

# Response duration tracks confidence

Hanbei Zhou ([hanbeizhou@g.harvard.edu](mailto:hanbeizhou@g.harvard.edu)), Rui Zhe Goh ([rgoh1@jhu.edu](mailto:rgoh1@jhu.edu)),  
Ian Phillips<sup>†</sup> ([ianbphillips@jhu.edu](mailto:ianbphillips@jhu.edu)), and Chaz Firestone<sup>†</sup> ([chaz@jhu.edu](mailto:chaz@jhu.edu))

Department of Psychological & Brain Sciences, Johns Hopkins University

<sup>†</sup> Joint senior authors

## Abstract

The timing of behavioral responses is among the most robust signatures of cognitive processing—with easier, more accurate, and more confident judgments typically made more quickly. But which aspects of response timing carry these signals? Whereas the time taken to *initiate* a response (e.g., the interval between stimulus onset and keypress) is among the most popular and informative measures in experimental psychology, another aspect of response timing is its *duration* (e.g., how long the key is held down before being released). Here, across 3 pre-registered experiments spanning perceptual and higher-cognitive tasks, we demonstrate that response duration reliably predicts subjective confidence, over and above response initiation time. When detecting faces (Experiment 1), judging motion coherence (Experiment 2), and answering trivia questions (Experiment 3), subjects held down a response key longer during trials in which they subsequently reported lower confidence in their judgments, as if making these judgments in a tentative fashion. In all cases, response duration accounted for variance in confidence that was not predicted by response initiation time. Our results suggest that response duration itself carries independent information about cognitive processing, advancing a useful and ubiquitously available tool to complement traditional response time measures in the study of cognition.

Version: 6/18/26; in press, *Proceedings of the National Academy of Sciences*

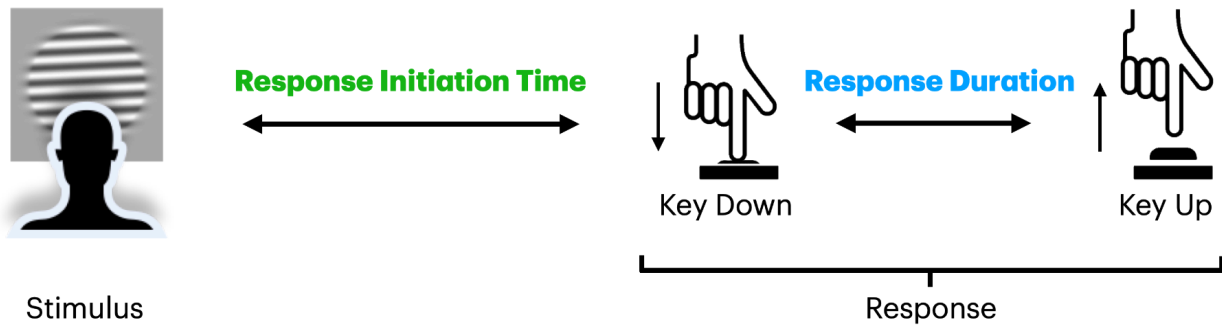
## Introduction

A foundational principle of experimental psychology is that the timing of external behaviors contains information about internal mental processes. This insight can be traced to at least 1868, when Franciscus Donders used the timing of behavioral responses to isolate different component processes of decision-making (1,2). Ever since, response timing has served as one of the most prominent measures in psychology, used to investigate phenomena ranging from basic motor skills (3) through attention allocation (4) to higher-level cognition (5).

Yet, response timing is more complex than it may appear. A typical response, such as a keypress, has at least two components: response initiation (pressing down the key) and response completion (releasing the key; Fig. 1). Most research using response timing focuses on *response initiation time*, i.e. the interval between stimulus appearance and response initiation. In contrast, the interval between response initiation and response completion—*response duration (RD)*—has been almost entirely neglected as a measure of cognitive processing. Such neglect is especially significant given the ubiquitous availability of this measure (response duration can be collected in any experiment with keypresses, without additional time or specialized equipment; cf. 6,7), as well as a recent demonstration that response duration varies independently of response initiation time (8).

