Perceptual decision confidence is sensitive to forgone physical effort expenditure

William Turner[1*#], Raina Angdias[1*], Daniel Feuerriegel[1], Trevor T-J Chong[2,3,4], Robert Hester[1], Stefan Bode[1, 5,]

[1]Melbourne School of Psychological Sciences, The University of Melbourne, Victoria 3010, Australia
[2] Turner Institute for Brain and Mental Health, Monash University, Victoria 3800, Australia
[3] Department of Neurology, Alfred Health, Melbourne, Victoria 3004, Australia
[4] Department of Clinical Neurosciences, St Vincent's Hospital, Victoria 3065, Australia
[5]Department of Psychology, The University of Cologne, Cologne, Germany
* These authors contributed equally to this work
[#] Corresponding author: William Turner, Email: williamfrancisturner@gmail.com, Address: Melbourne School of Psychological Sciences, The University of Melbourne, Parkville, Victoria, 3010, Australia

Abstract

Contemporary theoretical accounts of metacognition propose that action-related information is used in the computation of perceptual decision confidence. We investigated whether the amount of expended physical effort, or the 'motoric sunk cost' of a decision, influences perceptual decision confidence judgements in humans. In particular, we examined whether people feel more confident in decisions which required more effort to report. Forty-two participants performed a luminance discrimination task that involved identifying which of two flickering grayscale squares was brightest. Participants reported their choice by squeezing hand-held dynamometers. Across trials, the effort required to report a decision was varied across three levels (low, medium, high). Critically, participants were only aware of the required effort level on each trial once they had initiated their motor response, meaning that the varying effort requirements could not influence their initial decisions. Following each decision, participants rated their confidence in their choice. We found that participants were more confident in decisions that required greater effort to report. This suggests that humans are sensitive to motoric sunk costs and supports contemporary models of metacognition in which actions inform the computation of decision confidence.

Keywords: decision confidence, physical effort, motor costs, metacognition, sunk costs

1. Introduction

Every decision we make is associated with a degree of confidence (reflecting the subjective likelihood that a decision was correct or appropriate). Neural activity patterns in humans, monkeys, and rats correlate closely with confidence estimates derived from formal models, suggesting that metacognitive monitoring of decision behaviour occurs in these species (Bang & Fleming, 2018; Kepecs, Uchida, Zariwala, & Mainen, 2008; Middlebrooks & Sommer, 2011). Moreover, confidence estimates are also associated with patterns of learning and decision-making, suggesting that metacognitive information is used to guide behaviour (Folke, Jacobsen, Fleming, & De Martino, 2017; Kepecs et al., 2008; Middlebrooks & Sommer, 2011; Van Den Berg, Zylberberg, Kiani, Shadlen, & Wolpert, 2016). For example, rats abandon potential rewards when decisions are less certain (Kepecs et al., 2008), monkeys wager bets in a manner consistent with the use of metacognitive information to maximise rewards across time (Middlebrooks & Sommer, 2011), and humans take more care (i.e. gather more evidence) in making the second of two linked decisions when they are more confident in their first decision (Van Den Berg et al., 2016). Given the importance of confidence for guiding future behaviour, it is important to understand the factors that feed into decision confidence estimates.

One factor is the action associated with reporting the outcome of a decision. Intuitively, if the act of reporting a choice (e.g., pressing a button or moving a lever) is irrelevant to the decision itself, it should not affect decision confidence. However, an emerging view within the metacognition literature is that various sources of sensory- and action-related information are integrated when estimating decision confidence. According to a recent model by Fleming and Daw (Fleming & Daw, 2017), it may be beneficial for an organism to integrate action-related information when sensory evidence is limited, or feedback is absent.

Consistent with this view, a number of studies have provided evidence that action-related information can indeed affect perceptual confidence judgements (Faivre, Filevich, Solovey, Kühn, & Blanke, 2018; Faivre et al., 2020; Fleming et al., 2015; Palser, Fotopoulou, & Kilner, 2018; Pereira et al., 2020; Siedlecka, Hobot, et al., 2019; Siedlecka, Paulewicz, & Koculak, 2020; Siedlecka, Skóra, et al., 2019; Wokke, Achoui, & Cleeremans, 2020). For example, Fleming and colleagues (2015) applied single-pulse TMS to the dorsal premotor cortex before and after responses during a visual discrimination task. They found increased confidence when participants made a correct response that was congruent with the stimulation and decreased confidence when participants made a correct response incongruent with the stimulation. It has also been shown that metacognitive judgements in both perceptual and memory tasks tend to be more accurate (i.e. more closely correspond to the objective accuracy of a decision) when they follow a behavioural response (Pereira et al., 2020; Siedlecka, Skóra, et al., 2019). In addition, perceptual awareness ratings (Siedlecka, Hobot, et al., 2019) and perceptual confidence ratings (Siedlecka et al., 2020) have been shown to be higher following task-compatible cued motor responses, compared to task-neutral cued responses. Taken together, these studies broadly demonstrate that action-related information can affect perceptual confidence judgements.

Critically however, one question which previous studies did not address is whether fine-grained action information, such as the *degree of physical effort* expended to report a decision (i.e. the ‘motor cost’ of a decision), affects subsequent reports of decision confidence. Recently, it was shown that the presence of subthreshold muscle activation preceding a response, as well as the force with which the response (a keypress) was made, correlated positively with subsequent judgements of decision confidence (Gajdos, Fleming, Saez Garcia, Weindel, & Davranche, 2019). However, as muscle activation and response force were not experimentally manipulated, it is unclear whether higher confidence judgements simply co-occurred with greater muscle activation in the same trials. As such, it remains unclear whether the motor cost of a decision affects decision confidence.

Emerging evidence suggests that motor costs can also affect the way in which perceptual decisions are made. In an experiment by Hagura and colleagues (2017) participants moved one of two manipulanda to indicate their choice in a random dot motion task. Unbeknownst to the participants, physical resistance (i.e. motor costs) gradually increased for one manipulandum over the course of the experiment. Despite being unaware of this asymmetry, participants were biased against making responses that required more effort, and this bias carried over to a subsequent verbal-response task using the same stimuli. This suggest that motor costs can affect perceptual decision-making processes that are not strictly related to action selection. However, in this study the motor costs could be anticipated, and confidence ratings were not recorded. Consequently, it could not be determined whether expended motor costs (as opposed to anticipated motor costs) affected decision confidence.

The amount of effort one invests into reporting a decision can be thought of as a 'sunk cost'. Sunk cost errors are said to occur when individuals continue pursuing an action due to prior, and therefore irretrievable, investments (Arkes & Blumer, 1985). Recently, Sweis and colleagues (2018) showed that humans, rats, and mice are susceptible to a temporal sunk cost bias. In their experiment, subjects were offered to wait a short duration to obtain a reward in each trial. Critically, after accepting an offer, subjects were free to abandon the decision to wait at any point during the waiting period. Sweis and colleagues (2018) showed that the likelihood of continuing and obtaining the reward, rather than abandoning the decision, increased when more time had already been invested. Given this finding, we hypothesised that the degree of effort one invests into reporting a choice may similarly act as 'motoric sunk cost', which will increase decision confidence.

1.1 The Current Study

To investigate the relationship between expended motor costs and decision confidence, we employed a dynamic luminance discrimination task in which participants indicated which of two flickering grey squares was brightest. Participants reported their decision by squeezing one of two hand-held dynamometers. Critically, the effort required to report a choice (i.e. how hard participants needed to squeeze) was varied across three levels (low, medium, high). It was important to directly manipulate effort in this manner, since simply looking for associations between effort and confidence would not allow us to infer the directionality of any observed effect (i.e. positive associations could equally be driven by participants investing more effort into decisions they are highly confident are correct). The effort condition was also revealed only *after* participants had initiated their squeeze response, making it impossible for this information to influence the actual decisions they made. Each decision was followed by a confidence report (indicating how confident participants were in having responded correctly) ranging from 0% (certainly wrong) to 100% (certainly correct). Drawing from sunk cost theory and contemporary models of metacognition, we hypothesised that participants would be more confident in having responded correctly for decisions which they had invested greater effort into reporting.

