Examining the robustness of the relationship between metacognitive efficiency and metacognitive bias

Kai Xue [*], Medha Shekhar, Dobromir Rahnev

School of Psychology, Georgia Institute of Technology, Atlanta, GA, United States

ARTICLE INFO

ABSTRACT

We recently found a positive relationship between estimates of metacognitive efficiency and metacognitive bias. However, this relationship was only examined on a within-subject level and required binarizing the confidence scale, a technique that introduces methodological difficulties. Here we examined the robustness of the positive relationship between estimates of metacognitive efficiency and metacognitive bias by conducting two different types of analyses. First, we developed a new within-subject analysis technique where the original n-point confidence scale is transformed into two different (n-1)-point scales in a way that mimics a naturalistic change in confidence. Second, we examined the across-subject correlation between metacognitive efficiency and metacognitive bias. Importantly, for both types of analyses, we not only established the direction of the effect but also computed effect sizes. We applied both techniques to the data from three tasks from the Confidence Database (N > 400 in each). We found that both approaches revealed a small to medium positive relationship between metacognitive efficiency and metacognitive bias. These results demonstrate that the positive relationship between metacognitive efficiency and metacognitive bias is robust across several analysis techniques and datasets, and have important implications for future research.

1. Introduction

Metacognition refers to the ability to evaluate the accuracy of one's decisions. Accurate metacognition is important in many domains such as successful learning (Aguilar-Lleyda et al., 2020; Hainguerlot et al., 2018; Schunk & Ertmer, 2000) and the ability to engage in sequential decisions (van den Berg et al., 2016). However, despite its importance, metacognition is often found to be inefficient (Fleming et al., 2010; Maniscalco & Lau, 2012; Metcalfe & Shimamura, 1994; Shekhar & Rahnev, 2021a). Yet, the properties of metacognitive efficiency – the degree to which confidence judgments predict accuracy independently from task performance – are not yet well understood.

We recently revealed a previously unknown property of metacognitive efficiency (computed as the ratio meta-d'/d'; Maniscalco & Lau, 2012), namely that metacognitive efficiency increases for higher levels of confidence (Shekhar & Rahnev, 2021b). To investigate this relationship, our previous work developed an analysis technique where the original confidence scale was transformed into multiple 2-point confidence scales by moving a single cutoff. For example, a 4-point confidence scale would be transformed into 2-point confidence scales in three different ways such that the low confidence on the resulting 2-point scales consist of the original ratings 1, 1–2, or 1–3, respectively. This technique revealed that lower cutoffs (which result in higher average confidence) led to higher

---

estimates of metacognition efficiency.

However, our previous analysis technique has at least three limitations. First, it introduces methodological difficulties related to either empty cells or cells with small numbers of trials that lead to systematic misestimation of metacognitive efficiency. Critically, the resulting misestimation can produce the same qualitative results (higher estimated metacognitive efficiency for higher levels of confidence) even in the absence of a true effect and addressing this issue can require excluding substantial amount of data (Shekhar & Rahnev, 2021b). Moreover, these issues do not typically arise in empirical data since few studies collect confidence on 2-point scales, and the ones that do typically ensure that subjects use the ratings in a balanced way that generates no empty cells or cells with very small number of trials. Second, our previous technique does not allow for the computation of effect sizes related to the expected increase in metacognitive efficiency with manipulations of metacognitive bias (i.e., confidence level) in real experiments. For example, if a subject is to increase his or her confidence from one condition to the next in a real experiment, we may want to know how large and robust the corresponding increase in estimated metacognitive efficiency would be. However, our previous technique can only address this question for experiments with 2-point scales. Finally, our previous technique only addresses within-subject effects, leaving open the question of whether the positive relationship between metacognitive efficiency and metacognitive bias also holds across subjects.

Here we address these limitations in two different ways. First, we develop a new within-subject analysis technique of simulating a change in metacognitive bias, which addresses the first two limitations above. The technique reduces an n-point confidence scale to two different (n-1)-point confidence scales (Fig. 1a): either all ratings from 2 to n are reduced by one (resulting in relatively lower confidence), or only the rating n is reduced by one (resulting in relatively higher confidence). We illustrate this technique in Fig. 1b. By avoiding the creation of 2-point confidence scales, this technique substantially ameliorates the influence of empty cells or cells with small number of trials. The new technique mimics a situation where an experiment with (n-1)-point scale has two conditions – one with lower and one with higher confidence – for the same subject. Therefore, comparing the two different recodings of the original scale allows us to computed effect sizes related to the expected increase in metacognitive efficiency with manipulations of metacognitive bias. As such, our new technique addresses the first two limitations of the recoding technique we used previously (Shekhar & Rahnev, 2021b).

Second, to address the third limitation above, we analyze across-subject correlations between metacognitive efficiency and metacognitive bias. These correlations can be used to understand the extent to which the within-subject association between metacognitive efficiency and bias affect analyses conducted across subjects. As with the technique above, the across-subject correlations also naturally allow for the estimation of effect sizes.