These observations raise a crucial question: Does response duration carry meaningful information about the mind? Everyday experience suggests it might. For example, underconfidence in one's judgments may manifest not only in taking longer to initiate a response, but also in taking longer to complete it; consider hesitating before releasing a chess piece, or delaying lifting one's finger when clicking 'Send' on a high-stakes message. Does response duration provide reliable and independent information about confidence in our responses?

**The present studies: Response duration and confidence.** Here, across three experiments ranging from lower-level sensory processing to higher-level cognition, we ask whether response duration predicts confidence over and above response initiation time. Subjects made a simple judgment or decision (e.g., detecting motion, answering a trivia question), and indicated their response with a keypress. Following this, they rated their confidence in their answer. All three experiments revealed that response duration (i.e., how long subjects held down the key) reliably predicted confidence, independent of response initiation time.



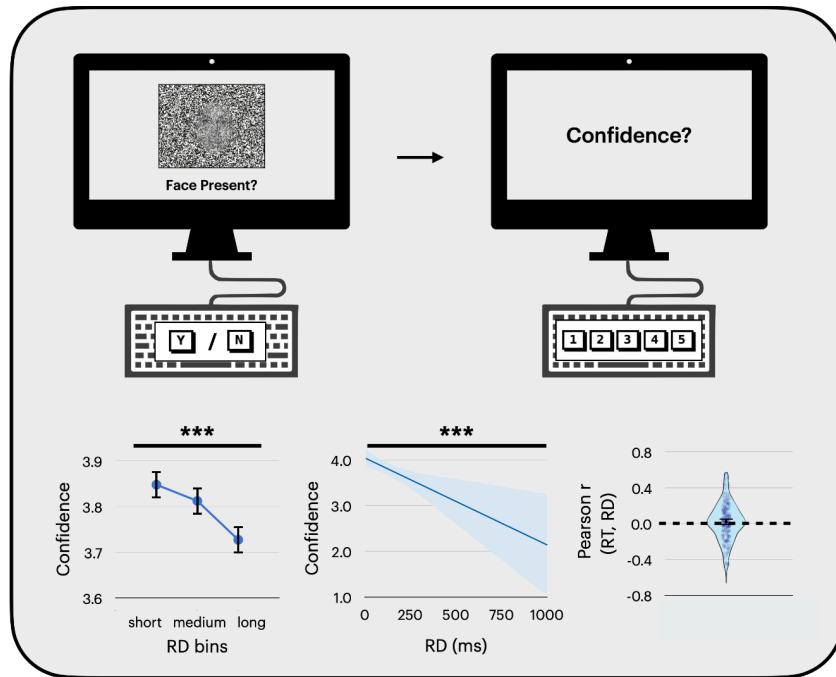
**Figure 1.** *Response Initiation Time and Response Duration.* A typical response, such as a keypress, consists of two stages: response initiation (i.e., KeyDown) and response completion (i.e., KeyUp). Response Initiation Time—often simply referred to as “response time” in the literature—is the interval between stimulus appearance and response initiation. In contrast, Response Duration (RD) is the interval between response initiation and response completion.

## Results

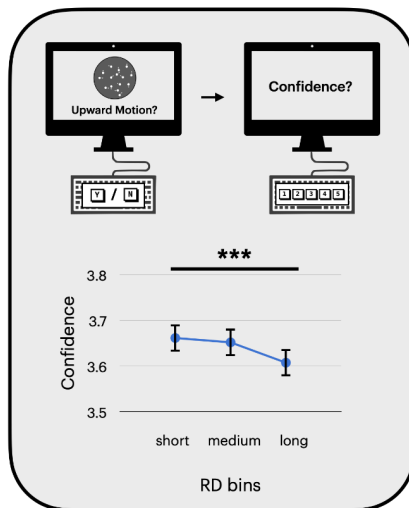
Experiment 1 comprised a face detection task. On each trial, a field of white noise either contained or did not contain a face, and subjects judged whether a face was present using a keypress. We recorded the duration between stimulus presentation and the first KeyDown event as response initiation time, and the duration between KeyDown and KeyUp events as response duration. After responding, subjects rated their confidence on a 1-5 scale, where 1 indicated low confidence and 5 indicated high confidence. Demonstrations are available at <https://perceptionresearch.org/responseduration>.