2. Materials and Methods

2.1. Participants

Fifty participants aged between 18 and 42 years (*M* = 23.9, *SD* = 4.23) were recruited via advertisements on campus and online. This sample size was chosen prior to collecting any data. We chose to approximately double the sample size used in previous studies which investigated associations between action and confidence (Fleming et al., 2015; Gajdos et al., 2019; Siedlecka, Hobot, et al., 2019) to ensure sufficient statistical power. Participants gave written informed consent prior to participation and were reimbursed $20 for their time. The experiment advertised reimbursement of $15 with the opportunity to earn an extra $5 to incentivise task performance, however all participants were ultimately paid the full amount. Participants were fluent in English, had normal or corrected-to-normal vision, and no history of neurological or psychiatric conditions. The study was approved by the Human Ethics Committee at the Melbourne School of Psychological Sciences, ID 1749955.3.

Five participants were excluded as the staircasing procedure (see below) did not produce sensible accuracy values for easy and hard difficulty conditions (i.e. the easy condition trials ended up being more difficult than hard condition trials). Two participants were excluded due to better performance on hard, rather than easy trials. One participant was excluded due to the lack of variability in their choices, as one response option was chosen in 85.13% (395/464) of completed trials, suggesting disengagement with the task. The final sample consisted of $N = 42$ participants aged between 19 and 42 years old ($M = 23.98$, $SD = 4.30$).

2.2 Materials

Stimuli consisted of two flickering grayscale squares (70 x 70 pixels, ~ 2.18 x 2.18 degrees of visual angle) presented side-by-side, equidistant from the centre and spaced 70 pixels apart horizontally. Individual frame RGB values were randomly sampled from Gaussian distributions centred around mean values that differed depending on the stimulus difficulty condition. There were two stimulus difficulty conditions (easy and hard). Mean RGB values for these two conditions were obtained from staircasing procedures (see below), meaning that the mean RGB values differed across participants. The difference in mean RGB values between the brighter and darker squares ranged from 11-34 ($M = 20.98$) in the easy stimulus condition, and 4-21 ($M = 12.36$) in the hard stimulus condition. The distributions of individual frame RGB values had standard deviations of 25.5 and were truncated to two standard deviations from the mean.

Stimuli were presented on a Sony Trinitron G420 CRT monitor (Resolution 1280 x 1024, Refresh Rate 75 Hz) that was gamma-corrected using a ColorCAL MKII Colorimeter. The paradigm was programmed in MATLAB 2015b using Psychophysics Toolbox Version 3.0.14 (Brainard, 1997; Kleiner et al., 2007). Participants used a pair of Biopac TSD121C Hand Dynamometers (one gripped in each hand) and a standard computer mouse and keyboard throughout the experiment. Participants were seated ~ 50 cm from the screen and performed the experiment in a darkened room. The dynamometers were affixed to a custom-made frame at a comfortable distance such that participants could grip them while resting their forearms on the table.

2.3 Procedure

2.3.1 Calibration phase

The hand dynamometers were calibrated to control for individual differences in hand strength. To calibrate the dynamometers, participants were instructed to squeeze the handles with as much force as possible. This was done to measure the force of their maximum voluntary contraction (MVC). A proportion of participants' MVC determined the amount of force participants needed to exert to submit a response in the three effort conditions (low = 20%, medium = 40%, high = 60%). Participants calibrated the dynamometers twice throughout the experiment—once prior to the experiment, and a second time mid-way through the experiment (between the fifth and sixth block) to control for fatigue.

Following the initial dynamometer calibration and prior to the main experiment, participants performed two short sessions in which interleaved staircase procedures were used to control for inter-individual variation in task aptitude. This included a three-down-one-up and a two-down-one-up staircase consisting of 200 trials of the luminance discrimination task. Participants responded using the left and right arrow keys of a keyboard and were provided visual feedback on the monitor ("correct" or "error"). Participants were not required to use dynamometers or report their decision confidence during the staircase procedure. The staircase procedure calibrated the mean brightness levels (i.e. stimulus difficulty) and achieved an average performance accuracy of 86.77% in the easy difficulty condition 75.75% in the hard difficulty condition.

2.3.2 Experiment phase

The main experiment consisted of 480 trials. The trial structure is depicted in Fig. 1. Participants completed ten blocks of 48 trials each with self-paced breaks in between. In each trial, a fixation point was presented for 500 ms, after which the stimuli appeared for 400 ms. Participants were asked to identify which of the two squares (left or right) was brighter overall. Participants were able to respond from 150 ms after stimulus onset. Following stimulus presentation (or upon squeezing a dynamometer, if participants responded before stimulus offset), two empty response columns and a red horizontal line (representing the amount of force required to submit a response) appeared on the screen. Importantly, the red horizontal line representing the required response force appeared only *after* participants indicated their choice by squeezing one of the dynamometers. This means that participants did not know how much effort would be required on a trial before they began responding. As participants continued to squeeze, a dynamic yellow bar filled the column according to the amount of force exerted. Participants were instructed to continue squeezing until the column was ‘filled’ to the red line (i.e. the response threshold) whereby a response would be submitted. Hence, the position of the red threshold determined the amount of force needed to submit a response, and this varied across three effort conditions (low, medium, and high). The three effort conditions were randomised within blocks. Participants were also prevented from changing their decision during this stage, as once one dynamometer was squeezed, the alternate dynamometer could not register a response. Participants were given 2,000 ms to respond. If participants were unable to respond in time, the feedback “Too Slow” appeared, and participants proceeded to the next trial (this occurred on ~3% of trials). Following response submission and a brief delay, participants were given 3,600 ms to report their confidence. Participants controlled a mouse with their right hand and clicked anywhere along a horizontal confidence scale ranging from 0% (certainly wrong) to 100% (certainly correct). The exact mid-point of the scale could not be selected to prevent participants from reporting a purely guessing response. If participants did not respond in time, the words "Too Slow" appeared and they proceeded to the next trial.

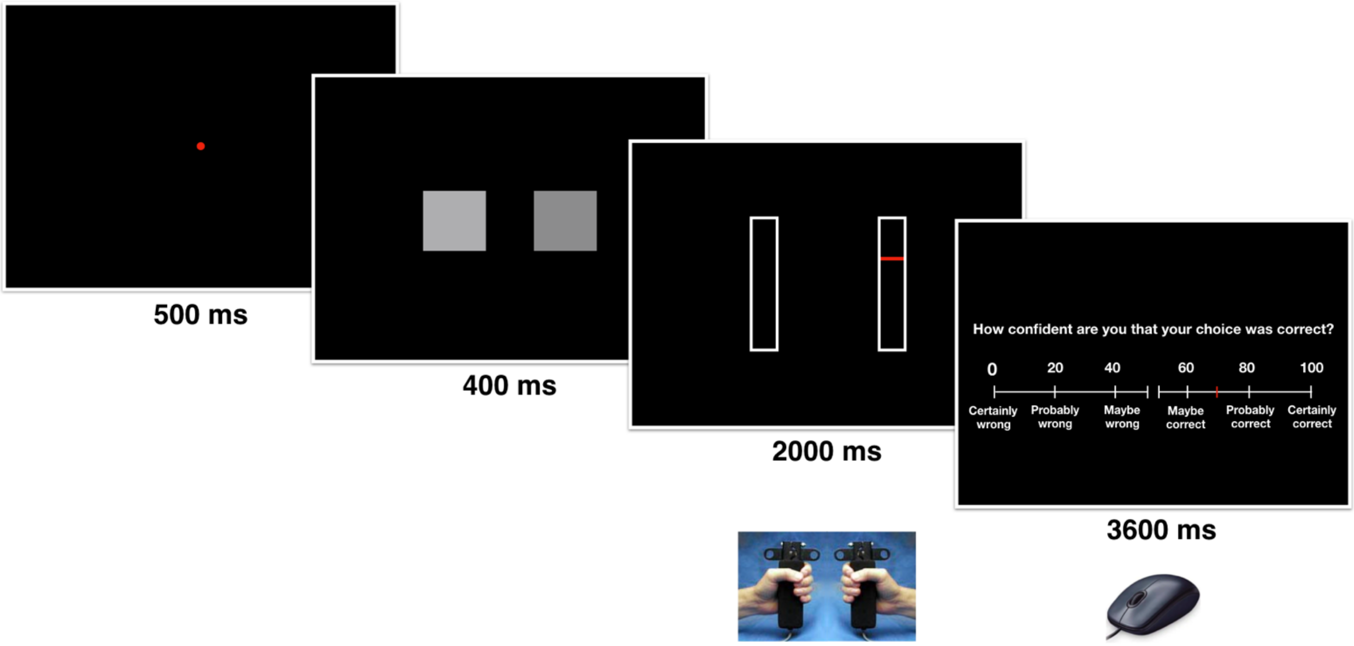

Fig. 1. Task schematic. A fixation point was presented for 500 ms. The stimuli were then presented for 400 ms. Once the participant squeezed one dynamometer, a red horizontal line appeared to indicate the amount of effort participants needed to exert to submit a response. As participants continued the squeeze, a dynamic yellow bar filled the column up to the red line, whereby a response was submitted. Participants were given 2,000 ms from stimulus onset to submit a decision. Then, a confidence scale appeared for 3,600 ms and participants needed to make a confidence judgment within that time. For their subsequent confidence reports, participants were able to click anywhere along the scale, excluding the absolute centre. The cursor controlled a dynamic red vertical line that provided visual feedback of the cursor's position along the scale. The red vertical line initially appeared at a random position along the scale on every trial.

