2012a) proposed that the population of potential cues associated with each item is largely shared by participants with the same background (see Koriat and Adiv 2016, for supporting evidence). Therefore, the confidence with which an answer is chosen, and the speed with which that answer is chosen, should increase with the consensuality of that answer – the likelihood that that answer will be chosen across participants.

A simulation experiment incorporating these assumptions yielded evidence for what Koriat et al. (2016) termed a Prototypical Majority Effect (PME): Responses that are endorsed by the majority of participants are associated with higher confidence and shorter response latency than minority responses, with the majority-minority differences in confidence and response speed increasing with the size of the majority. Indeed, a number of studies reviewed by Koriat et al. 2016 (see also Koriat and Adiv 2016) yielded a PME pattern for several different tasks.

The hypothesis tested in this article is particularly interesting in view of two observations. First, people tend to change their response and confidence when presented repeatedly with the same item. However, the response that they endorse with greater confidence is the more likely to be chosen by others. Second, recent evidence suggests that participants' confidence in their response predicts the majority response even when participants have no idea what other people choose, and even when they are wrong in predicting the majority response (Koriat et al. 2018). However, these results are consistent with SCM, assuming that S-L correspondence derives from people's tendency to sample their cues largely from a consensually shared population of cues.

In this article, I compare directly M-O correspondence with S-L correspondence for the three levels of processing mentioned earlier, examining the idea that confidence judgments tend to track same-level performance better than object-level performance: They monitor better the likelihood that other participants will make the same response than the likelihood that that response is correct. First, I review evidence indicating that such is the case for level A processing. Second, I present results that generalize this idea to level B and level C processing. Finally, I discuss the conditions under which this pattern of results is expected to emerge.

For Level-A participants, M-O correspondence is typically indexed by the Goodman-Kruskal gamma correlation (see Nelson 1984), which is a rank-order correlation between confidence and accuracy – whether the response is correct (scored as 1) or wrong (scored as 0). In a similar manner, S-L correspondence can be indexed by the gamma correlation between confidence in a response and the consensuality of that response – whether that response is the majority response (scored as 1) or the minority response (scored as 0).

In addition to confidence judgments, I also examine the results for response latency. Previous research has indicated that response speed is also diagnostic of object-level accuracy so that correct responses are made faster than incorrect responses (Ackerman and Koriat 2011; Robinson et al. 1997; Weidemann and Kahana 2016). However, I examine the possibility that response speed is also more diagnostic of same-level performance than of object-level performance.

In recent years, alternative measures to the gamma correlation have been proposed, derived from signal detection theory (SDT, Benjamin and Diaz 2008; Fleming and Lau 2014; Higham et al. 2009). These Type-2 SDT measures evaluate the accuracy of second-order judgments in parallel to the standard, Type-1 SDT measures that evaluate the accuracy of first-order judgments. In what follows, in addition to the gamma correlation, I will also use the Meta *d'* index proposed by Maniscalco and Lau (2014) in comparing M-O and S-L correspondence.

The results to be presented below are based on a reanalysis of previously reported data, taking advantage of data that permit a direct comparison between M-O and S-L correspondence, or on new experiments designed to examine that comparison. The main purpose of this article is to bring to the fore the idea that although confidence judgments are targeted explicitly at the accuracy of object-level performance, they actually track better same-level performance (Fig. 2).

Monitoring the accuracy of one's own responses

Let us consider the typical, level-A situation. As noted earlier, many studies have indicated that people are skilled at monitoring the accuracy of their answers and judgments for 2AFC tasks in a variety of domains. For these tasks, the confidence/accuracy (C/A) correlation, tapping M-O correspondence, is positive and sometimes quite high. Similarly, response speed is also generally diagnostic of accuracy.

Results reviewed by Koriat (2018), however, suggest that the C/A correlation is positive for consensually-correct (CC) items, for which most participants tend to choose the correct answer. However, when consensually-wrong (CW) items are used, for which most people tend to choose the wrong answer, the C/A correlation is actually *negative*: People are more confident when they are wrong than when they are right. This pattern was observed for a word-matching task (Koriat 1976), general-knowledge questions (Koriat 2008b), perceptual judgments (Koriat 2011), judgments of geographical relations (Koriat 2017), memory recognition judgments (DeSoto and Roediger III 2014; Kurdi et al. 2018), face recognition (Sampaio et al. 2017) and syllogistic reasoning (Bajšanski et al. 2019). Brewer and his associates also reported a negative C/A relationship for different types of so called "deceptive" items - those that tend to yield erroneous responses across participants (Brewer and Sampaio 2006; Brewer and Sampaio 2012; Brewer et al. 2005; Sampaio and Brewer 2009).

