



# Confidence in action: Differences between perceived accuracy of decision and motor response

Marta Siedlecka<sup>1</sup> · Marcin Koculak<sup>1</sup> · Borysław Paulewicz<sup>1</sup>

Accepted: 13 March 2021 / Published online: 26 April 2021  
© The Author(s) 2021

## Abstract

Each of our decisions is associated with a degree of confidence. This confidence can change once we have acted because we might start doubting our choice or even become convinced that we have made a mistake. In this study, we explore the relations between action and our confidence that our decision was correct or erroneous. Fifty-four volunteers took part in a perceptual decision task in which their decisions could either lead to action or not. At the end of each trial, participants rated their confidence that their decision was correct, or they reported that they had made an error. The main results showed that when given after a response, confidence ratings were higher and more strongly related to decision accuracy, and post-response reports of errors more often indicated actual errors. The results support the view that error awareness and confidence might be partially based on postaction processing.

**Keywords** Confidence · Error awareness · Metacognition · Action · Decision

Each of our decisions is associated with a degree of confidence, but this confidence can change after we have acted because we might start doubting our choice or even become convinced that we have made a mistake. In the study reported in this paper, we explore the relations between action and confidence that our decision was correct or incorrect.

Confidence in decisions and the ability to recognize our errors are manifestations of metacognition (i.e., processes that monitor and regulate other cognitive processes). Interestingly, studies on confidence and error awareness have mostly been run in parallel. In a typical study on confidence, participants make a series of forced choices about external stimuli or their memory content; subsequently, they rate their confidence. In such tasks, confidence is typically measured with a scale ranging from uncertainty to certainty in a given choice. Participants rate their confidence in having responded

correctly, and we have no knowledge about what rating they choose when they know they made a mistake. In studies on error awareness, forced-choice tasks are also used; they are usually easier than those used in confidence studies, but they elicit more errors due to time pressure. Importantly, in studies on error awareness participants do not typically report graded confidence; instead, they are asked to signal when they think they made a mistake or to classify each response as correct or incorrect (Rabbitt, 1968; Steinhäuser & Yeung, 2012; Wessel et al., 2011). To our knowledge, only a few studies have measured both confidence in being right and in being wrong—that is, by using a scale ranging from “certainly wrong” to “certainly correct” (Boldt & Yeung, 2015; Charles & Yeung, 2019; Scheffers & Coles, 2000). Confidence measured in this way has been shown to covary with neural activity associated with error commitment, thus leading some authors to suggest that error reports and confidence judgments result from the same metacognitive processes (Boldt & Yeung, 2015; Charles & Yeung, 2019).

What is the relation between action and metacognitive judgments? Studies have shown that confidence better predicts response accuracy when it is measured after a decision-related response (Pereira et al., 2020; Siedlecka et al., 2019; Wokke et al., 2019). Interpreted in the context of the observed relations between confidence level and action characteristics (Fleming et al., 2015; Susser & Mulligan, 2015), these results

---

This research has been supported by the National Science Centre in Poland with a Sonata grant given to M.S. (2017/26/D/HS6/00059).

✉ Marta Siedlecka  
marta.siedlecka@uj.edu.pl

<sup>1</sup> Consciousness Lab, Institute of Psychology, Jagiellonian University, Ingardena 6, Kraków 30-060, Poland

led to the conclusion that response-related signals inform confidence judgments. However, in these studies motor responses were indistinguishable from decisions, making it difficult to determine whether it is the action or the choice itself that improves metacognitive assessment (Kvam et al., 2015). On the other hand, when the type of the action that precedes a confidence judgment is manipulated experimentally, then certain types of responses (e.g., compatible with the decision-related stimulus characteristics) are shown to increase confidence (Filevich et al., 2020; Siedlecka et al., 2020) without affecting the relation between confidence ratings and decision accuracy (Filevich et al., 2020).

While confidence is usually measured as an assessment of decision accuracy (inferred from motor responses), errors are almost exclusively studied in the context of action control and performance monitoring. These studies are focused on how action selection and execution is evaluated and regulated; therefore, motor errors are of most interest. These errors are generally divided into two types: slips, which are unintended actions (Norman, 1981), and premature responses, which are responses that are given before stimulus processing has completed (Rabbitt, 2002). Such errors could be detected by the monitoring system by registering an action plan that is in competition with an already launched response (Rabbitt & Vyas, 1981) or by detecting conflict between several activated responses (Yeung et al., 2004). However, such errors are often unnoticed and unreported as the processes of error detection and correction can be relatively quick and automatic. Error awareness, on the contrary, is thought to result from a postaction, time-consuming accumulation of information from numerous sources, including sensory and proprioceptive feedback from the erroneous action, and interoception of the autonomic responses that accompany an error (Ullsperger et al., 2010; Wessel, 2012). However, it is not clear whether and how people can make accurate judgments about non-motor-type errors: for example, when an incorrect decision is not associated with any action.