As per our prediction, a linear mixed-effects model with random intercepts for subjects revealed a significant negative relationship between response duration and confidence ( $\beta = -0.000908$ ,  $SE = 0.000157$ ,  $t = -5.79$ ,  $P < 0.0001$ ): Subjects held the key down longer when less confident in their response. This relationship remained significant after accounting for response initiation time by adding it as a regressor; although response initiation time also predicted confidence as expected ( $\beta = -0.000033$ ,  $SE = 0.00000130$ ,  $t = -25.59$ ,  $P < 0.0001$ ), response duration captured variance in confidence that was not explained by this traditional response timing measure ( $\beta = -0.000928$ ,  $SE = 0.000155$ ,  $t = -5.99$ ,  $P < 0.0001$ ). Furthermore, we replicated the independence (i.e., lack of relationship) between response initiation time and response duration found by Pfister et al. (mean Pearson  $r = 0.0128$ ,  $SE = 0.0154$ ,  $t(193) = 0.803$ ,  $P = 0.423$ ). Finally, a model including both response duration and response initiation time predicted confidence more accurately than a model with response initiation time alone ( $\chi^2(1) = 35.90$ ,  $P < 0.0001$ ). These results demonstrate that response duration tracks confidence independently of response initiation time.

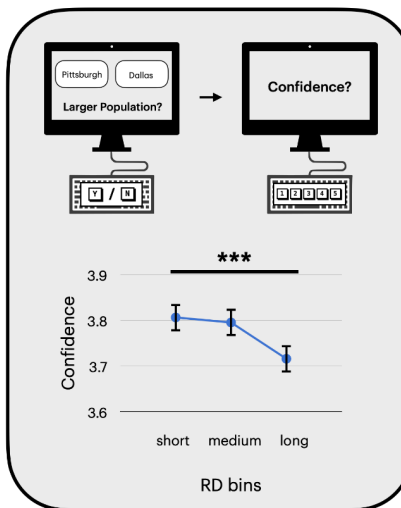
## Experiment 1



## Experiment 2



## Experiment 3



**Figure 2** *Stimuli, Procedures, and Results.* Top: In Experiment 1, a field of white noise either contained or did not contain a face, and subjects judged whether the face was present with a keypress. For visualization purposes, we grouped each subject's response durations into short, medium, and long bins; the average confidence for each bin decreases as response duration increases. A mixed-effects model with subjects as random intercepts revealed a correlation between confidence and response duration, with confidence decreasing as response duration increases. There is no significant correlation between response initiation time (RIT) and response duration. Bottom-Left: In Experiment 2, subjects saw an array of moving dots, and had to decide whether, as a group, the dots were moving upwards. Again, confidence decreases with the increase in response duration. Bottom-Right: In Experiment 3, subjects saw the names of two American cities and estimated which had a larger population. Response duration correlated negatively with confidence. All error bars represent 95% confidence intervals.

Experiments 2 and 3 generalized our findings further. Experiment 2 tested whether the link between response duration and confidence holds for other perceptual tasks. An independent group of subjects completed a motion detection task: Subjects saw an array of moving dots, and had to decide whether, as a group, the dots were moving upwards. As in Experiment 1, subjects gave their answer with a keypress, and then reported their confidence. Again, response duration correlated negatively with confidence ( $\beta = -0.000835$ ,  $SE = 0.000227$ ,  $t = -3.67$ ,  $P < 0.0005$ ). Moreover, this relationship remained significant when controlling for response initiation time ( $\beta = -0.000899$ ,  $SE = 0.000220$ ,  $t = -4.08$ ,  $P < 0.0001$ ). Thus, the relationship between response duration and confidence is not task-specific but generalizes more broadly.