We apply both techniques to data from three tasks (each N > 400) from two different papers (Haddara & Rahnev, 2021; Rouault, Seow, Gillan, & Fleming, 2018) made available as part of the Confidence Database (Rahnev et al., 2020). To anticipate, both analysis
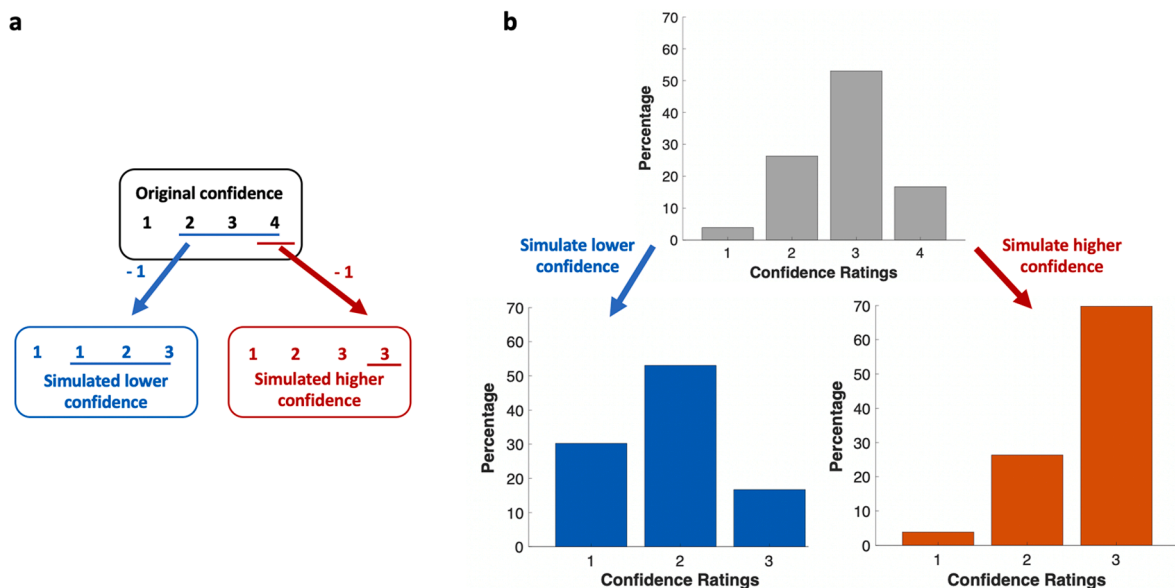
Fig. 1. Within-subject recoding technique. (a) An illustration of the recoding technique for a 4-point confidence scale that is recoded in two different ways into 3-point scales. The original confidence ratings are transformed by either recoding down confidence ratings from 2 to n or just rating n. On the left, we depict the process of simulating a low-confidence condition by reducing ratings 2 to 4 by 1. On the right, we depict the process of simulating a high-confidence condition by reducing only rating 4 by 1. (b) Example distribution of each confidence ratings after recoding. The panel shows the effects of the recoding technique when applied to the data from an example subject from Task 1. The original confidence rating distributions of this subject are shown in the figure above in gray (confidence ratings 1 to 4 were selected 3.9%, 26.3%, 53.1% and 16.7% of the time). In blue is the distribution of simulated low-confidence condition (confidence ratings 1 to 3 are now selected 30.2%, 53.1% and 16.7% of the time) and in red is the distribution of simulated higher confidence ratings (confidence ratings 1 to 3 are now selected 3.9%, 26.3% and 69.8% of the time).

techniques confirmed the positive relationship between metacognitive efficiency and metacognitive bias. The results were generally robust across all three tasks, and revealed effect sizes that ranged from small to medium. These findings extend our previous results in confirming the positive association between metacognitive bias and metacognitive efficiency, and clarify the robustness and strength of the relationship.

2. Methods

2.1. Dataset selection

To test the robustness of the relationship between metacognitive efficiency and metacognitive bias, we selected the data from the same three tasks that we analyzed in a recent publication (Rahnev, 2021). All three tasks were made available in the Confidence Database (Rahnev et al., 2020). Tasks 1 and 2 are from a dataset named "Haddara_2020" (Haddara & Rahnev, 2021), which includes 443 subjects. Task 3 is from a dataset named "Rouault_2018_Expt1" (Rouault et al., 2018), which includes 498 subjects.

2.2. Experimental designs

Details about Tasks 1 and 2 can be found in the original paper (Haddara & Rahnev, 2021). Briefly, subjects indicated whether the letters X or O (Task 1) or the colors red or blue (Task 2) were presented more frequently in a 7×7 grid. In Task 1, the letter that occurred more frequently was presented in 30 of the 49 locations. In Task 2, the color that occurred more frequently was presented in 27 of the 49 locations. Both tasks began with a fixation period (500 ms) and were followed by stimulus presentation (500 ms), untimed perceptual judgment, and untimed confidence rating provided on a 4-point scale (Fig. 2). The two tasks were adapted from (Rahnev et al., 2015). In Task 1, about half of the subjects received feedback immediately after each trial, and the other half did not receive any feedback. For the group with feedback, the word 'Correct' or 'Wrong' was presented on the feedback screen for 500 ms after each trial. For the group with no feedback, subjects saw a fixation cross for 500 ms. Here all subjects were pooled together regardless of whether they were provided with feedback or not. In Task 2, no subject received feedback. Task 1 had 330 trials per subject, whereas Task 2 had 150 trials per subject.