In the presented study, we compared metacognitive judgments between trials in which decisions were followed by action and in which they were not. We combined paradigms used for measuring confidence and error awareness in a task that allowed separation of decision and motor responses. One difficulty in testing the influence of response on metacognitive judgments is that manipulating the presence of motor response does not allow the accuracy of decisions that are not expressed with a behavioural response to be precisely assessed. In this task, participants were asked to act or not act, depending on their decision about the stimuli. This allowed identification of their covert choices, even when they did not overtly respond. Assuming that both types of metacognitive judgments could be affected by action-related information, we hypothesized that confidence and error reports would be more accurate following motor responses. We also expected higher confidence in decisions that were followed by actions.

## Methods

### Participants

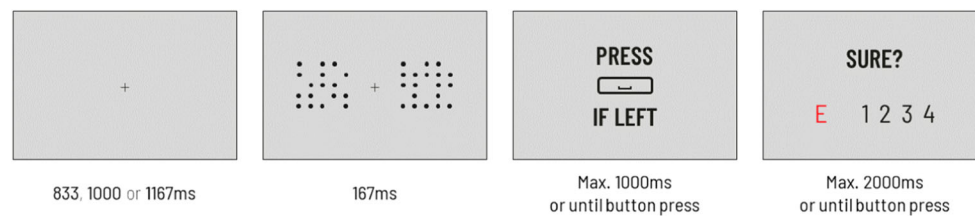
Fifty-four healthy student volunteers (six males), age 19–32 years ( $M = 20.72$ ,  $SD = 2.69$ ), took part in the experiment in return for credit points in a cognitive psychology course. All participants had normal or corrected-to-normal vision and gave written consent to participate in the study. The Ethical Committee of the Institute of Psychology, Jagiellonian University, approved the experimental protocol. Student participants were chosen because this is a group that is relatively easily available to the researchers at our university, and at the same time these participants are used to solving computer tasks. Because of the task novelty, we were not able to estimate sample size, so we decided to test a reasonably large group during available laboratory time (around 50 participants).

### Materials

We used a perceptual decision task, in which participants are asked to decide which of the two presented fields, left or right, contains more dots; then, they rate their confidence in this response (e.g., Boldt & Yeung, 2015). However, the task was modified such that it required a response to a question concerning whether there were more dots in one of the fields. Participants were instructed to press the space bar when there were more dots on one side, or not respond when there were more dots on the other side. In one block, participants used their left thumb to indicate a “left” decision and did nothing for “right” decisions, while in the other block they used their right thumb to indicate a “right” decision and did nothing for “left” decisions. This design allowed us to divide trials into two groups (later called conditions): those in which participants responded (Response condition) and those in which participants did not respond (No response condition). At the end of each trial, participants were asked to rate their confidence in their choice or report an error. The outline of the experimental trial is presented in Fig. 1.

The experiment was run on a PC using PsychoPy software (Peirce et al., 2019) with an LCD monitor (1,920 × 1,080 pixels resolution, 60-Hz refresh rate). Experimental stimuli were composed of two 10 × 10 grids filled with dots of equal size (160 × 160 px). Grids were placed at an equal distance of 120 pixels from the centre of the screen in the horizontal plane. In each grid, a predefined number of dots was placed randomly in each trial. Based on our previous pilot studies with the dot task (with responses for both decisions), the proportions of dots chosen for the experimental session were estimated to lead to an overall accuracy of 70%.

Participants responded by pressing buttons on a standard PC keyboard. The space bar was used for the response in the



**Fig. 1** Outline of the experimental trial. First, a fixation cross was displayed, followed by a brief presentation of the grid stimuli. Immediately after that, participants were asked to decide if there were more dots on one side (here: left) and to report this decision by either

decision task. To measure participants' confidence, we used a 4-point scale: 1 ("I am guessing"), 2 ("I am not confident"), 3 ("I am quite confident"), and 4 ("I am very confident"). Participants used their right hands and the keys U, I, O, P, respectively (labelled 1, 2, 3, 4 in Fig. 1). Additionally, participants could report that they had made an error by pressing the H key (labelled E in Fig. 1).