These results were taken by Koriat (2018) to imply that the positive C/A correlation that has been reported in many studies is an artifact of the fact that in these studies the accuracy of *first-order* judgments was much better than chance. This would be expected in view of people's adaptation to reality through evolution and learning. Thus, in terms of Koriat's (2012a) classification, the items used in most studies are mostly CC-type items. For example, for 2AFC general-knowledge questions drawn representatively from their reference classes, the percentage of correct answers is around 75% (see Koriat 2018). Similarly, a word that is retrieved from a studied list is much more likely to be correct than wrong. This is true even for studies that used DRM lists (Roediger III and McDermott 1995). In these studies, it was found that an item recalled had about a .90 probability of being correct (Koriat et al. 2011). Thus, even if no deliberate attempts are made to sample items representatively from their domain (see Gigerenzer et al. 1991; Juslin 1994), for many of the items used in most studies, people's responses are more likely to be correct than wrong, and for these items the C/A correlation would be expected to be positive.

The results indicating a positive C/A correlation for CC items but a negative correlation for CW items were described by Koriat (2008b; 2012a) in terms of the consensuality principle: Confidence judgments are correlated with the consensuality of the response irrespective of its accuracy. In terms of the proposal advanced in the present article (see Fig. 2), confidence judgments would seem to predict same-level performance, sometimes better than object-level performance. Here, however, I wish to compare directly M-O correspondence and S-L correspondence and to examine the conditions under which the two types of correspondence are aligned or diverge.

Experimental results

Level-a processing

I begin by examining the results for level-A processing, comparing the relative strength of the M-O and S-L relationships. To do so, I had to focus on studies that included a sufficiently large number of CW items. In Study 1, I examined what happens when 2AFC items are selected systematically to cover the full range of Object-Level Accuracy (OLA, the percentage of correct answers). To do so, I took advantage of the results of Koriat (2018) in which 120 participants performed five tasks requiring binary decisions, and indicated their confidence in each decision. One of these tasks involved Level B processing (which will be examined later). Here, however, I focus on the remaining four tasks, which involved Level-A processing. These tasks included general-information questions (24 CC items and 24 CW items), judgments of line lengths (7 CC and 7 CW items), judgments of the area of geometric shapes (15 CC and 15 CW items), and geography questions (8 CC and 8 CW items). These tasks were designated as 1, 2, 3, and 5, respectively in Koriat (2018, Table 2). Altogether there were 54 CC items, and 54 CW items that were matched in terms of item consensus – the percentage of participants choosing the consensual answer for each item. Thus, the percentage of correct answers averaged 68.41% for the CC items and 30.83% for the CW items.

As summarized in Table 1 (Study 1), the within-person C/A gamma correlation across the 108 items averaged only .05 [although it was still significant, $t(119) = 3.36$, $p < .005$]. In contrast, when each person's choice was scored in terms of its agreement with the majority choice, the gamma correlation averaged .25. This correlation was significant, $t(119) = 17.69$, $p < .0001$, and was also significantly higher than the C/A correlation (see Table 1). In fact, of the 120 participants, 102 yielded a pattern in which the S-L gamma correlation was higher than the M-O correlation, $p < .0001$ by a binomial test.

Table 1 also presents the respective results for response latency. The negative gamma correlation for S-L correspondence, was significant, $t(119) = 10.73$, $p < .0001$, and was significantly higher than the respective correlation for M-O correspondence. As can be seen in Table 1, the Meta *d*' index yielded the same pattern of results for both confidence and latency.

Note that in Study 1, OLA averaged 49.62 for the 108 items. Thus, across a set of items for which the majority response is not biased in favor of either the correct or the wrong answer, confidence judgments and response latency exhibited higher S-L correspondence than M-O correspondence.

Level-B processing

I turn next to Level B tasks. In these tasks, participants are asked to predict for each 2AFC item which of the two options is the one likely to be chosen by the majority of other participants.

I propose that even for level B processing, S-L correspondence can be higher than M-O correspondence. The rationale for this proposal is that in making a prediction about other people's choices, participants also sample cues that speak for one of the two predictions, and their confidence is based on the reliability with which the cues favor one of the two options. The cues sampled need not be the same as those involved in level-A processing. Tullis (2018), for example, obtained results suggesting that when participants make predictions about others' knowledge, they tend to rely on cues about their own knowledge, but the extent of that reliance differs for different judgment conditions. Assuming that the cues underlying predictions are

Table 1 Gamma and Meta d' for Meta-Object (M-O) and Same-Level (S-L) correspondence, and the t-test difference between them for confidence and response latency for Studies 1 to 5

Study	Variable	Measure	M-O	S-L	*t*-test
1	Confidence	Gamma	.05	.25	*t*(119) = 10.10, *p* < .0001
		Meta *d*’	.17	.69	*t*(119) = 7.87, *p* < .0001
	Response Latency	Gamma	+.01	−.14	*t*(119) = 8.79, *p* < .0001
		Meta *d*’	+.09	−.44	*t*(119) = 9.00, *p* < .0001
2	Confidence	Gamma	.29	.44	*t*(40) = 5.29, *p* < .0001
		Meta *d*’	.98	1.49	*t*(40) = 3.53, *p* < .01
	Response Latency	Gamma	−.17	−.25	*t*(40) = 3.00, *p* < .005
		Meta *d*’	−.50	−.82	*t*(40) = 2.83, *p* < .01
3	Confidence	Gamma	.03	.27	*t*(117) = 5.14 *p* < .0001
		Meta *d*’	0.12	1.09	*t*(119) = 5.13 *p* < .0001
	Response Latency	Gamma	+.002	−.18	*t*(119) = 4.48, *p* < .0001
		Meta *d*’	+.24	−.43	*t*(119) = 5.02, *p* < .0001
4	Confidence	Gamma	.003	.21	*t*(40) = 5.53, *p* < .0001
		Meta *d*’	.01	.59	*t*(40) = 6.64, *p* < .0001
	Response Latency	Gamma	+.01	−.12	*t*(40) = 3.36, *p* < .005
		Meta *d*’	−.05	.09	*t*(40) = 1.54, *p* = .14
5	Confidence	Gamma	.17	.25	*t*(30) = 2.83, *p* < .01
		Meta *d*’	−.11	.73	*t*(30) = 8.17, *p* < .0001
	Response Latency	Gamma	−.09	−.13	*t*(30) = 1.49, *p* = .16
		Meta *d*’	.02	-.40	*t*(30) = 3.86, *p* < .001