Details about Task 3 are available in the original paper (Rouault et al., 2018). Subjects identified which of two simultaneously presented black boxes had a higher number of white dots. One box was always half-filled (313 dots out of 625 positions), and the other box contained an increment of +1 to +70 dots compared to the first one (Fig. 2). All trials began with a fixation period for 1,000 ms, followed by 300-ms stimulus presentation, untimed perceptual judgment period, and untimed confidence rating provided on an 11-point scale. The confidence scale was organized such that the ratings lower than 6 indicated that the perceptual decision was made in error. More specifically, the different confidence ratings were labeled 1 = certainly wrong, 3 = probably wrong, 5 = maybe wrong, 7 = maybe correct, 9 = probably correct, 11 = certainly correct. No feedback was given, and each subject completed 210 trials.

All data were collected online using Amazon Mechanical Turk. The experiments were performed using jsPsych (version 5.0.3 for
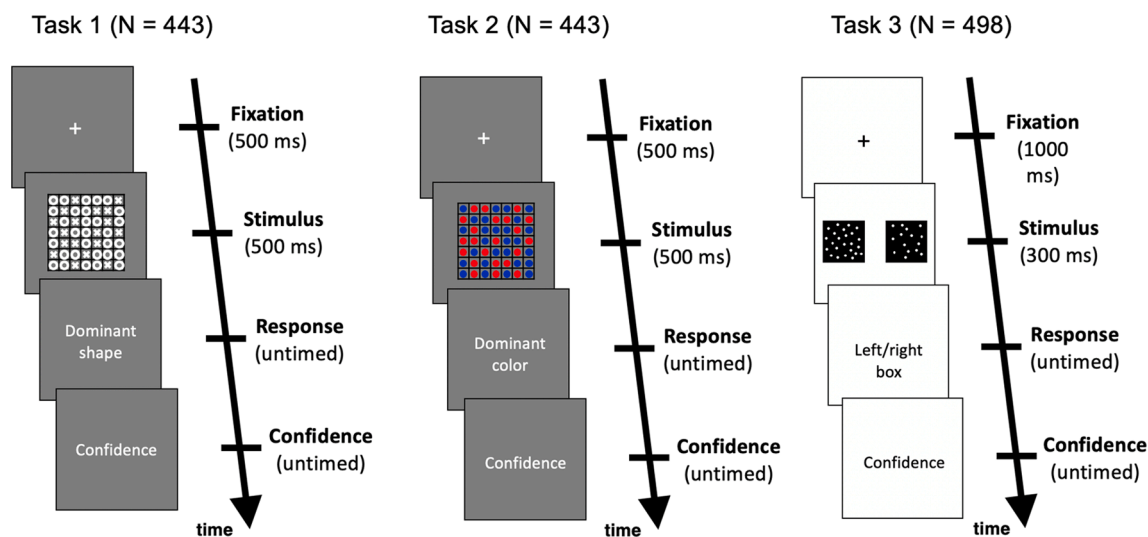
Fig. 2. Experimental designs for the three tasks. Subjects indicated whether the letters X or O occurred more frequently (Task 1), whether the colors red or blue occurred more frequently (Task 2), or whether the left or right box contained more white dots (Task 3). In all cases, each trial began with a fixation cross presented for either 500 ms (Tasks 1 and 2) or 1000 ms (Task 3). The stimulus was then presented for a short period (500 ms in Tasks 1 and 2; 300 ms in Task 3) and was followed by untimed response and confidence periods. Confidence was reported on a 4-point scale in Tasks 1 and 2, and on an 11-point scale in Task 3. Tasks 1–2 were originally reported in Haddara & Rahnev (2020); Task 3 was originally reported in Rouault et al. (2018).

Tasks 1 and 2, version 4.3 for Task 3).

2.3. Data preprocessing

We previously analyzed the same three tasks to investigate how bias depends on the individual variability in sensory recoding (Rahnev, 2020). Here we used the same criteria for preprocessing the data as in this previous paper. Specifically, for all tasks, we first removed all trials with reaction times lower than 200 ms or higher than 2 s. We then excluded subjects who had resulting accuracy lower than 55% correct or higher than 95% right. We further excluded subjects who only used less than three distinct confidence ratings. This was necessary because having only two distinct confidence ratings makes it impossible to apply our within-subject recoding technique that simulates a change in confidence.

Using these exclusion criteria, we removed 69 subjects in Task 1 (15.3%), 79 subjects in Task 2 (17.8%), and 5 subjects in Task 3 (1%). Note that the much larger rate of exclusion for Tasks 1 and 2 is due to the fact that the dataset Haddara_2020 included all subjects regardless of data quality, while the low exclusion rate for Task 3 is because the dataset Rouault_2018_Expt1 only included subjects that passed exclusion criteria that partially overlapped with the exclusion criteria here. Specifically, the original exclusion criteria used by the dataset Rouault_2018_Expt1 were: (1) removing subjects with lower than 55% correct task performance, (2) removing subjects who always selected the same confidence rating, and (3) removing trials with reaction times higher than 10 s or$>$3 standard deviations from the per-subject mean reaction time (Rouault et al., 2018).