Experiment 3 generalized our findings beyond perception, to a more cognitive task. On each trial, subjects were given the names of two American cities and estimated which had a larger population. Once again, response duration negatively predicted confidence: Subjects held down the response key longer when less confident in their choice ( $\beta = -0.001279$ ,  $SE = 0.000265$ ,  $t = -4.82$ ,  $P < 0.0001$ ). Furthermore, this effect held when controlling for response initiation time ( $\beta = -0.001419$ ,  $SE = 0.000255$ ,  $t = -5.56$ ,  $P < 0.0001$ ).

As in Experiment 1, Experiments 2 and 3 each found that (a) response duration and response initiation time were independent, and (b) a model including both response duration and response initiation time predicted confidence more accurately than a model with response initiation time alone (see SI Appendix for full analysis).

## Discussion

The present work demonstrates that a previously neglected aspect of response timing—the duration of responses—contains meaningful information about mental processing. Across three experiments spanning perceptual and cognitive tasks, the duration subjects held down a response key (rather than the time taken to initiate a response) reliably and independently predicted subjective confidence in the response itself.

Our findings illustrate how a familiar and ubiquitous aspect of psychological research—pressing a key in response to a stimulus—is not a stereotyped or inflexible event. Rather, this behavior carries information about higher-level psychological factors, in this case subjective confidence (cf. 9). In this way, our findings complement related work linking confidence to movement kinematics, such as the speed of one's finger when picking up a marble or pressing a touchscreen (6,7).

Our results also highlight the potential *usefulness* of response duration as a measure in psychology, perhaps especially for studying metacognition (10). Response duration is ubiquitous and freely available: It can be collected in any experiment with keypresses, without any additional burden of time, instruction, or specialized equipment—“a flexible and no-cost tool for psychological science” (8). (Indeed, some past datasets will already include measurements of response duration, providing a potentially rich and untapped source of extant data.) In certain contexts, response duration may even have advantages over traditional measures. Unlike retrospective confidence ratings solicited after a judgment, response duration tracks confidence at (or extremely close to) the moment of judgment, making it less cognitively demanding and perhaps less susceptible to post-decisional factors. Response duration can also easily be collected in populations for whom verbal or ratings-based confidence measures are inapplicable, such as non-human animals or pre-linguistic infants. Furthermore, in tasks where response initiation time is confounded with other experimental factors (or indeed is itself the key measure), response duration can serve as an implicit measure of confidence. Consider, for instance, visual search for a target among distractors; in such tasks, whereas response initiation time mostly reflects the duration of the search itself—i.e., the time taken to find the target—response duration remains independent of search time and so may offer a unique signal of post-search cognitive processes. Similarly, in temporal cognition research, timing tasks such as interval-production-by-waiting require subjects to initiate a response after a specified period has elapsed (e.g., 11); in such cases, response initiation time primarily reflects a temporal judgment, whereas response duration offers a potentially unique signal of non-temporal cognitive processing.

An important question for future research is the directionality of the present effects. It is natural to suppose that response time *reflects* confidence, such that we hesitate more because we are less confident. But confidence may also be influenced by response duration, such that we are less confident because we hesitate more (12, 13). Importantly, either possibility permits response duration to prove useful as described above.

While the present work focuses on confidence, one could also explore whether response duration carries information about other aspects of the mind. For example, response duration might reflect the speed of evidence accumulation. Indeed, one potential explanation of the relationship between response duration and confidence is that evidence accumulation continues after action initiation until a second action termination threshold is reached (cf. 14). Finally, an exciting avenue for future work would be to explore response duration in subjects other than human adults.

## Methods

Supplementary methods appear in *SI Appendix*. Studies were approved by the JHU IRB.