M-O correspondence refers to the extent to which confidence and response speed track the accuracy of the first-order response. S-L correspondence refers to the extent to which confidence and response speed track the likelihood that other participants will make the same first-order response

also sampled from a consensually-shared population of cues, confidence in predictions may also be expected to yield higher S-L correspondence than M-O correspondence.

Results reported by Koriat (2018) are consistent with this hypothesis. It was proposed that the positive C/A correlation that has been observed for the prediction of others’ responses is also due to the fact that people’s predictions are largely correct. Indeed, the analysis of 6 studies involving the prediction of others’ responses (Studies 6–11 in Project 1, Koriat 2018) indicated that for CC items, for which people’s *predictions* were correct, confidence was higher for the correct predictions than for the wrong predictions, whereas for a minority of items (CW), for which people’s predictions tended to be wrong, confidence was actually higher for the wrong predictions than for the correct predictions.

In Study 2, I compared directly M-O and S-L correspondence for Level-B processing, using the data of Koriat (2013). In that study, participants were presented with 60 2AFC items that measure personal preferences, for example, “Which sport activity would you prefer? (a) Jogging, (b) swimming”. For Blocks 1–5 (Self), participants chose for each question the option that reflects best their own preference, and indicated their confidence. In Block 6 (Other), they were asked to predict which of the two options would be preferred by the majority of other participants and to indicate their confidence in their prediction.

I first determined the majority preference for each item across Blocks 1–5 and used it as a criterion for assessing the accuracy of participants’ predictions in Block 6, scoring each prediction as correct (1) or wrong (0). Gamma correlation was then calculated between confidence in Block 6 and prediction accuracy. This correlation averaged .29 across participants, and was significant, $t(40) = 10.26$, $p <$ .0001. The respective correlation for response latency was −.17*,* $t(40) = 5.50$, $p <$ .0001.

To assess S-L correspondence, I first determined the prediction made by the majority of participants for each item in Block 6, and then scored each participant's prediction for its agreement with the consensual prediction (1 = agreed with it, 0 = disagreed). Gamma was then calculated between confidence in Block 6 and the agreement of the prediction with the consensual prediction. This correlation averaged .44 across participants. The respective correlation for response latency was −.25. As can be seen in Table 1, S-L correspondence was significantly higher than M-O correspondence for both confidence and response speed.

I used the same procedure in calculating Meta *d*'. The results yielded the same general pattern (see Table 1).

Thus, even for Level B processing, confidence and response latency were more strongly correlated with same-level performance than with object-level performance. This was true despite the fact that the percentage of correct predictions averaged 72.96% across items in Study 2, so that participants' predictions were overall more likely to be correct than wrong.

To obtain a comparison between M-O and S-L correspondence for a sample of items that is not biased in favor of the correct predictions, I took advantage of the results for a task included in Koriat (2018; Task 4 in Project 2), which was excluded from the analysis reported earlier. That task also involved predictions of personal preferences, but included 8 CC and 8 CW items roughly matched in terms of item consensus (averaged 51.04% across all items).

The analyses were based on 118 participants (the results of two participants who gave the same confidence judgments to all 8 items were eliminated). It can be seen (Table 1, Study 3) that the M-O gamma correlations for both confidence and response latency were very close to zero, and the S-L correlations were significantly higher than the respective M-O correlations. Meta *d*' for M-O correspondence was not significant for either confidence or response latency, $t(119) = 1.06$, $p = .30$, and $t(119) = 1.86$, $p = 08$, respectively. However, S-L correspondence was significant for both confidence, $t(119) = 6.69$, $p < .0001$, and response latency, $t(119) = 10.52$, $p < .0001$. For both confidence and response latency, S-L correspondence was also significantly higher than M-O correspondence (see Table 1).

I conducted two additional experiments in which I used a word-association task that allowed examination of Level B processing (Study 4) as well as Level C processing (Study 5) for the same stimuli. In Study 4, participants were given instructions describing the word association task in which people, who are presented with a stimulus word, are asked to respond with the first word that comes to mind. They were told that that they would be presented with several stimulus words, and their task is to predict for each stimulus word which of two response words participants are more likely to give as the first response to that stimulus word. The materials (in Hebrew) included 60 stimulus words and two potential word associates for each stimulus word, which differed in their associative strength according to Hebrew word association norms (Rubinsten et al. 2005). The two potential associates were selected on the basis of previous findings (e.g., Koriat and Bjork 2005; Koriat and Bjork 2006; Koriat et al. 2006) with the intention to yield a sufficient number of items for which participants' predictions of the stronger association would be likely to be wrong when compared to the Hebrew norms.