The ratings 1–6 in the original 11-point scale in Task 3 were specifically designated for error detection and were therefore rarely used. Therefore, we transformed the original 11-point scale into a 6-point scale such that the ratings between 1 and 6 were coded as "1″, while the ratings of 7 to 11 were coded as "2" to "6" by subtracting 5 from each. In addition, in control analyses, we excluded trials with confidence between 1 and 6 altogether. For these control analyses, we transformed the confidence ratings of 7 to 11 to a scale of 1 to 5 by subtracting 6 from each. The control analyses produced similar results to our main analyses (see Supplementary Results).

Since the reaction times and confidence may be closely intertwined (Kiani et al., 2014), we also performed control analyses with no filtering based on reaction times in order to ensure that the observed relationships were not specific to the set of exclusion criteria we used (see Supplementary Results).

2.4. Analyses

To overcome the limitations of our previous binarization technique (Shekhar & Rahnev, 2021b), we developed a new within-subject analysis technique of simulating a change in metacognitive bias. The technique recodes an n-point confidence scale to a (n-1)-point confidence scale in two different ways (Fig. 1). The first recoding decreases all ratings from 2 to n by one; the second recoding decreases only the rating n by one. For example, a 4-point confidence scale would be recoded such that the original ratings 1, 2, 3, and 4 become 1, 1, 2, and 3 (first recoding) or 1, 2, 3, and 3 (second recoding). The first recoding thus leads to lower average confidence than the second recoding. By comparing the metacognitive efficiency for the two types of recoding, we can then determine how the estimated metacognitive efficiency depends on metacognitive bias and estimate the effect size for this comparison.

In addition to the within-subject analyses, we also performed across-subject correlations between metacognitive efficiency and metacognitive bias. For these analyses, we computed metacognitive efficiency for each subject using the original data (without any recoding). We also computed the metacognitive bias of each subject as their average confidence. Finally, we performed across-subject Pearson correlations for these two measures.

Metacognitive efficiency was estimated by computing the measure meta-d'/d', also known as Mratio, developed by Maniscalco & Lau (2012). In control analyses, we further tested metacognitive sensitivity, meta-d'. For both measures, we tested whether a difference in estimated metacognitive ability exists between the two types of confidence recoding using paired-sample t-tests.

We confirm that we have reported all measures, conditions, and data exclusion used by each dataset. The sample size of each dataset was determined by the original authors (Haddara & Rahnev, 2021; Rouault et al., 2018).

2.5. Data and code availability

All data and codes for the analyses have been made freely available at https://osf.io/k8ghz/.

3. Results

We recently demonstrated the existence of a positive association between estimates of metacognitive efficiency and metacognitive bias using a within-subject analysis technique that relies on binarizing the original confidence scale (Shekhar & Rahnev, 2021b). Here we tested the robustness of the positive relationship between metacognitive efficiency and bias using two different techniques. First, we developed a different analysis technique to address the limitations of our original technique (Fig. 1). Second, we examined the across-subject correlation between metacognitive efficiency and metacognitive bias. We applied both techniques to data from three tasks (Fig. 2) reported in recent papers (Haddara & Rahnev, 2021; Rouault et al., 2018).

3.1. Within-subject analyses using new recoding technique

We developed a new within-subject recoding technique to simulate changes in confidence in a more naturalistic fashion. We first

confirmed that that the new technique successfully simulated a change in confidence (Fig. 3a). Analyzing the data from Task 1, we found that the recoding that simulates higher confidence (achieved by reducing only confidence rating n by one) indeed led to higher average confidence (mean confidence = 2.55) compared to the recoding that simulates lower confidence (achieved by reducing all confidence ratings from 2 to n by one; mean confidence = 2.05), with the difference being highly significant (t(393) = 41.3, $p = 3.8 \times 10^{-145}$, Cohen's d = 1.1). The same effects were also observed for Task 2 (mean confidence for simulated higher confidence: 2.42, mean confidence for simulated lower confidence: 1.95, t(388) = 34.9, $p = 5.3 \times 10^{-122}$, Cohen's d = 1.0) and Task 3 (mean confidence for simulated higher confidence: 2.86, mean confidence for simulated lower confidence: 2.17, t(494) = 78.5, $p = 3.8 \times 10^{-281}$, Cohen's d = 1.0). These results show that our recoding technique is effective in manipulating metacognitive bias by robustly increasing or decreasing confidence.

Having established the ability of our recoding technique to simulate a change in confidence, we then computed metacognitive efficiency in each of the two recoding conditions (Fig. 3b). We found that simulating higher confidence resulted in significantly higher estimated metacognitive efficiency than simulating lower confidence, and that this effect was significant in all three tasks (Task 1: t(386) = 6.0, $p = 4.0 \times 10^{-9}$, Cohen's d = 0.31; Task 2: t(352) = 5.6, $p = 5.4 \times 10^{-8}$, Cohen's d = 0.30; Task 3: t(486) = 3.2, $p = 0.002$, Cohen's d = 0.14). These results confirm the positive relationship between metacognitive efficiency and metacognitive bias. The effect sizes obtained (from 0.14 to 0.30) are considered to be small to medium, thus showing that the effect is of relatively modest magnitude.