On half of the trials, the confidence scale appeared one second after response submission (timing 1 condition), and in the remaining half, the confidence scale appeared 2.3 seconds after onset of the squares (timing 2 condition). These two timing conditions were used because the higher effort responses took longer to enact. As such, if the delay between response submission (i.e. the threshold being reached) and confidence scale onset were kept consistent across all trials, this would mean that there was a longer lag between stimulus onset and confidence report for higher, relative to lower effort trials. However, if the delay between stimulus onset and confidence scale onset were kept consistent across all trials, higher effort responses would leave a shorter gap between response submission and confidence report. Due to these conflicting confounds, both timing conditions were implemented, randomised across blocks, and explicitly modelled in the analyses.

2.3.3 Confidence ratings

Following Fleming and colleagues (2018), the confidence scale (Fig. 1) incorporated vertical lines and labels to mark 20% ("Probably wrong"), 40% ("Maybe wrong"), 60% ("Maybe correct"), and 80% ("Probably correct") confidence. Furthermore, participants were told that an additional reward of up to $5 could be earned based on task performance and the accuracy of their confidence ratings as calculated via a quadratic scoring rule (points = $100*[1 – (\text{correct}_i – \text{confidence}_i)]^2$). This was done to incentivise accurate responses and honest confidence ratings. Prior to the experiment, participants were familiarised with the confidence scale and scoring rule.

Note that we use the term 'confidence' to refer to confidence in having responded correctly, which is distinct from 'certainty' in the outcome of the response (i.e. the absolute distance from the centre of the confidence scale).

2.3.4 Procedure

Participants completed the initial calibration of the hand dynamometers with the experimenter present and then completed the staircase procedure and main experiment alone. Once participants completed the experiment, they were debriefed and received the monetary compensation.

2.4 Statistical Analysis

Data and analysis code will be made publicly available at https://osf.io/cg74z/ at the time of publication. Analyses were conducted using linear and generalised linear mixed-effects models. These were performed in *R* (version 3.5) with the *lme4* package (version 1.1; Bates, Mächler, Bolker, & Walker, 2015) and the *glmmTMB* package (version 1.0.1; Brooks et al., 2017). Continuous variables were centred and scaled, and missed responses were excluded.

2.4.1 Control analyses

Initial control analyses were conducted to ensure that the stimulus difficulty manipulation produced effects in the expected direction, and to examine whether the accuracy and timing of decisions differed across the three effort conditions. Although these effects were accounted for in the models and can technically be inferred from the mixed-effects model parameters, these analyses were reported for completeness.

To ensure that individuals were more accurate in easy as compared to hard trials, a likelihood ratio test was conducted between a generalised linear mixed model (GLMM) predicting accuracy from stimulus difficulty, and an intercept only null model. To ensure that participants responded faster on easy as compared to hard trials, a likelihood ratio test was conducted between a GLMM (Gamma family) with an identity link function (as recommended by Lo & Andrews, 2015), predicting response time from stimulus difficulty, and an intercept only null model. Furthermore, to ensure participants responded faster on correct as compared to error trials, a likelihood ratio test was conducted between a GLMM (Gamma family) with an identity link function predicting response time from accuracy, and an intercept only null model. Finally, likelihood ratio tests were also conducted to determine whether initial decision accuracy and response times differed significantly across the three effort conditions.

2.4.2 Mixed-effects models

To determine whether invested effort influenced confidence ratings, a linear mixed-effects regression model was used to predict decision confidence based on effort and a number of control variables. Mixed-effects models were used as the data had a multi-level structure; observations (i.e. confidence ratings) were nested within participants. As such, decision confidence and the predictors' effects on decision confidence would be more strongly correlated *within* participants than *between* participants (Fleming, Weil, Nagy, Dolan, & Rees, 2010). Mixed-effects models can account for the inherent dependence in our data due to individual-level differences, and better account for this variation. Therefore, participant ID was additionally included as a random intercept to allow average confidence ratings to differ for each participant. To account for variability in the effects of effort, accuracy, and stimulus difficulty across participants, random slopes were also included for these three variables as well as the interaction between accuracy and stimulus difficulty (see below for details on this interaction).

Initial response time (i.e. the time at which participants first began to squeeze) was defined as the time at which squeeze force first exceeded 10% MVC in each trial, and was included as a covariate in the model. Accuracy, stimulus difficulty, and timing condition (i.e. whether the onset of the confidence scale was time-locked to either stimulus onset or response offset) were also included as covariates in the model. These control variables were included either because they are known to be associated with decision confidence (response time, accuracy, stimulus difficulty; see Pleskac & Busemeyer, 2011) or to control for the effect of our manipulation of confidence scale onset (timing condition).

An interaction term between stimulus difficulty (i.e. evidence strength) and accuracy was also included in all models, as the distribution of confidence ratings reflected an established pattern in the metacognition literature (plotted in Appendix A) whereby increased evidence strength leads to increased confidence in correct responses, but to decreased confidence in error responses (i.e. the folded-X effect; Kepecs & Mainen, 2012). Interactions between effort condition and accuracy, effort condition and timing condition, and effort condition and stimulus difficulty were not included in the final model, as likelihood ratio tests indicated that models including these interactions did not fit the data significantly better than null models which did not include the interaction of interest.

A likelihood ratio test was conducted to compare the fit of a full model with effort as a predictor (model 1) to a null model which did not include effort as a predictor (model 2). Regression model structures are as follows:

(1) confidence ~ effort + accuracy*difficulty + timing + initialRT + (1 + effort + accuracy * difficulty | participant)

(2) confidence ~ accuracy*difficulty + timing + initialRT + (1 + effort + accuracy * difficulty | participant)

A post-hoc Tukey test was conducted using the emmeans package in R, to formally examine differences in confidence ratings between the three effort levels.

2.4.3 Time to threshold force analysis

In addition to the main analysis, we conducted an exploratory analysis examining the relationship between the time to threshold force (i.e. the time at which the force threshold was crossed relative to the initial response time) and decision confidence. For this analysis a likelihood ratio test was conducted between a full model containing the variable of interest (i.e. time to threshold force) and a null model (i.e. model 2 above). Both of these models included an additional random intercept for effort level. An additional exploratory analysis investigating the relationship between maximum recorded force and decision confidence is also reported in Appendix B.