The accuracy of participants' predictions of the dominant association averaged 45.16%. With regard to M-O correspondence, the average gamma correlation for both confidence and latency were around .0. In contrast, the respective correlations assessing S-L correspondence were significant, for both confidence, $t(40) = 8.21$, $p < .0001$, and response latency, $t(40) = 4.39$, $p < .0001$, and both were significantly higher than the respective M-O correlations. The Meta *d* index was also around .0 for both confidence and response latency, but only for confidence was the S-L correspondence significantly higher than the M-O correspondence (see Table 1).

Level-C processing

I turn next to level C processing, which involves predicting the predictions made by others. In study 5, I examined the prediction of others' predictions using the same materials as those used in Study 4. However, participants' task was to predict the predictions made by Study-4 participants. Participants were given a brief description of the procedure used in Study 4. They were told that their task was to guess for each item which of the two response words had been judged by the majority of Study-4 participants as the one most likely to be given as the first response to the stimulus word. (For details of the Method see Supplemental Materials).

The M-O gamma correlation was significant for both confidence judgments, $t(30) = 5.14$, $p < .0001$, and response latency, $t(30) = 3.05$, $p < .005$. The respective S-L correlation was significantly higher, but only for confidence; not for response latency. The Meta *d* index was not significant for M-O correspondence, either for confidence, $t(30) = 1.90$, $p = .08$, or for response latency, $t(30) = 0.36$, $p = .73$, and it was significantly higher for S-L correspondence than for M-O correspondence for both confidence and response latency.

In sum, the results on the whole yielded a consistent pattern in which S-L correspondence tended to be higher than M-O correspondence. This pattern was largely found for the three levels of processing investigated. This was so for both confidence judgments and response speed using either gamma or Meta *d'* as measures of correspondence.

Discussion

Confidence judgments have been used extensively in many different contexts. Typically, participants are asked to make a binary decision, and to indicate their degree of confidence in that decision. When the decision has a truth value, confidence is targeted specifically at the accuracy of that decision so that monitoring accuracy is appropriately evaluated in terms of the correspondence between confidence and the accuracy of the object-level (first-order) decision. In the present study I focused on one aspect of this correspondence – resolution or relative accuracy (Dunlosky and Metcalfe 2009; Koriat 2016; Lichtenstein et al. 1982). Resolution, which refers to the extent to which confidence discriminates between correct and wrong decisions, has been claimed to be critical for effective self-regulation because people generally rely on their confidence in a belief in deciding whether to act on that belief (Ackerman and Goldsmith 2011; Ariel et al. 2009; Gill et al. 1998; Goldsmith and Koriat 2008; Thiede and Dunlosky 1999; Tullis and Benjamin 2011). People also take into account experts' confidence in their advice in deciding whether to utilize that advice (Van Swol and Sniezek 2005), and more generally, the persuasiveness of a communication increases with the confidence with which it is expressed (Koriat 2012b; Koriat 2015; Pulford et al. 2018).

The conceptual scheme proposed by Nelson and Narens (1990) provided a useful framework for research on the monitoring and control processes that take place in the self-management of cognitive processes and behavior. This framework has been very influential (see Dunlosky and Metcalfe 2009) and several observations are consistent with the overall dynamics postulated in this framework. First, many studies have confirmed people's ability to monitor the accuracy of their beliefs and judgments: People generally endorse correct answers with higher confidence than wrong answers (see Koriat 2018). Second, consistent with the "monitoring-affects-control" hypothesis (Nelson and Leonesio 1988), people tend to rely, sometimes heavily, on their metacognitive judgments in the strategic regulation of their cognitive processes and behavior (Benjamin 2008; Goldsmith and Koriat 2008; Son and Metcalfe 2000). Finally, several studies indicate that participants' performance generally benefits from using the output of monitoring as a basis for regulation (Koriat and Goldsmith 1996; Metcalfe and Kornell 2003; Thiede et al. 2003).

Of course, many studies have demonstrated dissociations between metacognitive judgments and object-level performance (Benjamin and Bjork 1996; Benjamin et al. 1998; Bjork et al. 2013; Brewer and Sampaio 2012; Chandler 1994; Kelley and Lindsay 1993; Koriat 1995; Rhodes and Castel 2008; Roediger and DeSoto 2015). These demonstrations were obtained usually as part of the attempt to uncover the bases of metacognitive judgments. The work of Koriat (Koriat 2012a; Koriat 2018) on confidence judgments supplements these demonstrations in two important respects. First, that work shows that changes in the distribution of the items in terms of OLA produce C/A correlations that can vary all the way from a high positive correlation to a high negative correlation. Second, as the present study tried to show, confidence judgments actually predict better a different property than the property that these judgments are purported to monitor. Indeed, the results were quite consistent in showing that confidence judgments predict better S-L correspondence than M-O correspondence across the three processing levels investigated in this study. A similar pattern of results was obtained for response speed.