## Procedure

Participants were tested individually in a soundproof cabin with fixed lighting conditions. After receiving the general instructions, participants were told to place their fingers on the response keys so they would not look at the keyboard during the task. Participants were instructed to respond as accurately as they could. They were also told that the number of trials with correct "left" and "right" answers was equal in each block and that during the whole block they would be asked to respond to the same question ("More dots on the left?" or "More dots on the right?"). Participants were explicitly asked to avoid refraining from responding when they were not confident about the correct decision. They were told that after every 30 trials they would receive feedback about their accuracy and the percentage of trials in which they responded. The experimenters could see the feedback and would remind them about the instruction when they saw that one type of decision was reported more often than the other. Lastly, the experimenter asked participants to try to differentiate their confidence levels between trials even if they felt that the task was generally difficult.

The experiment consisted of three parts: two training sessions and the main task. A trial started with a fixation cross presented centrally on the screen for around one second; this was followed by a brief presentation of the grid stimuli with dots. After the stimuli disappeared, participants saw a screen prompting them to make a decision about the stimuli and to respond or not, according to whether the current block required the "left" or "right" decision to be reported. Finally, a screen with a confidence scale appeared on which participants rated their confidence in their decision or could signal that they made a mistake.

pressing the spacebar (here: if they decided "left") or not responding (here: if they decided "right"). Finally, participants rated their confidence in the correctness of their decision on a 4-point scale; alternatively, they could report that they had made an error

The first training session had a relatively low difficulty level as it was designed to accustom participants to the experimental procedure. This was achieved through presenting the grids with dots for 334 ms, with a fixed dot ratio of 70:30 and 60:40. Participants were allowed 2.5 seconds for stimulus-related decisions, and 3.5 seconds were allowed for the confidence rating and error report. Subsequently, participants went through a second training session that looked like the main task, except for the ratio of the dots (55:45). We introduced variable fixation cross presentation time (833, 1,000, or 1,167 milliseconds) to prevent perceptual entrainment and minimize the effects of expectations. In the main task, the ratio of the dots between the grids was changed and three levels of decision difficulty were introduced: 55:45, 52:48, and 51:49. The time available for the decision was 1 second, while 2 seconds were allowed for the confidence rating and error report.

There were two blocks in the main experimental task: responding when the decision was "left" and responding when the decision was "right." The order of blocks was counterbalanced. The two blocks were bundled into 13 series of 30 trials (summing up to 390 trials per block and 780 trials in the whole experiment). After each series, the participant was provided with a feedback screen (accuracy and proportion of left/right responses) which lasted at least 10 seconds, but they could extend this break if needed.

During this procedure, EEG and EMG recordings were also acquired but these were unrelated to the hypotheses tested in the current study and are not presented here.

## Results

Statistical analyses were run in the R statistical environment (R Core Team, 2019), using linear and logistic mixed-regression models. The models were fitted using the lme4 and lmerTest packages (Bates et al., 2015; Kuznetsova et al., 2017). Series of models with all possible random effects and their interactions were fitted to estimate confidence level and response times (a linear mixed model) and accuracy (a logistic mixed model).

## Data set preparation

The trials in which participants did not rate their confidence were omitted from analysis. We also excluded the data of one participant who differed in scale-use variance from the rest of the group (based on visual inspection of the sorted confidence-rating variances scores plot). The data of 53 participants (38,733 observations) were included in the final data set. Three participants did not succeed in adapting to the new instruction (i.e., responding in the case of a “left” decision in the current block after they were supposed to respond in the case of a “right” decision in the previous block) and their performance in the second block was at or below chance level. Their data from the second block were not included in the analyses. We identified these participants by fitting a Bayesian mixture model, assuming that the participants’ accuracy scores come from a mixture of two distributions: one with an unknown probability of a correct response and one with chance probability.<sup>1</sup>

## Performance in the decision task

The general statistics related to task performance are presented in Table 1. In order to compare accuracy level between Response and No-response trials, we fitted a mixed logistic regression model (with fixed and random effect of condition). The results indicated that participants’ accuracy was higher when they responded compared to trials in which they did not respond ( $z = 2.99, p < .01$ ). Signal detection analysis using mixed logistic regression with the probit function revealed that Response trials were associated with higher discrimination sensitivity ( $z = 2.51, p < .05$ ), but we did not detect significant differences in response bias ( $z = -0.55, p = .58$ ).

Task difficulty manipulation resulted in 87% of correct responses in easy trials, 69% in medium trials, and 60% in hard trials. A mixed logistic regression model with random participant-specific intercept revealed that accuracy differed significantly between hard- and medium-difficulty trials ( $z = -15.73, p < .001$ ), and between easy and medium trials ( $z = 34.69, p < .001$ ). Participants more often chose not to respond than to respond ( $z = -2.92, p < .05$ ), which was revealed by a mixed logistic regression model with random intercept. In order to determine whether the frequency of response and no-response trials differed between the levels of difficulty, we fitted a mixed logistic regression model with random effect of participant-specific intercept. We found differences neither between easy and medium trials ( $z = 0.46, p =$

.65), nor between hard and medium trials ( $z = -0.26, p = .80$ ), nor between easy and hard trials ( $z = -0.72, p = .47$ ).