3. Results

3.1 Control Analyses

Control analyses were conducted to ensure that the stimulus difficulty manipulation produced the expected effects on behaviour, and to examine whether the accuracy and timing of decisions differed significantly across the effort conditions. As expected, the proportion of correct responses was higher on easy (M = 86.8%, SD = 0.06) compared to hard trials (M = 75.7%, SD = 0.08; likelihood ratio test, $\chi^2(1) = 403.82$, $p < .001$), and participants' response times were faster on easy (M = 713 ms, SD = 169 ms), compared to hard trials (M = 740 ms, SD = 182 ms; likelihood ratio test, $\chi^2(2) = 73.29$, $p < .001$). Participants also responded faster when making correct responses (M = 715 ms, SD = 170 ms) compared to errors (M = 782 ms, SD = 207 ms; likelihood ratio test, $\chi^2(1) = 225.98$, $p < .001$). There was no evidence of a significant difference in participants' accuracy rates between the three effort conditions (M_{low} = 81.1%, M_{med} = 81%, M_{high} = 81.8%; likelihood ratio test, $\chi^2(2) = 1.74$, $p = .419$). There was evidence of a significant difference in response times ($Mean_{low}$ = 732 ms, $Mean_{med}$ = 731 ms, $Mean_{high}$ = 717 ms; likelihood ratio test, $\chi^2(2) = 19.68$, $p < .001$). The most likely reason for this is that, because the effort threshold took longest to reach in the high effort condition, slow responses were more likely to be missed – leading to a slight artificial increase in response times. Given this, it was important to include response time as a covariate in the main analysis model (see below) to rule out that response times were driving the observed effects. We also conducted an additional analysis (reported in Appendix C) where we sub-sampled the dataset and matched response times and miss-rates across conditions to confirm that condition-wise differences in response time were not the cause of condition-wise differences in confidence.

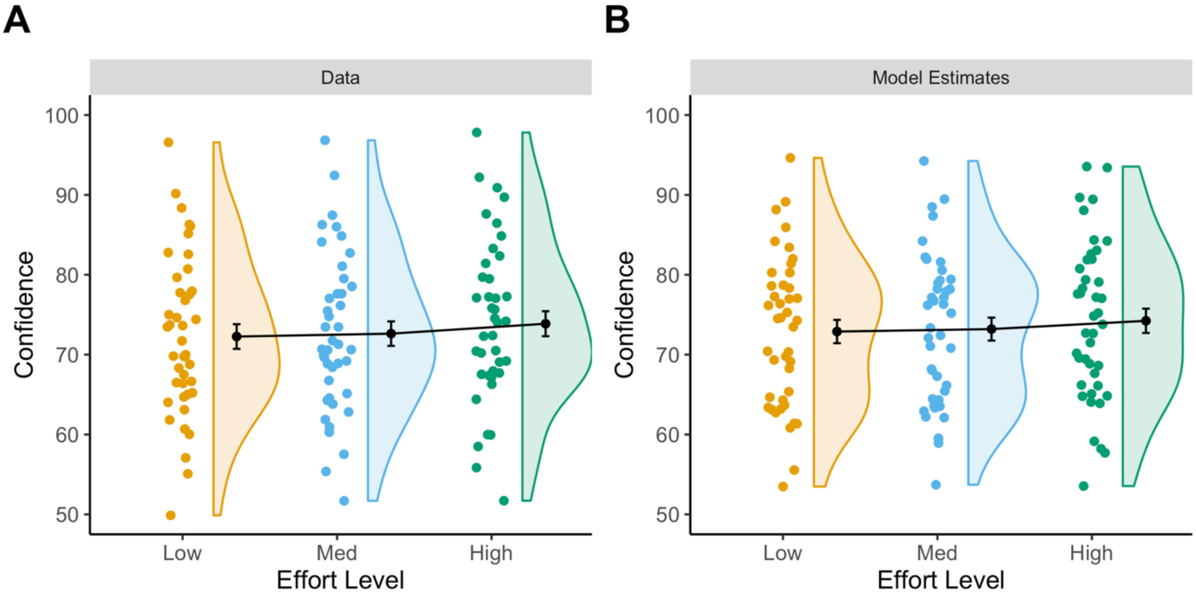

Fig. 2. Mean confidence ratings across the three effort conditions. A) Participants' mean confidence ratings across all trials for each effort level. Each coloured point represents the mean confidence rating from an individual participant. B) Estimated mean confidence ratings from the mixed effects model across the three effort levels. In all plots mean confidence ratings (black dots) are connected by black lines. Error bars indicate the standard error of the mean. For reference, a confidence rating of 0 represents a confidence level of 'Certainly Wrong', whilst a rating of 100 represents a confidence level of 'Certainly Correct'. The raincloud plots were made using code from (Allen, Poggiali, Whitaker, Marshall, & Kievit, 2019). Note, the apparent bimodal distribution of the predicted confidence ratings from the model in Fig. 2B (particularly for the low and medium effort conditions) is simply due to variability in the model predictions. If the seed of the random number generator in R is changed, then this apparent bimodality disappears. For simplicity we have left the RNG seed equal to 1.

3.2 Confidence Ratings

To determine whether effort was a significant predictor of decision confidence, a likelihood ratio test was used to compare a full model including the main predictor of interest (i.e. effort condition) to a null model which did not include this predictor. The logic of the test is that if the model with effort is a better fit to the data, then effort is a significant predictor of decision confidence. The full model fit the data significantly better than the null model (likelihood ratio test, $\chi^2(2) = 10.47$, $p = .005$). In the full model (summarised in Table 1), high effort was a predictor of increased confidence ($p = .002$), however medium effort was not (*p* = .579). The distribution of confidence ratings (Fig. 2A) reflects these results, as high effort showed a larger positive effect on confidence ratings compared to both medium and low effort. A post-hoc Tukey test showed that confidence ratings were significantly higher in high compared to low effort trials (*p* = .006) and high compared to medium effort trials (*p* = .029), but did not significantly differ between low and medium effort trials (*p* = .844).

As expected, confidence was increased for fast responses (p < .001 see Table 1). There was also a significant interaction between accuracy and stimulus difficulty (i.e. evidence strength). This reflects the ‘folded-X’ effect (Kepecs & Mainen, 2012), whereby increases in evidence strength are associated with increased confidence in correct decisions but decreased confidence in incorrect decisions. This has been widely reported in previous studies and is a feature predicted by many models of metacognition (e.g., Fleming & Daw, 2017).

Though linear mixed-effects methods are commonly used in the confidence literature for multi-level data structures (Fleming et al., 2018; Gajdos et al., 2019), it has been suggested that a generalised linear model that assumes a beta distribution is more appropriate for modelling doubly bounded continuous data (Verkuilen & Smithson, 2012). Our results were consistent when using generalised linear (beta) mixed-effects model analyses (Appendix D).

Table 1. Estimates from the Full Linear Mixed-Effects Model

Fixed effects	Estimate	*CI*	*p*
(Intercept)	75.73	71.51 – 79.95	**<.001**
Medium Effort	0.17	-0.42 – 0.75	.579
High Effort	0.91	0.33 – 1.49	**<.002**
Timing (Stimulus-locked)^a	0.52	0.07 – 0.97	**.025**

Accuracy (Correct)	13.79	9.74 – 17.84	**<.001**
Difficulty (Easy)	-6.52	-8.45 – -4.59	**<.001**
RT	-0.02	-0.02 – -0.02	**<.001**
Accuracy*Difficulty	9.47	6.99 – 11.96	**<.001**

[a] Timing (Stimulus-locked) refers to the timing 2 condition whereby confidence scale onset occurred 2.3 seconds following the onset of the squares.

3.3 Time to Threshold Analysis

Having determined that decision confidence was positively associated with the level of effort required to report a decision, we then conducted an additional exploratory analysis examining whether there was an association between decision confidence and the time it took to reach the force threshold relative to the initial response time (‘time to threshold force’). The full model fit the data significantly better than the null model, indicating a significant negative effect of time to threshold force on decision confidence (Fig. 3; likelihood ratio test: $\chi^2(1) = 113.11, p < .001$). Further analyses looking at the relationship between maximum recorded force and decision confidence on each trial are reported in Appendix B.