These results are consistent with the assumption of SCM that subjective confidence in a choice is based on the reliability with which that choice is supported across the cues consulted in making the choice. Assuming that these cues are sampled from a population of cues that is largely shared across participants, confidence in a choice should correlate with inter-subject consensus in making that choice.

What are the implications of these results for the Nelson and Narens' (1990) conceptual scheme? Koriat (2018) argued that because of people's adaptation to the world through evolution and learning, the responses to 2AFC items are more likely to be correct than wrong for many domains. Therefore, when items are sampled representatively from their reference classes, as recommended by proponents of the ecological approach (see Brunswik 1955; Dhami et al. 2004; Gigerenzer et al. 1991; Hoffrage and Hertwig 2006), M-O correspondence should be relatively high. However, it will be largely comparable to S-L correspondence. Only when the accuracy of first-order responses is not better than chance should M-O correspondence be low whereas S-L correspondence should remain high. This is the general picture that emerges from the results although, somewhat surprisingly, in some cases (see Study 1 and Study 2), S-L correspondence was still higher than M-O correspondence even across items for which OLA was not below chance.

Thus, although the results obtained so far place some constraints on the applicability of the Nelson-Narens conceptual framework, this framework is still very useful for most real-life situations. In fact, Koriat (2018) proposed that both the self-consistency heuristic assumed to underlie subjective confidence (Koriat 2012a), and the accessibility heuristic assumed to underlie the feeling-of-knowing (Koriat 1993; Koriat 1995), have been specifically tailored to the probabilistic structure of the environment, for which first-order judgments tend to be correct by and large. Other heuristics capitalize on regularities that exist in the "internal ecology" (e.g., that easily-learned items are better remembered, see Koriat 2008a), and these heuristics have been found to develop during childhood (Koriat et al. 2009a; Koriat et al. 2009b), suggesting that metacognitive heuristics are learned (Unkelbach 2006).

The foregoing discussion implies that the Nelson-Narens conceptual framework should be largely useful across many real-life conditions. It is only under certain special conditions (e.g., Brewer and Sampaio 2012; DeSoto and Roediger III 2014; Koriat 1995; Sampaio et al. 2017) that the results should deviate from what would be expected by this framework.

What happens in these conditions? First, metacognitive judgments may be deceitful to the extent of being counterdiagnostic of accuracy. This is what happens for the relatively infrequent CW items. However, even for these items, confidence judgments may still track the consensuality of the response yielding the peculiar pattern documented in this study: Rather than monitoring the property that they are purported to track, they predict the responses of others who face the same task. However, consistent with the monitoring-affects-control hypothesis (Nelson and Leonesio 1988), participants still rely on these judgments in regulating their cognitive processes and behavior (Fischhoff et al. 1977; Koriat and Goldsmith 1996; Pansky and Goldsmith 2014). In that case, monitoring-based regulation may actually be detrimental. For example, when participants were asked to wager money on their answer, they placed larger wagers on the correct answers for CC items, thus maximizing their earnings (Koriat 2011). For CW items, in contrast, they lost money by betting heavily on the wrong choices. Also, decisions made jointly by a group tend to be more accurate than the decisions made by the individual members. This is partly due to the fact that for each issue, the more confident members tend to have greater impact on the group decision. However, group discussion was found to improve decision accuracy only for CC items, whereas for CW items it proved detrimental to accuracy (Koriat 2015).

The foregoing discussion implies an interesting consortium involving three major dimensions: consensus, confidence and accuracy. In the natural ecology, a positive correlation exists between the three components such that both consensus and confidence are strongly correlated, and both track the accuracy of the response. In SCM, these correlations are explained in terms of the distributed wisdom of crowds (Koriat and Sorka 2017): The shared population of cues that people draw on in making their first-order judgments generally converges on the correct judgment. Hence accuracy and consensus go hand in hand. When we step aside from the common conditions characteristic of the natural ecology, this consortium breaks down. What characterizes a "misleading" or "deceptive" item (see Gigerenzer et al. 1991; Koriat 2017) is that the distributed wisdom of crowds from which people sample their cues converges on the wrong answer. In that case, the positive links between confidence and accuracy and between consensus and accuracy break down. However, confidence still tracks the consensuality of the judgment.

Acknowledgments The preparation of this article was supported by Grant 2013039 from the United States—Israel Binational Science Foundation to Asher Koriat and Norbert Schwarz. I am grateful to Miriam Gil for her help in the analyses of the results, and to Etti Levran (Merkine) for her help in copyediting.

Compliance with ethical standards

Conflict of interest The author declares that he has no conflict of interest.

References

Ackerman, R., & Goldsmith, M. (2011). Control over grain size in memory reporting—With and without satisficing knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 1224–1245. https://doi.org/10.1037/a0012938.

Ackerman, R., & Koriat, A. (2011). Response latency as a predictor of the accuracy of children's reports. *Journal of Experimental Psychology: Applied, 17*, 406–417. https://doi.org/10.1037/a0025129.