Mratio values tend to be noisier due to the division of meta-d' by d', which can lead to extreme Mratio values for small values of d'. Because our recoding technique does not affect the decisions, both recoding conditions have the same d', and therefore a more robust analysis would involve comparing meta-d' rather than Mratio. We thus repeated the above analyses using meta-d'. As with Mratio, we found significantly higher meta-d' in the recoding condition that simulates higher confidence (Task 1: t(386) = 8.2, $p = 3.5 \times 10^{-15}$, Cohen's d = 0.42; Task 2: t(352) = 6.4, $p = 6.1 \times 10^{-10}$, Cohen's d = 0.34; Task 3: t(486) = 4.8, $p = 2.1 \times 10^{-6}$, Cohen's d = 0.22). As expected, the meta-d' were less noisy than the Mratio values, which resulted in larger effect sizes (0.22 to 0.42) though they remained in the small to medium range.

3.2. Across-subject correlations between Mratio and confidence

Our analysis technique above extended our previous findings that a within-subject increase in confidence results in a corresponding increase in Mratio (Shekhar & Rahnev, 2021b) by showing that these results are robust to a more naturalistic confidence increase and by estimating the associated effect size. Here we further extended our previous results by testing the robustness of the relationship between metacognitive bias and metacognitive efficiency across subjects.
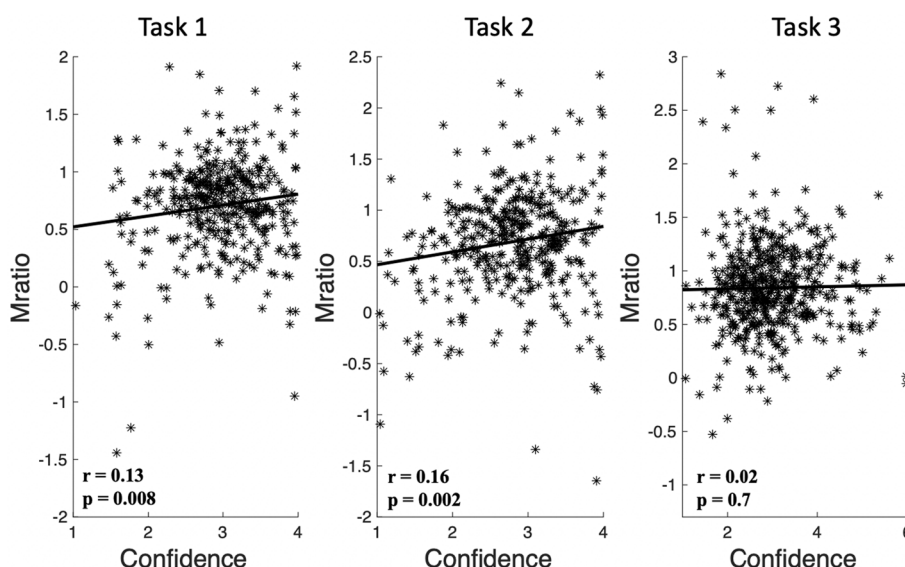
To do so, we examined the across-subject correlation of metacognitive efficiency and metacognitive bias. We found a significantly positive correlation between metacognitive efficiency and average confidence for both Task 1 (r = 0.13, $p = 0.008$) and Task 2 (r = 0.16, $p = 0.002$) but not for Task 3 (r = 0.02, $p = 0.7$; Fig. 4). These data do not reveal why the results for Tasks 1 and 2 differ from the results for Task 3 but one possible explanation is that only Task 3 features substantial variability in the difficulty of individual trials. Such variability in difficulty leads to inflated estimates of metacognitive ability across all measures of metacognition including Mratio (Rahnev & Fleming, 2019). This effect could have resulted in not just larger but also more variable Mratio values, thus reducing the power for finding a positive association with metacognitive bias.

To interpret the effect sizes obtained for Tasks 1 and 2 (r of 0.13 to 0.17), it is helpful to compare them to the effect sizes for other across-subject correlations that involve a more intuitive relationship (Funder & Ozer, 2019). For example, the correlation between confidence and Mratio in our data is of the same approximate magnitude as the correlation between perceptual sensitivity (d') and
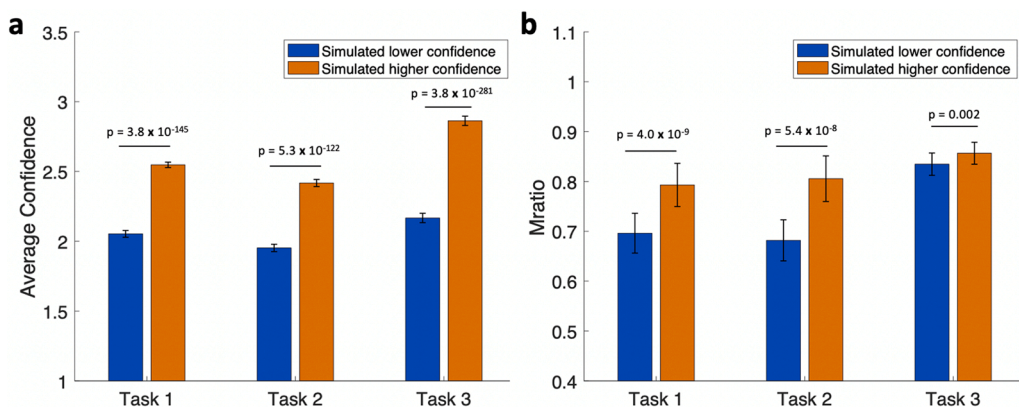
Fig. 3. Results obtained with our new within-subject recoding technique. (a) Average confidence for the two recoding conditions. The recoding simulating higher confidence (obtained by recoding down only the highest rating, n, in an n-point scale) led to higher confidence than the recoding simulating lower confidence (obtained by recoding down the ratings 2 to n). This relationship was significant for all three tasks, confirming that our technique of simulating higher or lower confidence indeed leads to a difference in overall confidence. (b) Estimated metacognitive efficiency for the two recoding conditions. The recoding simulating higher confidence led to higher estimated Mratio than the recoding simulating lower confidence. This relationship was significant for all three tasks, confirming the positive relationship between metacognitive efficiency and confidence.