**Data, Materials, and Software Availability.** Data and materials are available at <https://osf.io/srtf4>

## References

1. Donders FC (1868) La vitesse des actes psychiques. *Archives néerlandaises des sciences exactes et naturelles* 3:296–317.
2. Roelofs A (2018) One hundred fifty years after Donders: Insights from unpublished data, a replication, and modeling of his reaction times. *Acta Psychol (Amst)* 191:228–233.
3. Fitts PM, Peterson JR (1964) Information capacity of discrete motor responses. *J Exp Psychol* 67:103–112.
4. Neisser U (1964) Visual search. *Scientific American* 210:94–102.
5. Greenwald AG, Banaji MR (1995) Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychol Rev* 102:4–27.
6. Patel, D., Fleming, S. M., & Kilner, J. M. (2012). Inferring subjective states through the observation of actions. *Proc R Soc B* 279:4853–4860.
7. Dotan, D., Meyniel, F., & Dehaene, S. (2018). On-line confidence monitoring during decision making. *Cognition* 171:112–121.
8. Pfister R, Neszvényi B, Kunde W (2023) Response durations: A flexible, no-cost tool for psychological science. *Curr Dir Psychol Sci* 32:160–166.
9. Oancea G, Maniscalco B, Peters MAK, Chapman CS (2025) Measuring motor awareness and metacognition at the start, middle, and end of a reaching movement. *Conscious Cogn* 132:103878.
10. Fleming SM (2024) Metacognition and confidence: A review and synthesis. *Annu Rev Psychol* 75:241–268.
11. Penton-Voak IS, Edwards H, Percival A, Wearden JH (1996) Speeding up an internal clock in humans? Effects of click trains on subjective duration. *J Exp Psychol Anim Behav Process* 22:307–320.
12. Kiani, R., Corthell, L., & Shadlen, M. N. (2014). Choice certainty is informed by both evidence and decision time. *Neuron* 84:1329–1342.
13. Fleming, S. M., Maniscalco, B., Ko, Y., Amendi, N., Ro, T., & Lau, H. (2015). Action-specific disruption of perceptual confidence. *Psycho Sci* 26:89–98.
14. Resulaj A, Kiani R, Wolpert DM, Shadlen MN (2009) Changes of mind in decision-making. *Nature* 461:263–266.
15. DeBruine L, Jones B (2017) Face Research Lab London Set. figshare. <https://doi.org/10.6084/m9.figshare.5047666.v5>.

## Supplementary Methods

All experiments reported here were pre-registered. Readers can view all experiments at <https://perceptionresearch.org/responseduration>.

### Experiment 1

Experiment 1 tested whether response duration predicts confidence in a face detection task.

#### *Participants*

We recruited 200 subjects from the online platform Prolific; data from 196 subjects were successfully received by the server. An additional 2 subjects were excluded in accordance with our pre-registered exclusion criteria (specified below): 1 subject was excluded for failing two or more catch trials, and 1 for having more than 4 excluded trials. 194 subjects remained after exclusion.

#### *Browser Compatibility Screening*

The experiment asked subjects to detect faces contained in a field of dynamic white noise. To ensure the white noise was displayed, subjects completed four screening trials, two of which contained white noise and two of which did not. On each trial, subjects judged whether there was white noise presented on the screen. Subjects had to answer all four trials correctly before proceeding to the experiment.

#### *Stimuli and Procedure*

On each trial, subjects were presented with a field of white noise which either contained or did not contain a face. Faces varied in opacity, as specified below. The more opaque a face is, the easier it is to detect. Subjects were asked to indicate whether there was a face by pressing 'Y' (for yes) or 'N' (for no) on their keyboard. We recorded the duration between stimulus presentation and the first KeyDown event as the response initiation time, and the duration between KeyDown and KeyUp events as the response duration. The stimuli stayed on the screen until a response was made. After responding, subjects were asked to indicate their confidence on a 1-5 scale, where 1 indicates low confidence and 5 indicates high confidence.

There were six opacities for the faces (0.5, 0.25, 0.175, 0.1, 0.05, and 0.025). The opacities were within the range [0, 1], where opacity of 1 is completely opaque and 0 is completely transparent. We used 10 different faces sampled from the Face Research Lab London Set (15), which were converted to grayscale and cropped to remove non-facial elements such as hair and background. Each subject completed 120 trials, 60

of which contained a face (6 opacities x 10 faces) and 60 of which did not, presented in random order.

We used the trials with an opacity of 0.5 as catch trials, since those faces were very easily visible. Any subject who failed two or more such catch trials was excluded. Additionally, we excluded trials with incomplete data and trials where subjects pressed both 'Y' and 'N' response keys. Any subject with more than four such excluded trials was excluded completely.