In order to determine whether trials that required “left” and “right” decisions differed in accuracy, we fitted a mixed logistic regression model with random effect of condition and participant-specific intercept. We did not detect statistically significant differences ( $z = -0.51, p = .61$ ).

## Error reports

A mixed logistic model with fixed and random effect of condition and participant-specific intercept did not reveal a significant difference in the number of reports between the conditions ( $z = 1.85, p = .06$ ). However, error reports were more accurate when they followed motor response: a mixed logistic model with fixed and random effect of condition and participant-specific intercept fitted to data from trials in which an error was reported revealed that in Response trials, the actual error rate was significantly lower than in No response trials ( $z = -7.48, p < .001$ ). The model-based estimations of the proportions of incorrect decisions are .56 (No response) and .85 (Response).

## Confidence ratings

In the following analyses, we omitted trials in which participants reported errors (because a person could either rate their decision confidence or report an error). We compared levels of confidence between conditions using a linear mixed model with fixed and random effect of condition and participant-specific intercept. The analyses showed that reported confidence was higher in Response trials compared with trials in which participants did not respond (see Table 2). When fixed and random effects of accuracy and its interaction with condition were also included in the model, we found that the effect was significant for both correct and erroneous responses (no detected interaction between accuracy and condition),  $t(52) = -0.13, p = .89$ .

One potential confounder of the relation between confidence and response is task difficulty, which might influence both the probability of giving a response and the level of confidence. Therefore, we fitted another model in which we controlled for the dot ratio to find out whether the effect remains statistically significant. We assumed that a decision is easier when the difference between the number of dots on both sides is larger. We fitted a linear mixed model with confidence rating as a dependent variable and fixed effects of condition, difficulty (three levels), and random effect of condition and participant-specific intercept (the model with all random effects did not converge). The model is presented in Table 3. Intercept refers to the No response condition and the lowest level of difficulty (easy). Independently of difficulty, the Response coefficient estimates the average effect of condition on confidence level. Importantly, this effect remains statistically significant. The other two coefficients estimate the

<sup>1</sup> We used only principled data exclusion criteria. It is possible to further clean the data—for example, by excluding participants that performed worse in the second than the first block, or by removing a subset of initial trials in which performance was lower than in the other trials. However, this would require arbitrary decisions concerning how much lower accuracy would have to be or how much slower the responses would have to be to exclude a subset of trials or participants from further analyses.

**Table 1** Descriptive statistics: Decision accuracy, number of trials, number of error reports, and error rate when reporting an error in the response and no-response conditions

	Decision accuracy	Number of trials	Frequency of error report	Error rate when reporting error
Response	72%	18,506 (48%)	535 (2.9%)	84%
No response	70%	20,227 (52%)	501 (2.5%)	56%

difference in confidence between the lower level of difficulty and the two other levels; they show that confidence is significantly lower in medium and difficult trials compared to easy trials.

### The relation between decision accuracy and confidence rating

In this analysis, we tested whether the relationship between decision accuracy and confidence ratings differed between the two conditions by comparing the extent to which confidence ratings predicted accuracy in the decision task. Here, we used logistic regression analysis because we did not want to rely on strong untested assumptions about the source of confidence that are implicit in popular alternatives based on signal detection theory (Paulewicz et al., 2020).

We fitted a mixed logistic regression model to decision accuracy data with condition (two levels), confidence rating (four levels), and their interaction as fixed and random effects. To improve the readability of the regression coefficients, confidence ratings were centred on the lowest value (“guessing”), and the reference condition was the No response. Therefore, the regression slope reflects the relation between accuracy and confidence rating, while the intercept informs about the difference between chance level and decision accuracy when participants reported “guessing” in the No response condition.

The intercept was not significant; we did not detect a difference between participants’ decision accuracy when reporting guessing and chance level in the No response condition ( $z = 1.92, p = .05$ ). This lack of significant difference from chance-level performance when reporting guessing was also not significantly related to condition ( $z = -0.89, p = .38$ ). However, the model revealed a statistically significant

**Table 2** The linear mixed-model analysis of the relationship between the presence of response and confidence rating

	Estimate	SE	df	<i>t</i>	<i>p</i>
Intercept	1.74	0.08	52.01	22.93	<.001***
Response	0.08	0.02	49.18	3.45	<.01**

Note. Model syntax: Confidence rating ~ Condition + (Condition | Participant)

\*\* $p < .01$ . \*\*\* $p < .001$

relationship between decision accuracy and confidence rating in the No response condition ( $z = 13.64, p < .001$ ), and this relationship was significantly stronger in the Response condition ( $z = 2.88, p < .01$ ). The model coefficients are presented in Table 4, and model fit, together with percentage of correct responses for each confidence rating and frequency of each rating are presented in Fig. 2.