Fig. 4. Across-subject correlation between metacognitive efficiency and confidence. There was a significantly positive across-subject correlation between metacognitive efficiency and confidence for Tasks 1 and 2, but not for Task 3. Each dot on the scatter plot represents a single subject.

confidence (Task 1: r = 0.15, p = 0.002; Task 2: r = -0.06, p = 0.25; Task 3: r = 0.2, p = 7.1 × 10⁻⁵; Supplementary Fig. 1). In other words, across subjects, the strength of the association between higher confidence and higher Mratio is about as high as the strength of the association between higher sensitivity and higher confidence.

4. Discussion

We examined the robustness of the strength and robustness of the positive relationship between metacognitive efficiency and metacognitive bias (Shekhar & Rahnev, 2021b). First, we developed a new technique to simulate a naturalistic, within-subject increase or decrease in confidence in existing data. Second, we examined the across-subject correlations between metacognitive efficiency and metacognitive bias. Both approaches confirmed the existence of a positive relationship of metacognitive efficiency and metacognitive bias in three large datasets with the size of the effect being small to medium. These results show that the positive relationship is robust across several analysis techniques and datasets.

A central goal of the current paper was to estimate the strength of the association between metacognitive efficiency and meta-cognitive bias. While the across-subject correlation of ~ 0.15 is typically considered small, there have been recent attempts for a more nuanced understanding of association strength (Funder & Ozer, 2019). Based on both simulations and comparisons to intuitively understood effects (e.g., effect of nonsteroidal anti-inflammatory drugs, such as ibuprofen, on pain is r = 0.14), Funder and Ozer proposed the following classification: "an effect-size r of 0.10 indicates an effect that is still small at the level of single events but potentially […] consequential, an effect-size r of 0.20 indicates a medium effect that is of some explanatory and practical use even in the short run." In addition, we found that, across subjects, the strength of the relationship between metacognitive efficiency and metacognitive bias is comparable to the strength of the relationship between confidence and perceptual sensitivity (d'). Therefore, we can conclude that while the relationship between metacognitive efficiency and metacognitive bias is small to medium, its size is likely to have practical implications and is comparable to other relationships that are of practical importance (e.g., we tend to rely more on confident eye-witnesses because we believe that they are likely more accurate).

What explains the relationship between metacognitive efficiency and metacognitive bias? We previously proposed that this relationship could be explained by postulating the existence of metacognitive noise (Bang et al., 2019; Maniscalco & Lau, 2016; Shekhar & Rahnev, 2021b; van den Berg et al., 2017) sampled from a lognormal distribution, which leads to a model that we named the "lognormal meta noise model" (Shekhar & Rahnev, 2021b). A lognormal distribution has the property that its variance increases with its mean, which can be thought of as signal-dependent multiplicative noise (Dosher & Lu, 1999; Lu et al., 2002; Lu & Dosher, 2008). Therefore, lognormal metacognitive noise implies that the confidence criteria far from the decision criterion are noisier than the confidence criteria located near the decision criterion. Practically, this effect implies that using the inside confidence criteria (which is equivalent to having high confidence) would lead to less noisy confidence ratings and therefore higher estimated metacognitive efficiency. However, while the lognormal meta noise model can certainly account of the positive relationship between metacognitive efficiency and metacognitive bias, it is possible that other models (Barrett et al., 2013; Rausch et al., 2020; Rausch & Zehetleitner, 2017) can account for this relationship too. Our results here are thus limited to establishing the robustness of the empirical relationship between metacognitive efficiency and metacognitive bias, but cannot pinpoint the source of this relationship.

What are the implications of our findings for Mratio as a measure of metacognitive ability? It is now well appreciated the meta-cognitive performance naturally depends on first-order performance (Fleming & Lau, 2014) and Mratio has been found to successfully

measure metacognitive ability independent of first-order performance in both simulations (Barrett et al., 2013) and empirical data (Shekhar & Rahnev, 2021b). If our lognormal meta noise model is correct, then there is also a natural relationship between metacognitive performance and metacognitive bias. The findings here thus simply reflect this relationship and do not imply that Mratio is a "bad" measure any more than any measure that does not correct for first-order performance is a "bad" measure. Yet, it is clearly desirable to have a measure that controls for both the effects of first-order performance and metacognitive bias, and our results show that Mratio does not accomplish this task. Future work should thus attempt to either implement a modified version of Mratio that can control for the effect of metacognitive bias or create a new measure that does.