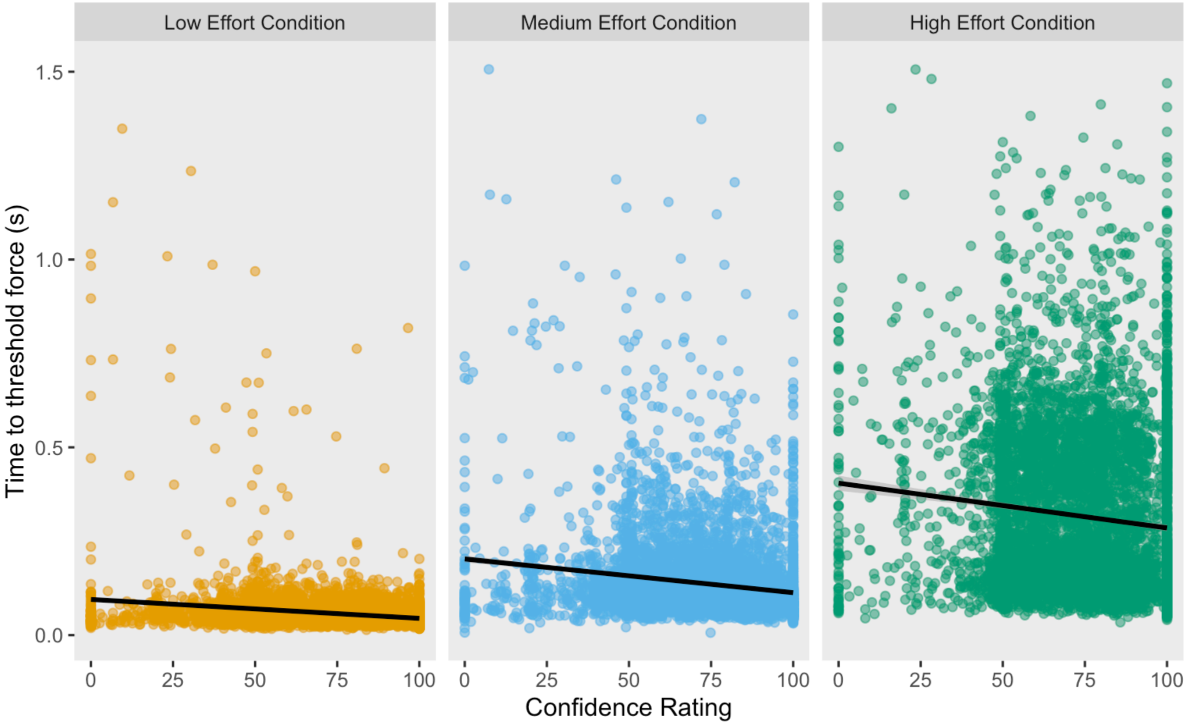

Fig. 3. Associations between decision confidence and time to threshold force, within each effort condition. For illustrative purposes the black lines were fit using a simple regression model which predicted confidence from time to threshold force.

4. Discussion

We investigated whether the 'motoric sunk cost' of a decision (i.e. the amount of effort one has invested into reporting a decision) affects decision confidence (i.e. how confident one feels in having responded correctly). In support of our hypothesis, we found that increases in the amount of effort required to report a choice were associated with increased confidence. This suggests that humans are sensitive to a 'motoric sunk cost effect', whereby decisions which one has invested more effort into reporting are judged as more likely to be correct. Additional, exploratory single-trial analyses revealed that decision confidence was also negatively associated with the time it took to reach the response force threshold, relative to the initial response time ('time to threshold force'). In other words, more vigorous responses were associated with higher confidence. Taken together these findings suggest that various sources of action-related information feed into judgements of decision confidence, consistent with contemporary models of metacognition (Fleming & Daw, 2017).

This study sits within a growing body of literature which shows associations between action-related information and metacognitive judgements (Faivre et al., 2018, 2020; Fleming et al., 2015; Palser et al., 2018; Pereira et al., 2020; Siedlecka, Hobot, et al., 2019; Siedlecka et al., 2020; Siedlecka, Skóra, et al., 2019; Wokke et al., 2020). By directly manipulating the amount of effort required to report a decision we have shown that confidence depends, in part, on fine-grained representations of one's own actions. This supports Fleming and Daw's (Fleming & Daw, 2017) model of metacognitive judgements and is consistent with the notion that multiple sources of sensory and motoric information can be exploited to refine confidence estimates.

Fleming and Daw (2017) hypothesised that actions can inform decision confidence. However, they did not specify the exact effect that variations in decision-related motor costs would have on confidence. Our results help clarify this by showing that expended effort influences decision confidence (i.e. it increases confidence in a decision being correct). One interpretation of this finding is that expended effort is used as a heuristic (i.e. a proxy for decision accuracy) that informs confidence judgements. Investing more effort into a decision might be interpreted post-hoc as a signal that the decision is likely to be correct. In a similar vein, it has been shown that faster response times predict increased confidence in a decision, as quick responses potentially indicate that a decision is more likely to be correct (Kiani, Corthell, & Shadlen, 2014). Indeed, this effect was also present in our data. Taken together, both the effects of effort and response speed on subsequent confidence judgements reinforce the notion that various sources of action-information can act as additional cues regarding decision accuracy, particularly when sensory and decision-related information is limited or ambiguous. Fast response times may act as a signal that a decision was easily made (so likely to be correct), whilst effort invested into reporting a decision may act as a 'sunk cost' which also inflates decision confidence.

Notably, a related body of literature has shown that change-of-mind decisions—rapid decision reversals (Resulaj, Kiani, Wolpert, & Shadlen, 2009)—are sensitive to anticipated motor costs. In particular, it has been shown that individuals are less likely to change their minds when it is more effortful to do so (Burk, Ingram, Franklin, Shadlen, & Wolpert, 2014; Moher & Song, 2014). In these studies, participants moved their hand towards a leftward or rightward target box to indicate their choices during a random dot motion task (Burk et al., 2014; Moher & Song, 2014). On trials where the distance between the two targets was larger and revising a decision mid-movement would incur a larger motoric and temporal cost, the frequency of changes-of-mind was reduced. As confidence has been used as a proxy for change-of-mind decisions (i.e. high confidence is associated with a lower likelihood of changing one's mind, and vice versa; Fleming, 2016; Folke et al., 2017), this could suggest that, when anticipating more costly changes-of-mind, confidence in an initial choice was increased. Crucially, our experiment suggests that, in addition to *anticipated* effort, *expended* effort can also increase confidence. In essence, this serves as a demonstration of a 'motoric sunk cost effect' in humans, similar to a novel temporal sunk cost effect which has recently been reported (Sweis et al., 2018). While anticipated effort might bias confidence already during the decision process, potentially to restrict energy expenditure linked to costly changes-of-mind, expended effort might be linked to a different mechanism and serve as post-hoc evidence, in addition to the sensory information, which feeds into the metacognition evaluation process. Critically, whilst such an effect acts as a bias in the current experimental context, in more real world scenarios it may often be useful, and even rational (c.f. Fleming & Daw, 2017), for decision-makers to take this information into account when making metacognitive judgements.

The observation that confidence ratings did not significantly differ between the low and medium effort conditions raises the question of whether expended effort has a graded effect on decision confidence. One possible reason why confidence ratings did not significantly differ between the low and medium effort conditions is that participants may not have experienced a substantially larger effort cost in the medium effort condition compared to the low effort condition. Effort discounting studies have shown that incremental increases in effort expenditure have a greater impact on perceived costs when individuals are closer to their maximum level of exertion (Chong et al., 2018; Hartmann, Hager, Tobler, & Kaiser, 2013; Stevens & Mack, 1959). Given that confidence was significantly increased on high compared to medium and low effort trials, and that confidence on medium effort trials was quantitatively higher than confidence on low effort trials, we conclude that expended effort does affect decision confidence.

To better differentiate between effort conditions, future studies could incorporate additional effort levels (e.g., six effort levels at 5% increments) and utilise sustained contractions (see Chong et al., 2018 for an example), rather than brief, ballistic contractions. This might allow differences between effort increments—even at lower levels—to become more salient, and tease out graded effects to determine whether the pattern in the effort discounting literature (e.g., a parabolic/concave relationship between actual and subjective effort costs) extends to the effect of physical effort on confidence as well.

4.1 Limitations

Our results should be interpreted with the following limitations in mind. Since participants were given a visual indication as to how much effort they were exerting on each trial, it is not possible to determine whether the effect of expended effort was driven by proprioceptive feedback, the visual cue, or a combination of both. It is possible that simply believing that they had expended more effort after seeing a visual cue was enough to affect participants' decision confidence (either unconsciously or as a form of demand characteristic). However, if this were the case, it is unclear why there was no statistically significant difference in confidence between the low and medium effort conditions, but there was a statistically significant difference in confidence between the medium and high conditions. Given that the position of the threshold line increased by equal increments between the low and medium, and the medium and high conditions, if participants were influenced by the visual cue, we would expect their confidence ratings to also change by the same amount across effort levels. Instead, participants displayed a greater increase in confidence between the medium and high effort conditions, compared to the low and medium effort conditions, consistent with a parabolic/concave relationship between actual and subjective effort costs (Chong et al., 2018; Hartmann et al., 2013; Stevens & Mack, 1959).