Alba, J. W., & Marmorstein, H. (1987). The effects of frequency knowledge on consumer decision making. *Journal of Consumer Research, 14*, 14–25. https://doi.org/10.1086/209089.

Ariel, R., Dunlosky, J., & Bailey, H. (2009). Agenda-based regulation of study-time allocation: When agendas override item-based monitoring. *Journal of Experimental Psychology: General, 138*, 432–447. https://doi.org/10.1037/a0015928.

Armelius, K. (1979). Task predictability and performance as determinants of confidence in multiple-cue judgments. *Scandinavian Journal of Psychology, 20*, 19–25. https://doi.org/10.1111/j.1467-9450.1979.tb00678.

Bajšanski, I., Žauhar, V., & Valerjev, P. (2019). Confidence judgments in syllogistic reasoning: The role of consistency and response cardinality. *Thinking & Reasoning, 25*(1), 14–47. https://doi.org/10.1080/13546783.2018.1464506.

Baranski, J. V., & Petrusic, W. M. (1998). Probing the locus of confidence judgments: Experiments on the time to determine confidence. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 929–945. https://doi.org/10.1037/0096-1523.24.3.929.

Barr, D. J., & Keysar, B. (2005). Mindreading in an exotic case: The normal adult human. In B. F. Malle & S. D. Hodges (Eds.), *Other minds: How humans bridge the divide between self and other* (pp. 271–283). New York: Guilford Press.

Benjamin, A. S. (2008). Memory is more than just remembering: Strategic control of encoding, accessing memory, and making decisions. In A. S. Benjamin & B. H. Ross (Eds.), *The psychology of learning and motivation: Skill and strategy in memory use* (Vol. 48, pp. 175–223). London: Academic Press. https://doi.org/10.1016/S0079-7421(07)48005-7.

Benjamin, A. S., & Bjork, R. A. (1996). Retrieval fluency as a metacognitive index. In L. M. Reder (Ed.), *Implicit memory and metacognition* (pp. 309–338). Mahwah, NJ: Erlbaum.

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*, 55–68. https://doi.org/10.1037/0096-3445.127.1.55.

Benjamin, A. S., & Diaz, M. (2008). Measurement of relative metamnemonic accuracy. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of memory and metamemory* (pp. 73–94). New York, NY: Psychology Press.

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.

Brewer, W. F., & Sampaio, C. (2006). Processes leading to confidence and accuracy in sentence recognition: A metamemory approach. *Memory, 14*, 540–552. https://doi.org/10.1080/09658210600590302.

Brewer, W. F., & Sampaio, C. (2012). The metamemory approach to confidence: A test using semantic memory. *Journal of Memory and Language, 67*, 59–77. https://doi.org/10.1016/j.jml.2012.04.002.

Brewer, W. F., Sampaio, C., & Barlow, M. R. (2005). Confidence and accuracy in the recall of deceptive and nondeceptive sentences. *Journal of Memory and Language, 52*, 618–627. https://doi.org/10.1016/j.jml.2005.01.017.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review, 62*, 193–217. https://doi.org/10.1037/h0047470.

Chandler, C. C. (1994). Studying related pictures can reduce accuracy, but increase confidence in a modified recognition test. *Memory & Cognition, 22*, 273–280. https://doi.org/10.3758/BF03200854.

DeSoto, K. A., & Roediger, H. L., III. (2014). Positive and negative correlations between confidence and accuracy for the same events in recognition of categorized lists. *Psychological Science, 25*, 781–788. https://doi.org/10.1177/0956797613516149.

Dhami, M. K., Hertwig, R., & Hoffrage, U. (2004). The role of representative design in an ecological approach to cognition. *Psychological Bulletin, 130*, 959–988. https://doi.org/10.1037/0033-2909.130.6.959.

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. 249–276). Mahwah, NJ: Erlbaum.

Dunlosky, J., & Metcalfe, J. (2009). *Metacognition*. Thousand Oaks, CA: Sage.

Dunning, D. (2007). Prediction: The inside view. In E. T. Higgins & A. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (2nd ed., pp. 69–90). New York: Guilford.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance, 3*, 552–564. https://doi.org/10.1037/0096-1523.3.4.552.

Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience, 8*, 443. https://doi.org/10.3389/fnhum.2014.00443.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98*, 506–528. https://doi.org/10.1037/0033-295X.98.4.506.

Gill, M. J., Swann, W. B., Jr., & Silvera, D. H. (1998). On the genesis of confidence. *Journal of Personality and Social Psychology, 75*, 1101–1114. https://doi.org/10.1037/0022-3514.75.5.1101.

Goldsmith, M., & Koriat, A. (2008). The strategic regulation of memory accuracy and informativeness. In A. Benjamin & B. Ross (Eds.), *Psychology of learning and motivation* (*Memory use as skilled cognition*) (Vol. 48, pp. 1–60). San Diego, CA: Elsevier. https://doi.org/10.1016/s0079-7421(07)48001-x.

Higham, P. A., Perfect, T. J., & Bruno, D. (2009). Investigating strength and frequency effects in recognition memory using type-2 signal detection theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 57–80. https://doi.org/10.1037/a0013865.