An important implication of our findings is that any manipulation that changes one's overall confidence level should be expected to produce a spurious change in estimated metacognitive efficiency. In other words, if the same person is to increase their confidence due to any feature of the instructions or the task, then there may be a spurious increase in that person's estimated metacognitive ability. For example, a recent paper gave feedback that essentially penalized the use of low confidence ratings, and observed an increase in both overall confidence and metacognitive efficiency (Carpenter et al., 2019). Given our current findings, it is possible that at least a portion of the metacognitive increase in that study was due to the increase in confidence, and indeed a follow-up paper that replicated Carpenter et al.'s design but used feedback that did not penalize low confidence did not find a change in metacognitive efficiency (De Gardelle et al., 2020). A number of other studies have examined how metacognition changes over the course of training (Bang, Shekhar, & Rahnev, 2019; Schwiedrzik, Singer, & Melloni, 2011; Zizlsperger, Kümmel, Haarmeier, & Cropper, 2016; Guggenmos et al., 2016), or how it correlates with factors such as brain volume (Allen et al., 2017; Fleming, Weil, Nagy, Dolan, & Rees, 2010; McCurdy et al., 2013; Rahnev, Koizumi, McCurdy, D'Esposito, & Lau, 2015) or age (Palmer et al., 2014), but typically did not try to control for potential effects of overall confidence (that is, metacognitive bias). To avoid confounding metacognitive bias and metacognitive efficiency, future studies should minimize or control for changes in one variable when examining the other.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

References

Aguilar-Lleyda, D., Lemarchand, M., & de Gardelle, V. (2020). Confidence as a Priority Signal. *Psychological Science, 31*(9), 1084–1096. https://doi.org/10.1177/0956797620925039.

Allen, M., Glen, J. C., Müllensiefen, D., Schwarzkopf, D. S., Fardo, F., Frank, D., … Rees, G. (2017). Metacognitive ability correlates with hippocampal and prefrontal microstructure. *NeuroImage, 149*, 415–423. https://doi.org/10.1016/j.neuroimage.2017.02.008.

Bang, J. W., Shekhar, M., & Rahnev, D. (2019). Sensory noise increases metacognitive efficiency. *Journal of Experimental Psychology: General, 148*(3), 437–452. https://doi.org/10.1037/xge0000511.

Barrett, A. B., Dienes, Z., & Seth, A. K. (2013). Measures of metacognition on signal-detection theoretic models. *Psychological Methods, 18*(4), 535–552. https://doi.org/10.1037/a0033268.

Carpenter, J., Sherman, M. T., Kievit, R. A., Seth, A. K., Lau, H., & Fleming, S. M. (2019). Domain-general enhancements of metacognitive ability through adaptive training. *Journal of Experimental Psychology. General, 148*(1), 51–64. https://doi.org/10.1037/xge0000505.

De Gardelle, V., Faivre, N., Filevich, E., Reyes, G., Rouy, M., Sackur, J., & Vergnaud, J.-C. (2020). *Role of feedback on metacognitive training.*. https://doi.org/10.23668/PSYCHARCHIVES.3138.

Dosher, B. A., & Lu, Z.-L. (1999). Mechanisms of perceptual learning. *Vision Research, 39*(19), 3197–3221. https://doi.org/10.1016/S0042-6989(99)00059-0.

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

Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J., & Rees, G. (2010). Relating introspective accuracy to individual differences in brain structure. *Science (New York, N. Y.), 329*(5998), 1541–1543. https://doi.org/10.1126/science.1191883.

Funder, D. C., & Ozer, D. J. (2019). Evaluating Effect Size in Psychological Research: Sense and Nonsense. Advances in Methods and Practices in Psychological Science, 2(2), 156–168. DOI: 10.1177/2515245919847202.

Guggenmos, M., Wilbertz, G., Hebart, M. N., & Sterzer, P. (2016). Mesolimbic confidence signals guide perceptual learning in the absence of external feedback. ELife, 5, e13388. DOI: 10.7554/eLife.13388.

Haddara, N., & Rahnev, D. (2021). *The impact of feedback on perceptual decision making and metacognition: Reduction in bias but no change in sensitivity.* Psychological Science.

Hainguerlot, M., Vergnaud, J.-C., & de Gardelle, V. (2018). Metacognitive ability predicts learning cue-stimulus associations in the absence of external feedback. *Scientific Reports, 8*(1), 5602. https://doi.org/10.1038/s41598-018-23936-9.

Kiani, R., Corthell, L., & Shadlen, M. N. (2014). Choice certainty is informed by both evidence and decision time. *Neuron, 84*(6), 1329–1342. https://doi.org/10.1016/j.neuron.2014.12.015.

Lu, Z.-L., & Dosher, B. A. (2008). Characterizing observers using external noise and observer models: Assessing internal representations with external noise. *Psychological Review, 115*(1), 44–82. https://doi.org/10.1037/0033-295X.115.1.44.