### Response times

For exploratory reasons, we analyzed reaction times in the Response condition. Correct responses were given faster than incorrect responses,  $t(52.19) = -6.31, p < .001$  (a mixed linear model analysis with fixed and random effect of accuracy and participant-specific intercept). Analysis of incorrect response times showed that reported errors were faster than unreported errors,  $t(54.14) = 4.84, p < .001$ , a mixed linear model with random intercept. Slower responses were followed by lower confidence ratings,  $t(47) = -6.34, p < .001$  (a linear mixed model with fixed and random effect of confidence).

We also compared confidence rating time between conditions, taking decision accuracy into consideration. We assumed that in no-response trials, it is possible that participants’ responses were not omitted on purpose, but were delayed due to decision-related uncertainty. In such a case, confidence ratings times should be longer in the No response condition than in the Response condition, regardless of whether a decision was correct or incorrect. We fitted a linear mixed regression model with confidence rating time as a dependent variable, condition and decision accuracy as fixed effects, and random effect of condition and participant-specific intercept (the

**Table 3** The linear mixed-model analysis of the relationship between condition, difficulty, and confidence rating

	Estimate	SE	df	<i>t</i>	<i>p</i>
Intercept	1.96	0.08	51.01	25.64	<.001***
Response	0.08	0.02	49.41	3.39	.001**
Difficulty medium	-0.36	0.03	51.80	-12.03	<.001***
Difficulty hard	-0.30	0.03	52.36	-12.10	<.001***

Note. Model syntax: Confidence rating ~ Condition + Difficulty + (Condition | Participant)

\*\* $p < .01$ . \*\*\* $p < .001$

**Table 4** The mixed logistic regression model analysis of the relationship between decision accuracy and confidence rating in both conditions

	Estimate	SE	z	p
Intercept	0.12	0.06	1.92	.05
Response	−0.06	0.06	−0.89	.38
Confidence	0.47	0.03	13.64	< .001***
Confidence: Response	0.11	0.04	2.88	< .01**

Note. Model syntax: Decision Accuracy ~ Confidence rating × Condition + (Confidence rating × Condition | Participant)

\*\* $p < .01$ . \*\*\* $p < .001$

model with random effects of both decision accuracy and condition did not converge). Confidence rating times were longer for incorrect decisions in No response trials compared to Response trials,  $t(68.01) = -2.85$ ,  $p = .006$ , but we did not detect a difference in the case of correct decisions,  $t(54.74) = 0.73$ ,  $p = .47$ . The model coefficients are presented in Table 5.

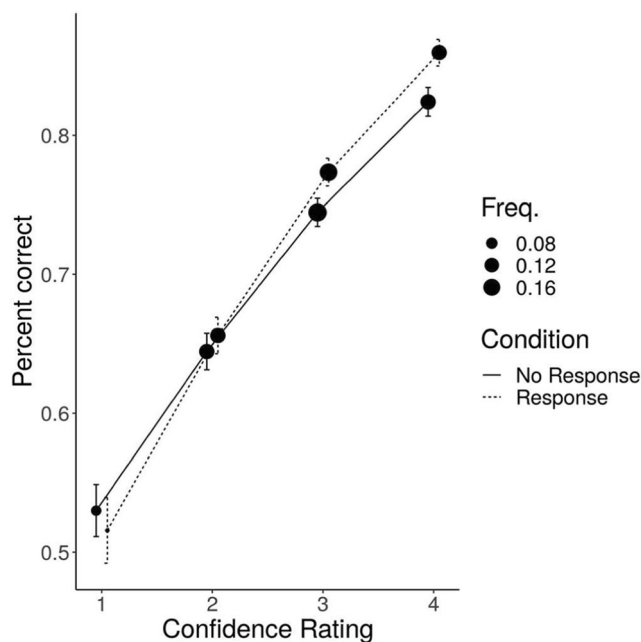
## Discussion

In this experiment, we compared confidence judgments and error reports between trials in which, depending on their decision about the stimulus, participants did or did not respond. The data showed that when given after a response, confidence

ratings predicted decision accuracy better and reports of errors more often indicated an actual error. The data also showed that after responding participants reported higher confidence in both correct and incorrect decisions and independently of task difficulty. Additionally, analyses of reaction times showed that slower responses were followed by lower confidence ratings, while fast responses were associated with high confidence that a decision was correct or incorrect.