Consistent with the overarching view that actions inform decision confidence (Fleming & Daw, 2017), we also found that measures of squeeze force trajectories (i.e. time to threshold force and maximum recorded force) were related to decision confidence. This suggests that even when the visual cue is controlled for, fine-grained response information is still reliably associated with decision confidence. Critically however, since unlike the effort condition manipulation, time to threshold force and maximum force were not directly manipulated within each effort condition, the implications of these associations are ultimately unclear. It is possible that slightly more vigorous responses led to greater confidence, or that more confident decisions led to slightly more vigorous responding.

Whilst we cannot unequivocally conclude that the effect was driven by proprioceptive feedback alone, this interpretation seems most plausible. Whether this effect remains when motor costs are manipulated without providing exogenous cues should be investigated in future studies. However, such manipulations are not trivial. Simply removing the visual cue, or replacing it with an auditory cue, will introduce response uncertainty (i.e. uncertainty about how close one is to locking in a response), which will lead to different response dynamics (i.e. repeated bursts of squeezing to make up for missing the force threshold) that, in turn, may themselves influence confidence judgements.

One final potential limitation of this study is that participants were more likely to exceed the response time deadline in high effort trials, as it took longer to reach the required squeeze force threshold. This gives rise to a potential confound, as only relatively quick initial responses would have been recorded (i.e. if participants were slow to start squeezing, then their response would not be recorded). Since response time is known to negatively correlate with confidence (Kiani et al., 2014), a potential concern is that condition-wise differences in response times may have given rise to the condition-wise differences in confidence. However, in the mixed-effects models, when effects of response time were controlled for, motoric effort nevertheless had a significant effect. Moreover, the effect of effort on confidence still remained after matching response times and miss-rates across conditions (see Appendix C). Finally, response time was also controlled for in the within-condition analyses, and time to threshold force and maximum recorded force were nevertheless consistently associated with decision confidence. Given this, we conclude that motor information can influence decision confidence, independently of response times.

4.2 Summary

Here, we have shown that confidence in a perceptual decision depends, in part, on the ‘motoric sunk cost’ incurred from reporting the decision. In other words, we have shown that individuals tend to report higher confidence in decisions for which they had invested greater effort into reporting. This demonstration of a ‘motoric sunk cost effect’ supports contemporary models of metacognition in which action information feeds into confidence estimates. Our findings lend further support to the notion that fine-grained representations of action-related information are indeed used for computations of perceptual confidence judgements.

5. Acknowledgements

This work was supported by an Australian Research Council (ARC) Discovery Project Grant [DP160103353] to SB and RH and an Australian Government Research Training Program (RTP) Scholarship to WT.

6. Declaration of interest

We have no competing interests.

7. Supplementary material

Data and analysis code for this paper are available at https://osf.io/cg74z/.

References

Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., & Kievit, R. A. (2019). Raincloud plots: A multi-platform tool for robust data visualization. *Wellcome Open Research*, *4*, 1–40. https://doi.org/10.12688/wellcomeopenres.15191.1

Arkes, H. R., & Blumer, C. (1985). The psychology of sunk cost. *Organizational Behavior and Human Decision Processes*, *35*(1), 124–140. https://doi.org/10.1016/0749-5978(85)90049-4

Bang, D., & Fleming, S. M. (2018). Distinct encoding of decision confidence in human medial prefrontal cortex. *Proceedings of the National Academy of Sciences*, *115*(23). https://doi.org/10.1073/pnas.1800795115

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

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Brooks, M. ., Kristensen, K., Van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., … Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for Zero-inflated Generalized Linear Mixed Modeling. *The R Journal*, *9*((2)), 378–400. Retrieved from http://orbit.dtu.dk/files/154739064/Publishers_version.pdf

Burk, D., Ingram, J. N., Franklin, D. W., Shadlen, M. N., & Wolpert, D. M. (2014). Motor effort alters changes of mind in sensorimotor decision making. *PLoS ONE*, *9*(3), e92681. https://doi.org/10.1371/journal.pone.0092681

Chong, T. T. J., Apps, M. A. J., Giehl, K., Hall, S., Clifton, C. H., & Husain, M. (2018). Computational modelling reveals distinct patterns of cognitive and physical motivation in elite athletes. *Scientific Reports*, *8*(1), 1–11. https://doi.org/10.1038/s41598-018-

30220-3

Faivre, N., Filevich, E., Solovey, G., Kühn, S., & Blanke, O. (2018). Behavioral, modeling, and electrophysiological evidence for supramodality in human metacognition. *Journal of Neuroscience*, *38*(2), 263–277. https://doi.org/10.1523/JNEUROSCI.0322-17.2017

Faivre, N., Vuillaume, L., Bernasconi, F., Salomon, R., Blanke, O., & Cleeremans, A. (2020). Sensorimotor conflicts alter metacognitive and action monitoring. *Cortex*, *124*, 224–234. https://doi.org/10.1016/j.cortex.2019.12.001

Fleming, S. M. (2016). Changing our minds about changes of mind. *ELife*, *5*, e14790.

Fleming, S. M., & Daw, N. D. (2017). Self-Evaluation of Decision-Making: A General Bayesian Framework for Metacognitive Computation. *Psychological Research*, *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

Fleming, S. M., Van Der Putten, E. J., & Daw, N. D. (2018). Neural mediators of changes of mind about perceptual decisions. *Nature Neuroscience*, *21*(4), 617–624. https://doi.org/10.1038/s41593-018-0104-6

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

Folke, T., Jacobsen, C., Fleming, S. M., & De Martino, B. (2017). Explicit representation of confidence informs future value-based decisions. *Nature Human Behaviour*, *1*(1), 17–19. https://doi.org/10.1038/s41562-016-0002

Gajdos, T., Fleming, S. M., Saez Garcia, M., Weindel, G., & Davranche, K. (2019). Revealing subthreshold motor contributions to perceptual confidence. *Neuroscience of*

Consciousness, *2019*(1), 1–8. https://doi.org/10.1093/nc/niz001

Hagura, N., Haggard, P., & Diedrichsen, J. (2017). Perceptual decisions are biased by the cost to act. *ELife*, *6*, 1–20. https://doi.org/10.7554/eLife.18422

Hartmann, M. N., Hager, O. M., Tobler, P. N., & Kaiser, S. (2013). Parabolic discounting of monetary rewards by physical effort. *Behavioural Processes*, *100*, 192–196. https://doi.org/10.1016/j.beproc.2013.09.014

Kepecs, A., & Mainen, Z. F. (2012). A computational framework for the study of confidence in humans and animals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1594), 1322–1337. https://doi.org/10.1098/rstb.2012.0037

Kepecs, A., Uchida, N., Zariwala, H. A., & Mainen, Z. F. (2008). Neural correlates, computation and behavioural impact of decision confidence. *Nature*, *455*(7210), 227–231. https://doi.org/10.1038/nature07200

Kiani, R., Corthell, L., & Shadlen, M. N. (2014). Choice certainty is informed by both evidence and decision time. *Neuron*. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4271191/pdf/nihms647759.pdf

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3. *Perception*, *36*(14), 1–16. https://doi.org/10.1068/v070821

Lo, S., & Andrews, S. (2015). To transform or not to transform: using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, *6*(August), 1–16. https://doi.org/10.3389/fpsyg.2015.01171

Middlebrooks, P. G., & Sommer, M. A. (2011). Metacognition in monkeys during an oculomotor task. *Journal of Experimental Psychology: Learning Memory and Cognition*, *37*(2), 325–337. https://doi.org/10.1037/a0021611

Moher, J., & Song, J.-H. (2014). Perceptual decision processes flexibly adapt to avoid change-of-mind motor costs. *Journal of Vision*, *14*(8), 1–13.

https://doi.org/10.1167/14.8.1

Palser, E. R., Fotopoulou, A., & Kilner, J. M. (2018). Altering movement parameters disrupts metacognitive accuracy. *Consciousness and Cognition*, *57*(May 2017), 33–40. https://doi.org/10.1016/j.concog.2017.11.005