Lu, Z.-L., Lesmes, L. A., & Dosher, B. A. (2002). Spatial attention excludes external noise at the target location. *Journal of Vision, 2*(4), 4. https://doi.org/10.1167/2.4.4.

Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and Cognition, 21*(1), 422–430. https://doi.org/10.1016/j.concog.2011.09.021.

Maniscalco, B., & Lau, H. (2016). The signal processing architecture underlying subjective reports of sensory awareness. *Neuroscience of Consciousness, 2016*(1). https://doi.org/10.1093/nc/niw002.

McCurdy, L. Y., Maniscalco, B., Metcalfe, J., Liu, K. Y., de Lange, & Lau, H. (2013). Anatomical coupling between distinct metacognitive systems for memory and visual perception. *Journal of Neuroscience, 33*(5), 1897–1906. https://doi.org/10.1523/JNEUROSCI.1890-12.2013.

Metcalfe, J., & Shimamura, A. P. (Eds.). (1994). *Metacognition: Knowing about Knowing*. The MIT Press. https://doi.org/10.7551/mitpress/4561.001.0001.

Palmer, E. C., David, A. S., & Fleming, S. M. (2014). Effects of age on metacognitive efficiency. *Consciousness and Cognition, 28*, 151–160. https://doi.org/10.1016/j.concog.2014.06.007.

Rahnev, D. (2021). Response Bias Reflects Individual Differences in Sensory Encoding. *Psychological Science, 32*, 1157–1168.

Rahnev, D., Desender, K., Lee, A. L. F., Adler, W. T., Aguilar-Lleyda, D., Akdoğan, B., Arbuzova, P., Atlas, L. Y., Balcı, F., Bang, J. W., Bègue, I., Birney, D. P., Brady, T. F., Calder-Travis, J., Chetverikov, A., Clark, T. K., Davranche, K., Denison, R. N., Dildine, T. C., … Zylberberg, A. (2020). The Confidence Database. Nature Human Behaviour, 4(3), 317–325. DOI: 10.1038/s41562-019-0813-1.

Rahnev, D., & Fleming, S. M. (2019). How experimental procedures influence estimates of metacognitive ability. *Neuroscience of Consciousness, 2019*(1), niz009. https://doi.org/10.1093/nc/niz009.

Rahnev, D., Koizumi, A., McCurdy, L. Y., D'Esposito, M., & Lau, H. (2015). Confidence Leak in Perceptual Decision Making. *Psychological Science, 26*(11), 1664–1680. https://doi.org/10.1177/0956797615595037.

Rausch, M., & Zehetleitner, M. (2017). Should metacognition be measured by logistic regression? *Consciousness and Cognition, 49*, 291–312. https://doi.org/10.1016/j.concog.2017.02.007.

Rausch, Manuel, Zehetleitner, Michael, Steinhauser, Marco, & Maier, Martin E. (2020). Cognitive modelling reveals distinct electrophysiological markers of decision confidence and error monitoring. *NeuroImage, 218*, 116963. https://doi.org/10.1016/j.neuroimage.2020.116963.

Rouault, M., Seow, T., Gillan, C. M., & Fleming, S. M. (2018). Psychiatric Symptom Dimensions Are Associated With Dissociable Shifts in Metacognition but Not Task Performance. *Biological Psychiatry, 84*(6), 443–451. https://doi.org/10.1016/j.biopsych.2017.12.017.

Schunk, D. H., & Ertmer, P. A. (2000). *Self-Regulation and Academic Learning. In Handbook of Self-Regulation* (pp. 631–649). Elsevier.

Schwiedrzik, C. M., Singer, W., & Melloni, L. (2011). Subjective and objective learning effects dissociate in space and in time. *Proceedings of the National Academy of Sciences, 108*(11), 4506–4511. https://doi.org/10.1073/pnas.1009147108.

Shekhar, M., & Rahnev, D. (2021a). Sources of Metacognitive Inefficiency. *Trends in Cognitive Sciences, 25*(1), 12–23. https://doi.org/10.1016/j.tics.2020.10.007.

Shekhar, M., & Rahnev, D. (2021b). The nature of metacognitive inefficiency in perceptual decision making. *Psychological Review, 128*(1), 45–70. https://doi.org/10.1037/rev0000249.

van den Berg, R., Yoo, A. H., & Ma, W. J. (2017). Fechner's law in metacognition: A quantitative model of visual working memory confidence. *Psychological Review, 124*(2), 197–214. https://doi.org/10.1037/rev0000060.

van den Berg, R., Zylberberg, A., Kiani, R., Shadlen, M. N., & Wolpert, D. M. (2016). Confidence Is the Bridge between Multi-stage Decisions. *Current Biology: CB, 26*(23), 3157–3168. https://doi.org/10.1016/j.cub.2016.10.021.

Zizlsperger, Leopold, Kümmel, Florian, Haarmeier, Thomas, & Cropper, Simon J (2016). Metacognitive Confidence Increases with, but Does Not Determine. *Visual Perceptual Learning. PLOS ONE, 11*(3), e0151218. https://doi.org/10.1371/journal.pone.0151218.