The results support previous findings showing that confidence judgments are more strongly related to decision accuracy when measured after a person responds (Pereira et al., 2020; Siedlecka et al., 2019; Wokke et al., 2019); this suggests that motor response provides metacognitive processes with additional information such as deliberation time, motor fluency, or internal accuracy feedback (Kiani et al., 2014; Susser & Mulligan, 2015; Ullsperger et al., 2010). Moreover, the results are consistent with findings from studies in which the occurrence of motor response was experimentally manipulated, showing that confidence in decisions is increased by action paired with decision-relevant stimulus characteristics (Filevich et al., 2020; Siedlecka et al., 2020). Since confidence was higher after both correct and incorrect motor responses, this supports the interpretation that carrying out a response increases confidence by indicating successful completion of the decisional process (Siedlecka et al., 2020).

To our knowledge, this is the first study to compare the accuracy of error reports between decisions that were and were not followed by motor responses. The data showed that even though participants reported a similarly small number of errors in both conditions, their error reports were more accurate after they had responded. That means that participants in no-response trials were more often mistakenly convinced that they had made an erroneous decision. On one hand, inaccurate error reports together with lower decision accuracy and confidence level in no-response trials could be interpreted as another sign of higher uncertainty. However, we would expect that in the case of high decision uncertainty, participants would rather report guessing than an error (Scheffers & Coles, 2000). Studies on confidence, including this one, show that slower and more difficult decisions are associated with lower confidence (Desender et al., 2017; Kiani et al., 2014). At the same time, studies on error awareness indicate that errors are reported when sufficiently strong evidence (compared with noise and evidence for an already given response) passes the awareness threshold (Ullsperger et al., 2010; Wessel, 2012). Interestingly, even though there are theories explaining how and when error awareness emerges, not much is known about the possible causes of error illusions. On one hand, if a decision-related action increases confidence by indicating that the decision process has completed, lack of action following the decision might lead to internal error signals. At the same time, without action, accuracy assessment is based on fewer sources of information. Moreover, in some



**Fig. 2** Model fit for the relationship between decision accuracy and confidence ratings in each condition. The position of filled circles represents average accuracy for each scale point. The bars represent 95% confidence intervals. The frequency describes the proportion of each confidence rating

**Table 5** The linear mixed-model analysis of the relationship between decision accuracy, condition, and confidence rating time

	Estimate	SE	df	<i>t</i>	<i>p</i>
Intercept (Incorrect decision: No response)	0.58	0.02	55.36	34.56	<.001***
Intercept (Correct decision: No response)	0.50	0.02	52.57	29.79	<.001***
Incorrect decision: Response	−0.04	0.01	68.01	−2.85	.006**
Correct decision: Response	0.01	0.01	54.47	0.73	.47

Note. Model syntax: Confidence rating time  $\sim -1 +$  Decision Accuracy / Condition + (Condition | Participant)

\*\**p* < .01. \*\*\**p* < .001

cases, this information might be unreliable—for example, when a chosen option is not associated with any action plan, even weak activation of a response linked to an unchosen option might bias the judgment towards an error report.

The results of this study are consistent with the view that both confidence judgments and error reports are informed by postdecisional processing. In most confidence models, however, postdecisional processing refers to ongoing decision-related evidence and not to action-related information (e.g., Moran et al., 2015; Pleskac & Busemeyer, 2010). One explanation of the role of action in metacognitive judgments is offered by the hierarchical Bayesian model of self-evaluation, according to which such judgments are based on second-order inference about the performance of the decision system. Crucially, decision and confidence are computed separately in this model, and action provides information about the decision process that is not otherwise accessible to the metacognitive system (Fleming & Daw, 2017). We propose that confidence and error judgments are the results of metacognitive processes that consist of multiple monitoring and regulating loops that occur at consecutive stages of stimuli identification, decision-making, and action preparation and execution (see also Paulewicz et al., 2020). Therefore, each stage of this process, before and after action, is monitored and might trigger regulation. For example, studies have shown that the monitoring system can detect, inhibit, and correct erroneous response activation (so-called partial errors) before the response is given, which therefore helps avoid error commission (Burle et al., 2002; Scheffers et al., 1996).