Pereira, M., Faivre, N., Iturrate, I., Wirthlin, M., Serafini, L., Martin, S., … Millán, J. del R. (2020). Disentangling the origins of confidence in speeded perceptual judgments through multimodal imaging. *Proceedings of the National Academy of Sciences*, *117*(15), 201918335. https://doi.org/10.1073/pnas.1918335117

Pleskac, T. J., & Busemeyer, J. R. (2011). “Two-stage dynamic signal detection: A theory of choice, decision time, and confidence”: Erratum. *Psychological Review*, *118*(1), 56–56. https://doi.org/10.1037/a0022399

Resulaj, A., Kiani, R., Wolpert, D. M., & Shadlen, M. N. (2009). Changes of mind in decision-making. *Nature*, *461*(7261), 263–266. https://doi.org/10.1038/nature08275

Siedlecka, M., Hobot, J., Skóra, Z., Paulewicz, B., Timmermans, B., & Wierzchoń, M. (2019). Motor response influences perceptual awareness judgements. *Consciousness and Cognition*, *75*. https://doi.org/10.1016/j.concog.2019.102804

Siedlecka, M., Paulewicz, B., & Koculak, M. (2020). Task-related motor response inflates confidence. *BioRxiv*.

Siedlecka, M., Skóra, Z., Paulewicz, B., Fijałkowska, S., Timmermans, B., & Wierzchoń, 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

Stevens, J. C., & Mack, J. D. (1959). Scales of apparent force. *Journal of Experimental Psychology*, *58*(5), 405–413. https://doi.org/10.1037/h0046906

Sweis, B., Abram, S. V., Schmidt, B. J., Seeland, K. D., MacDonald III, A. W., Thomas, M.

J., & Redish, D. A. (2018). Sensitivity to “sunk costs” in mice, rats, and humans. *Science*, *361*(6398). https://doi.org/doi:10.1126/science.aar8644.

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*, *26*(23), 3157–3168. https://doi.org/10.1016/j.cub.2016.10.021

Verkuilen, J., & Smithson, M. (2012). Mixed and mixture regression models for continuous bounded responses using the beta distribution. *Journal of Educational and Behavioral Statistics*, *37*(1), 82–113. https://doi.org/10.3102/1076998610396895

Wokke, M. E., Achoui, D., & Cleeremans, A. (2020). Action information contributes to metacognitive decision-making. *Scientific Reports*, *10*(1), 1–15. https://doi.org/10.1038/s41598-020-60382-y

Appendix A

The Folded-X Interaction Effect

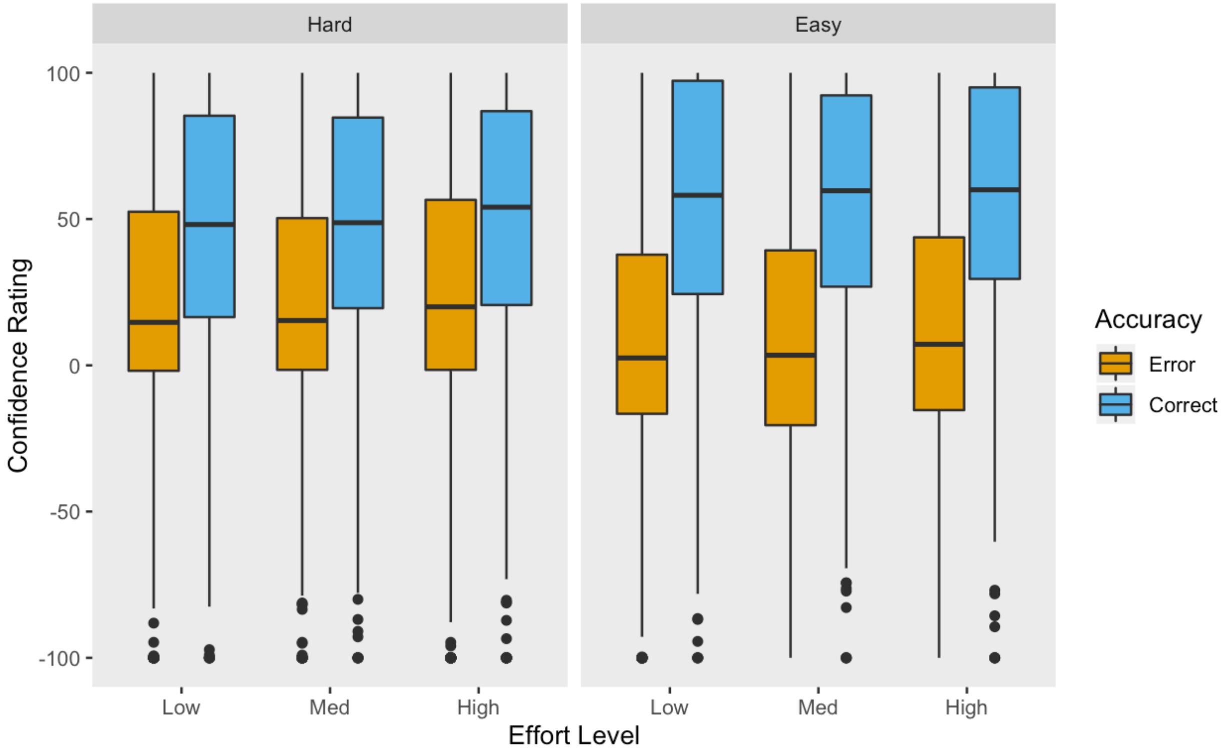

Fig. A.1. Distributions of confidence ratings for correct and error trials across the three effort levels for hard and easy difficulty conditions: The results show the 'folded-X' interaction pattern of confidence judgements. That is, as compared to hard trials (low evidence strength), reported confidence in easy trials (high evidence strength) tended to be higher for correct trials, but lower for error trials. This provided a rationale for including the accuracy*difficulty interaction in the model as a control variable.

Appendix B

Maximum Recorded Force and Decision Confidence

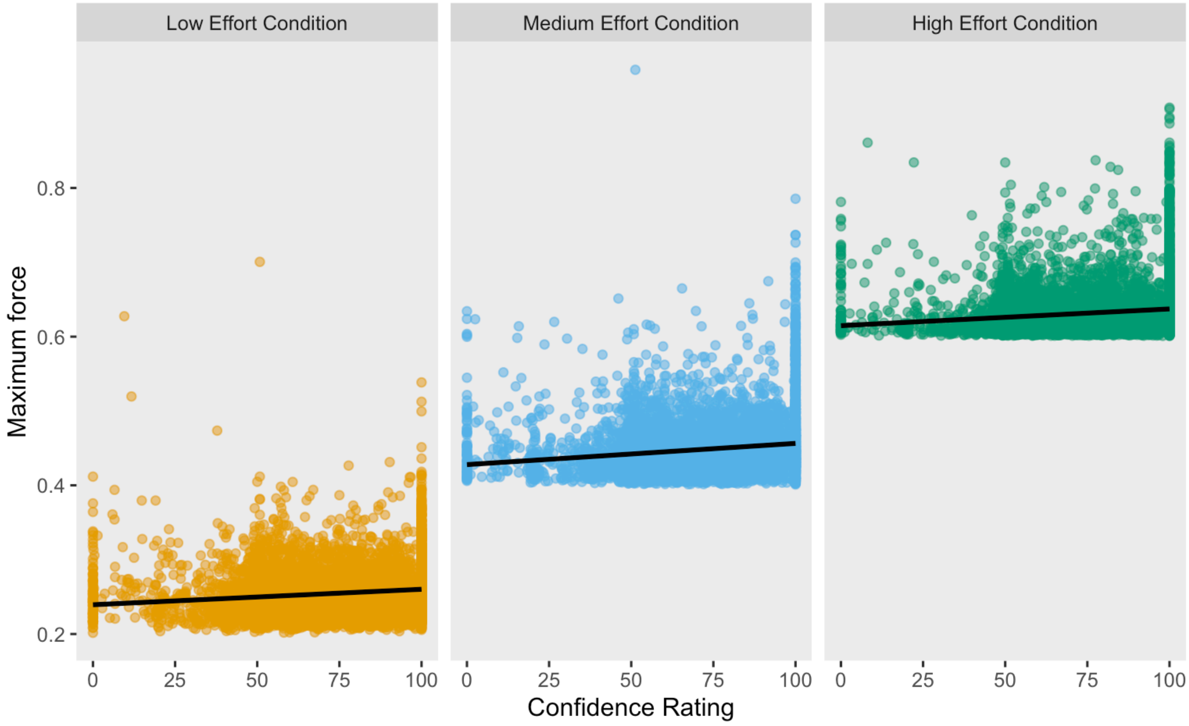

Fig. B.1. Associations between maximum recorded force and decision confidence, within each effort condition. For illustrative purposes the black lines were fit using a simple regression model which predicted confidence from maximum recorded force.