Hoffrage, U., & Hertwig, R. (2006). Which world should be represented in representative design? In K. Fiedler & P. Juslin (Eds.), *Information sampling and adaptive cognition* (pp. 381–408). Cambridge, UK: Cambridge University Press.

Juslin, P. (1994). The overconfidence phenomenon as a consequence of informal experimenter-guided selection of almanac items. *Organizational Behavior and Human Decision Processes, 57*, 226–246. https://doi.org/10.1006/obhd.1994.1013.

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

Keynes, J. M. (1936). *The general theory of employment, interest and money.* New York, NY: Harcourt, Brace and Company.

Koriat, A. (1976). Another look at the relationship between phonetic symbolism and the feeling of knowing. *Memory & Cognition, 4*, 244–248. https://doi.org/10.3758/BF03213170.

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review, 100*, 609–639. https://doi.org/10.1037/0033-295X.100.4.609.

Koriat, A. (1995). Dissociating knowing and the feeling of knowing: Further evidence for the accessibility model. *Journal of Experimental Psychology: General, 124*, 311–333. https://doi.org/10.1037/0096-3445.124.3.311.

Koriat, A. (2008a). Easy comes, easy goes? The link between learning and remembering and its exploitation in metacognition. *Memory & Cognition, 36*, 416–428. https://doi.org/10.3758/MC.36.2.416.

Koriat, A. (2008b). Subjective confidence in one's answers: The consensuality principle. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 945–959. https://doi.org/10.1037/0278-7393.34.4.945.

Koriat, A. (2011). Subjective confidence in perceptual judgments: A test of the self-consistency model. *Journal of Experimental Psychology: General, 140*, 117–139. https://doi.org/10.1037/a0022171.

Koriat, A. (2012a). The self-consistency model of subjective confidence. *Psychological Review, 119*, 80–113. https://doi.org/10.1037/a0025648.

Koriat, A. (2012b). When are two heads better than one and why? *Science, 336*, 360–362. https://doi.org/10.1126/science.1216549.

Koriat, A. (2013). Confidence in personal preferences. *Journal of Behavioral Decision Making, 26*, 247–259. https://doi.org/10.1002/bdm.1758.

Koriat, A. (2015). When two heads are better than one and when they can be worse: The amplification hypothesis. *Journal of Experimental Psychology: General, 144*, 934–950. https://doi.org/10.1037/xge0000092.

Koriat, A. (2016). Metacognition: Decision-making processes in self-monitoring and self-regulation. In G. Keren & G. Wu (Eds.), (Vol. 1, pp. 356–379). Malden, MA: Wiley–Blackwell.

Koriat, A. (2017). Can people identify "deceptive" or "misleading" items that tend to produce mostly wrong answers? *Journal of Behavioral Decision Making, 30*, 1066–1077. https://doi.org/10.1002/bdm.2024.

Koriat, A. (2018). When reality is out of focus: Can people tell whether their beliefs and judgments are correct or wrong? *Journal of Experimental Psychology: General, 147*, 613–631. https://doi.org/10.1037/xge0000397.

Koriat, A., Ackerman, R., Lockl, K., & Schneider, W. (2009a). The easily learned, easily-remembered heuristic in children. *Cognitive Development, 24*, 169–182. https://doi.org/10.1016/j.cogdev.2009.01.001.

Koriat, A., Ackerman, R., Lockl, K., & Schneider, W. (2009b). The memorizing-effort heuristic in judgments of learning: A developmental perspective. *Journal of Experimental Child Psychology, 102*, 265–279. https://doi.org/10.1037/a0016374.

Koriat, A., & Adiv, S. (2016). The self-consistency theory of subjective confidence. In J. Dunlosky & S. Tauber (Eds.), *The Oxford handbook of metamemory* (pp. 127–147). New York: Oxford.

Koriat, A., Adiv, S., & Schwarz, N. (2016). Views that are shared with others are expressed with greater confidence and greater fluency independent of any social influence. *Personality and Social Psychology Review, 20*, 176–193. https://doi.org/10.1177/1088868315585269.

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

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*, 187–194. https://doi.org/10.1037/0278-7393.31.2.187.

Koriat, A., & Bjork, R. A. (2006). Illusions of competence during study can be remedied by manipulations that enhance learners' sensitivity to retrieval conditions at test. *Memory & Cognition, 34*, 959–972. https://doi.org/10.3758/BF03193244.

Koriat, A., Fiedler, K., & Bjork, R. A. (2006). Inflation of conditional predictions. *Journal of Experimental Psychology: General, 135*, 429–447. https://doi.org/10.1037/0096-3445.135.3.429.

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103*, 490–517. https://doi.org/10.1037/0033-295X.103.3.490.

Koriat, A., Pansky, A., & Goldsmith, M. (2011). An output-bound perspective on false memories: The case of the Deese-Roediger-McDermott (DRM) paradigm. In A. Benjamin (Ed.), *Successful remembering and successful forgetting: A Festschrift in honor of Robert a. Bjork* (pp. 297–328). London, UK: Psychology Press.