The main limitation of this study is its observational assessment of the relation between confidence level and action. Although confidence is often assumed to reflect the assessment of the whole decisional process, it might determine the outcome of this process, and more importantly, if and how people will act on their decisions. The choice of action might be affected by the results of monitoring the decision process, and confidence judgment at the end of a trial might be the result of combined preresponse and postresponse levels of

confidence. In this study we cannot exclude the possibility that in some of the trials participants chose not to respond when they were not confident in their decision. However, the relation between the presence of a motor response and confidence remained significant when decision accuracy and trial difficulty were controlled for. This suggests that confidence level was higher in the response compared with the no-response condition, even in easy trials in which uncertainty was low and in trials in which a corrected decision was made.

Importantly, although it could be the level of confidence in a decision that affects a person's decision about responding or not, confidence that one has committed an error can only occur after this decision or action has been done. It is plausible that the differences in the accuracy of error reports were due to different types of errors, which were in turn related to predecisional confidence. While mistakes in No response trials could have mostly resulted from high uncertainty about stimulus characteristics, in response trials we might have observed slips and premature responses. However, as there was no prepotent response in our procedure (both “left/right” decisions and overt/covert choices had the same probability of being correct), this explanation would suggest that incorrect responses were given quickly because participants were confident in their decisions, and then this confidence changed into confidence that an error had been committed.

Another potential limitation of this study is that we did not embed the error report into the confidence scale; thus, we did not collect ratings of confidence in having committed an error. Based on our pilot data, we expected the number of responses for each scale point on the error side of the confidence scale to be very low, therefore making it difficult to estimate the accuracy of error reports. Additionally, it is not clear how to analyze and interpret such data—for example, should we equate low error confidence with error report or with low confidence in decision generally? There does not seem to be a confidence theory explaining the relations between confidence in having made an error and confidence in having given a correct response.

To summarize, the results of this study support the view that action plays an important role in shaping subjective confidence

and allows better assessment of decision accuracy. It would be informative in future studies to find a way of probing subjective confidence at different stages of the decision-making process to find out how it changes before and after decisions and actions, or how it predicts commitment to action.

**Acknowledgement** M.S. and M.K. conceived the plan of the study. M.K. prepared the experimental procedure and ran the tests. B.P. analyzed the data in collaboration with M.S. M.S. wrote the manuscript, and B.P. and M.K. provided comments.