We examined the relationship between the maximum recorded force on each trial and decision confidence. For this analysis, it is important to note that the dynamometers were programmed to stop recording once the initial force threshold was crossed. However, the testing computer only received a new sample (a 15 ms sample of data recorded at 1,000 Hz) from the dynamometers every 15 ms. As a result, the maximum recorded force was different on each trial, even though the threshold crossing ultimately triggered the dynamometers to stop recording. This allowed us to examine whether maximum recorded force was meaningfully related to decision confidence. Nevertheless, given that this is an imperfect measure of the maximum force applied to the dynamometers in each trial (i.e. it is very likely that on some trials participants continued to squeeze after the dynamometers stopped recording), we have chosen to report these results here, rather than in the main text.

For this analysis, a likelihood ratio test was conducted between a full model, containing maximum recorded force as a predictor, and a null model which did not contain maximum recorded force but was otherwise identical (see R code at https://osf.io/cg74z/ for full details). These analyses revealed that decision confidence was positively associated with the maximum recorded force (Fig. B.1; likelihood ratio test: $\chi^2(1) = 9.93$, $p = .002$).

Appendix C

Matching response times and miss rates across the effort levels.

Considering the percentage of missed trials (i.e. trials in which a response was not recorded), it is apparent that participants were slightly more likely to miss responses on high effort trials (5.16% of trials) compared to the low (1.89%) and medium (1.93%) effort trials. As a result, response times tended to be slightly faster on high effort trials compared to low and medium effort trials. This is because it took longer to reach the response threshold on high effort trials, so slow responses were more likely to be missed. As can be seen in Fig C1 (below) this leads to a slight speeding of high effort responses in the 0.9 quantile of the response time distribution. Analysing RTs across the effort levels we find that there was a small but significant negative association between RT and effort level (likelihood ratio test, $\chi^2(2) = 19.68$, $p < .001$). Given this, it was important to include response time as a covariate in the main analysis.

To ensure that the effects we observed were not due to differences in response time, we also conducted an additional analysis on a subset of the original data. We first removed a percentage of the slowest responses in the low and medium effort conditions, equal to the difference in the percentage of missed trials between the low and medium conditions and the high effort condition. Specifically, we removed the slowest 3.27% of trials in the low condition, and the slowest 3.23% of trials in the medium condition. This left 6353 trials in the low condition, 6354 trials in the medium condition, and 6354 trials in the high condition.

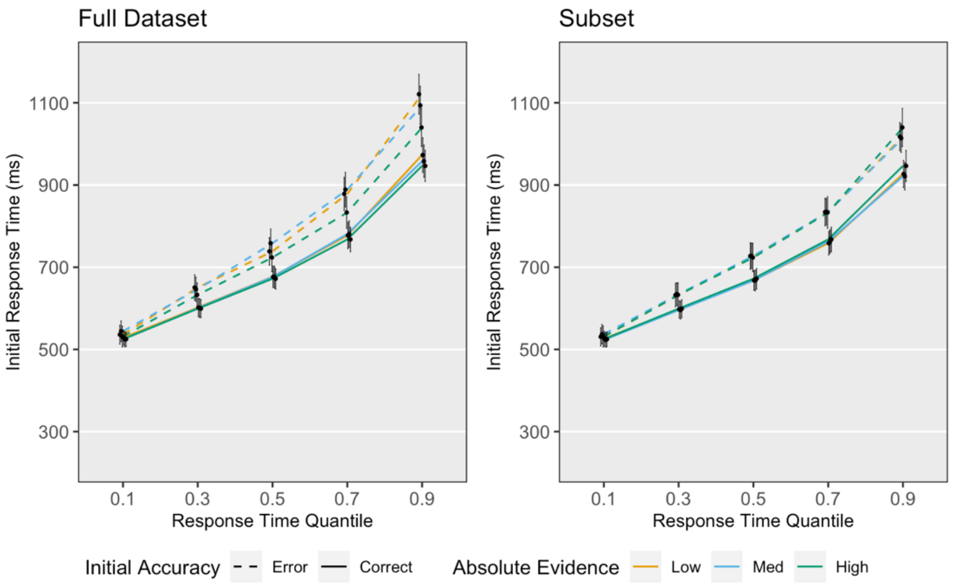

Fig. C.1. Response time quantiles calculated from the full dataset and the sub-setted data. These plots were created by vincentizing correct and error RT quantiles across participants, within the three effort levels. After sub-setting, response times were more closely matched across the effort conditions and the negative trend between RT and effort is removed.

As intended, after sub-sampling the data, response times were no longer significantly different between the effort conditions (likelihood ratio test, $\chi^2(2) = 0.81$, $p = 0.67$; see Fig C.1). Moreover, accuracy was not significantly different between the three effort conditions (likelihood ratio test, $\chi^2(2) = 0.30$, $p = 0.86$). Critically however, there was still a significant effect of effort on decision confidence, with participants being more confident in high effort responses (likelihood ratio test, $\chi^2(2) = 12.29$, $p = 0.002$). This indicates that the effect of effort on confidence was not simply driven by differences in response time or the proportion of missed responses across conditions.

Because of the differences in miss-rates between the effort levels, one additional concern might be that participants may have gradually learned to associate high effort trials with high decision confidence. However, analysing only responses in the first ~10% of experimental trials (i.e. the first 50 trials) of the sub-setted dataset for each participant, we still observed a significant effect of effort on decision confidence (likelihood ratio test, $\chi^2(2)$

= 7.37, p = 0.025). This indicates that the current effects were also not driven by a gradual, learned association between high effort and high confidence.

Appendix D

Generalised Linear Mixed Effects Models

Though linear mixed-effects methods are commonly used in the confidence literature for multi-level data structures, a potential problem with conventional linear models is that they do not appropriately address the non-normally distributed nature of confidence rating data. It has been suggested that a generalised linear model with a beta distribution can overcome these issues, and that beta distributions are more appropriate for modelling doubly bounded continuous data (Verkuilen & Smithson, 2012). To ensure that the effects were robust across these approaches, additional analyses were conducted with generalised linear models using a beta distribution. Note, the model did not converge with all random slopes included, so we removed the random slope for effort level but left in the slope for the interaction between accuracy and difficulty (when just a random slope for effort was included the model also failed to converge).

The likelihood ratio test demonstrated that effort was a significant predictor of confidence, $\chi^2(2) = 8.09$, $p = .018$. Hence, the beta model also supported the main hypothesis that effort is a significant predictor of increased confidence. Similar to the linear mixed-effects models described in the main text, the model with effort (Table D.1) showed that high effort was significant ($p = .014$) but medium effort was not ($p = .998$). Confidence ratings were also higher for correct, relative to error trials ($p < .001$) and faster RTs ($p < .001$). Finally, when analysing just the sub-set of data (see Appendix C), effort was still a significant predictor of confidence, $\chi^2(2) = 10.40$, $p = .005$.

Table D.1. Estimates from the Full Generalised Linear (Beta Distribution) Mixed-Effects Model

Fixed effects	Estimate	*CI*	*p*
(Intercept)	1.54	1.17 – 2.03	**0.002**
Medium Effort	1.00	0.97 – 1.04	0.998
High Effort	1.05	1.01 – 1.08	**0.014**
Timing (Stimulus-locked)	1.06	1.03 – 1.09	**<0.001**
RT	0.74	0.72 – 0.75	**<0.001**
Accuracy (Correct)	2.24	1.64 – 3.06	**<0.001**
Difficulty (Easy)	0.67	0.58 – 0.77	**<0.001**
Accuracy*Difficulty	1.74	1.46 – 2.08	**<0.001**