Koriat, A., & Sorka, H. (2017). The construction of category membership judgments: Towards a distributed model. In H. Cohen & C. Lefebvre (Eds.), *Handbook of categorization in cognitive science* (2nd ed., pp. 773–794). Amsterdam, Netherlands: Elsevier.

Kurdi, B., Diaz, A. J., Wilmuth, C. A., Friedman, M. C., & Banaji, M. R. (2018). Variations in the relationship between memory confidence and memory accuracy: The effects of spontaneous accessibility, list length, modality, and complexity. *Psychology of Consciousness: Theory, Research, and Practice, 5*, 3–28. https://doi.org/10.1037/cns0000117.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306–334). New York, NY: Cambridge University Press.

Maniscalco, B., & Lau, H. (2014). Signal detection theory analysis of type 1 and type 2 data: Meta-d', response-specific meta-d', and the unequal variance SDT model. In S. M. Fleming & C. D. Frith (Eds.), *The cognitive neuroscience of metacognition* (pp. 25–66). Berlin, Germany: Springer. https://doi.org/10.1007/978-3-642-45190-4_3.

Metcalfe, J. (2009). Metacognitive judgments and control of study. *Current Directions in Psychological Science, 18*, 159–163. https://doi.org/10.1111/j.1467-8721.2009.01628.x.

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

Metcalfe, J., & Kornell, N. (2003). The dynamics of learning and allocation of study time to a region of proximal learning. *Journal of Experimental Psychology: General, 132*, 530–542. https://doi.org/10.1037/0096-3445.132.4.530.

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

Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the "labor-in-vain effect.". *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 76–686. https://doi.org/10.1037/0278-7393.14.4.676.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 1–45). New York, NY: Academic Press.

Nelson, T. O., & Narens, L. (1994). Why investigate metacognition? In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 1–25). Cambridge, MA: MIT Press.

Pansky, A., & Goldsmith, M. (2014). Metacognitive effects of initial question difficulty on subsequent memory performance. *Psychonomic Bulletin & Review, 21*, 1255–1262. https://doi.org/10.3758/s13423-014-0597-2.

Pulford, B. D., Colman, A. M., Buabang, E. K., & Krockow, E. M. (2018). The persuasive power of knowledge: Testing the confidence heuristic. *Journal of Experimental Psychology: General, 147*, 1431–1444. https://doi.org/10.1037/xge0000471.

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). New York: Oxford UP.

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

Robinson, M. D., Johnson, J. T., & Herndon, F. (1997). Reaction time and assessments of cognitive effort as predictors of eyewitness memory accuracy and confidence. *Journal of Applied Psychology, 82*, 416–425. https://doi.org/10.1037/0021-9010.82.3.416.

Roediger, H. L., & DeSoto, K. A. (2015). Understanding the relation between confidence and accuracy in reports from memory. In S. D. Lindsay, C. M. Kelley, A. P. Yonelinas, & H. L. Roediger III (Eds.), *Remembering: Attributions, processes, and control in human memory: Papers in honor of Larry L. Jacoby* (pp. 347–367). New York, NY: Psychology Press.

Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 803–814. https://doi.org/10.1037/0278-7393.21.4.803.

Rubinsten, O., Anaki, D., Henik, A., Drori, S., & Faran, Y. (2005). Norms for free associations in the Hebrew language. In A. Henik, O. Rubinsten, & D. Anaki (Eds.), *Word norms for the Hebrew language (in Hebrew) (pp. 17–34)*. Ben Gurion University of the Negev.

Sampaio, C., & Brewer, W. F. (2009). The role of unconscious memory errors in judgments of confidence for sentence recognition. *Memory & Cognition, 37*, 158–163. https://doi.org/10.3758/MC.37.2.158.

Sampaio, C., Reinke, V., Mathews, J., Swart, A., & Wallinger, S. (2017). High confidence in falsely recognizing prototypical faces. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 71*, 1348–1356. https://doi.org/10.1080/17470218.2017.1329844.

Slovic, P. (1966). Cue-consistency and cue-utilization in judgment. *The American Journal of Psychology, 79*, 427–434. https://doi.org/10.2307/1420883.

Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 204–221. https://doi.org/10.1037/0278-7393.26.1.204.

Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology, 95*, 66–73. https://doi.org/10.1037/0022-0663.95.1.66.

Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1024–1037. https://doi.org/10.1037/0278-7393.25.4.1024.

Tullis, J. G. (2018). Predicting others' knowledge: Knowledge estimation as cue utilization. *Memory & Cognition*, 1–16. https://doi.org/10.3758/s1342.

Tullis, J. G., & Benjamin, A. S. (2011). On the effectiveness of self-paced learning. *Journal of Memory and Language, 64*, 109–118. https://doi.org/10.1016/j.jml.2010.11.002.

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

Van Swol, L. M., & Sniezek, J. A. (2005). Factors affecting the acceptance of expert advice. *British Journal of Social Psychology, 44*, 443–461. https://doi.org/10.1348/014466604X17092.

Weidemann, C. T., & Kahana, M. J. (2016). Assessing recognition memory using confidence ratings and response times. *Royal Society Open Science, 3*, 150670. https://doi.org/10.1098/rsos.150670.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.