We would like to thank Justyna Hobot and Magdalena Senderecka for commenting on the manuscript, and Kinga Ciupińska and Alicja Krzyżewska for their help with data collection.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Boldt, A., & Yeung, N. (2015). Shared neural markers of decision confidence and error detection. *Journal of Neuroscience*, 35(8), 3478–3484. <https://doi.org/10.1523/JNEUROSCI.0797-14.2015>
- Burle, B., Possamaï, C. A., Vidal, F., Bonnet, M., & Hasbroucq, T. (2002). Executive control in the Simon effect: An electromyographic and distributional analysis. *Psychological Research*, 66(4), 324–336. <https://doi.org/10.1007/s00426-002-0105-6>
- Charles, L., & Yeung, N. (2019). Dynamic sources of evidence supporting confidence judgments and error detection. *Journal of Experimental Psychology: Human Perception and Performance*, 45(1), 39–52. <https://doi.org/10.1037/xhp0000583>
- Desender, K., Van Opstal, F., & Van den Bussche, E. (2017). Subjective experience of difficulty depends on multiple cues. *Scientific Reports*, 7(1), 44222. <https://doi.org/10.1038/srep44222>
- Filevich, E., Koß, C., & Faivre, N. (2020). Response-related signals increase confidence but not metacognitive performance. *Eneuro*. <https://doi.org/10.1523/ENEURO.0326-19.2020>
- Fleming, S. M., & Daw, N. D. (2017). Self-evaluation of decision-making: A general Bayesian framework for metacognitive computation. *Psychological Review*, 124(1), 91–114. <https://doi.org/10.1037/rev0000045>
- Fleming, S. M., Maniscalco, B., Ko, Y., Amendi, N., Ro, T., & Lau, H. (2015). Action-specific disruption of perceptual confidence. *Psychological Science*, 26(1), 89–98. <https://doi.org/10.1177/0956797614557697>
- 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>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>
- Kvam, P. D., Pleskac, T. J., Yu, S., & Busemeyer, J. R. (2015). Interference effects of choice on confidence: Quantum characteristics of evidence accumulation. *Proceedings of the National Academy of Sciences of the United States of America*, 112(34), 10645–10650. <https://doi.org/10.1073/pnas.1500688112>
- Moran, R., Teodorescu, A. R., & Usher, M. (2015). Post-choice information integration as a causal determinant of confidence: Novel data and a computational account. *Cognitive Psychology*, 78, 99–147. <https://doi.org/10.1016/j.cogpsych.2015.01.002>
- Norman, D. A. (1981). Categorization of action slips. *Psychological Review*, 88(1), 1–15. <https://doi.org/10.1037/0033-295X.88.1.1>
- Paulewicz, B., Siedlecka, M., & Koculak, M. (2020). Confounding in studies on metacognition: A preliminary causal analysis framework. *Frontiers in Psychology*, 11, 1933. <https://doi.org/10.3389/fpsyg.2020.01933>
- Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51, 195–203. <https://doi.org/10.3758/s13428-018-01193-y>
- Pereira, M., Faivre, N., Iturrate, I., Wirthlin, M., Serafini, L., Martin, S., Desvachez, A., Blanke, O., Van De Ville, C., & Millan, J. D. R. (2020). Disentangling the origins of confidence in speeded perceptual judgments through multimodal imaging. *Proceedings of the National Academy of Sciences of the United States of America*, 117(15), 8382–8390. <https://doi.org/10.1073/pnas.1918335117>
- Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: A theory of choice, decision time, and confidence. *Psychological Review*, 117, 864–901. <https://doi.org/10.1037/a0019737>
- R Core Team. (2019). R: A language and environment for statistical computing [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rabbitt, P. (1968). Three kinds of error-signalling responses in a serial choice task. *The Quarterly Journal of Experimental Psychology*, 20, 179–188. <https://doi.org/10.1080/14640746808400146>
- Rabbitt, P. (2002). Consciousness is slower than you think. *The Quarterly Journal of Experimental Psychology Section A*, 55(4), 1081–1092. <https://doi.org/10.1080/02724980244000080>
- Rabbitt, P., & Vyas, S. (1981). Processing a display even after you make a response to it: How perceptual errors can be corrected. *The Quarterly Journal of Experimental Psychology*, 33(3), 223–239. <https://doi.org/10.1080/14640748108400790>
- Scheffers, M. K., & Coles, M. G. (2000). Performance monitoring in a confusing world: Error-related brain activity, judgments of response accuracy, and types of errors. *Journal of Experimental Psychology: Human Perception and Performance*, 26(1), 141–151. <https://doi.org/10.1037/0096-1523.26.1.141>
- Scheffers, M. K., Coles, M. G., Bernstein, P., Gehring, W. J., & Donchin, E. (1996). Event-related brain potentials and error-related processing: An analysis of incorrect responses to go and no-go stimuli. *Psychophysiology*, 33(1), 42–53. <https://doi.org/10.1111/j.1469-8986.1996.tb02107.x>
- Siedlecka, M., Paulewicz, B., & Koculak, M. (2020). Task-related motor response inflates confidence [Preprint]. *BiorXiv*. <https://doi.org/10.1101/2020.03.26.010306>
- Siedlecka, M., Skóra, Z., Paulewicz, B., Fijałkowska, S., Timmermans, B., & Wierchoń, M. (2019). Responses improve the accuracy of confidence judgements in memory tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45(4), 712–723. <https://doi.org/10.1037/xlm0000608>
- Steinhauser, M., & Yeung, N. (2012). Error awareness as evidence accumulation: Effects of speed-accuracy trade-off on error signaling.



- Frontiers in Human Neuroscience*, 6. <https://doi.org/10.3389/fnhum.2012.00240>
- Susser, J. A., & Mulligan, N. W. (2015). The effect of motoric fluency on metamemory. *Psychonomic Bulletin & Review*, 22(4), 1014–1019. <https://doi.org/10.3758/s13423-014-0768-1>
- Ullsperger, M., Harsay, H. A., Wessel, J. R., & Ridderinkhof, K. R. (2010). Conscious perception of errors and its relation to the anterior insula. *Brain Structure and Function*, 214(5/6), 629–643. <https://doi.org/10.1007/s00429-010-0261-1>
- Wessel, J. R. (2012). Error awareness and the error-related negativity: Evaluating the first decade of evidence. *Frontiers in Human Neuroscience*, 6. <https://doi.org/10.3389/fnhum.2012.00088>
- Wessel, J. R., Danielmeier, C., & Ullsperger, M. (2011). Error Awareness Revisited: Accumulation of Multimodal Evidence from Central and Autonomic Nervous Systems. *Journal of Cognitive Neuroscience*, 23(10), 3021–3036. <https://doi.org/10.1162/jocn.2011.21635>
- Wokke, M. E., Achoui, D., & Cleeremans, A. (2019). Action information contributes to metacognitive decision-making. *BioRxiv*, 657957. <https://doi.org/10.1101/657957>
- Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychological Review*, 111(4), 931–959. <https://doi.org/10.1037/0033-295X.111.4.931>

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