### *Analysis*

We used a linear mixed-effects model to estimate the relationship between response duration and confidence, with response duration as a fixed effect and subjects as random intercepts. The analysis revealed a significant negative correlation between response duration and confidence: Subjects held the key down longer on trials where they subsequently reported lower confidence in their response ( $\beta = -0.000908$ ,  $SE = 0.000157$ ,  $t = -5.79$ ,  $P < 0.0001$ ).

To determine whether this relationship between response duration and confidence was dependent on response initiation time, we fitted an additional mixed-effects model including both response duration and response initiation time as fixed effects, with subjects as random intercepts. This model isolates the variance in confidence uniquely predicted by response duration, by accounting for response initiation time. This model revealed that both response duration and response initiation time significantly correlated with confidence (response initiation time:  $\beta = -0.000033$ ,  $SE = 0.00000130$ ,  $t = -25.59$ ,  $P < 0.0001$ ; response duration:  $\beta = -0.000928$ ,  $SE = 0.000155$ ,  $t = -5.99$ ,  $P < 0.0001$ ). However, response duration remained a significant predictor of confidence even when response initiation time was included in the model, suggesting that response duration captured variance in confidence not explained by response initiation time.

Moreover, we computed a within-subject Pearson correlation between response initiation time and response duration across trials to verify the independence between response duration and response initiation time found by Pfister et al. (2023). The mean correlation was not significantly different from zero (mean Pearson  $r = 0.0128$ ,  $SE = 0.0154$ ,  $t(193) = 0.803$ ,  $P = 0.423$ ), suggesting in line with previous work that response duration and response initiation time are independent of one another.

Additionally, we tested whether adding response duration as a regressor improves overall prediction of confidence relative to response initiation time alone. To do this, we compared two linear mixed-effects models: one with response initiation time as the sole predictor of confidence, and the other including both response initiation time and

response duration. An ANOVA revealed that the model including both predictors provides a significantly better fit to reported confidence than the model including only response initiation time ( $\chi^2(1) = 35.90, P < 0.0001$ ).

Finally, we examined the relationship between response duration and confidence while controlling for accuracy. We fit an additional mixed-effects model that predicts confidence with both response duration and trial-by-trial accuracy as fixed effects, with subjects as random intercepts. This model revealed that both response duration and accuracy significantly correlated with confidence (Accuracy:  $\beta = 0.675371, SE = 0.01707926, t = 39.54331, P < 0.0001$ ; Response duration:  $\beta = -0.000782, SE = 0.00015199, t = -5.14267, P < 0.0001$ ), suggesting that response duration captures variance in confidence that is not explained by accuracy.

Together, these results suggest that response duration tracks confidence above and beyond traditional response timing measures.

## **Experiment 2**

Experiment 2 tested whether the link between response duration and confidence generalizes to other perceptual tasks, here a motion detection task.

### *Participants*

We recruited 200 subjects from the online platform Prolific; data from 199 subjects were successfully received by the server. An additional 13 subjects were excluded in accordance with our pre-registered exclusion criteria: 12 for failing at least two catch trials, and 1 for having more than 4 excluded trials). 186 subjects remained after exclusion.

### *Stimuli and Procedure*

On each trial, subjects were presented with a gray circle filled with moving dots. The dots were either all moving randomly or some moving randomly and some moving upwards. The proportion of upwards-moving dots varied across trials, as specified below. Subjects were asked to indicate whether as a group the dots were moving upwards by pressing 'Y' (for yes) or 'N' (for no) on their keyboard. Again, we recorded the duration between stimulus presentation and the first KeyDown event as the response initiation time, and the duration between KeyDown and KeyUp events as the response duration. The stimuli stayed on the screen until a response was given. After responding, subjects were asked to indicate their confidence on a 1-5 scale, where 1 means low confidence and 5 means high confidence.

There were six proportions of upward-moving dots (0.7, 0.2, 0.15, 0.1, 0.03, and 0.01). Each subject completed 120 trials, 60 of which contained upward motion (6 proportions x 10 repetitions) and 60 of which did not, presented in random order.

We used the trials with a proportion of 0.7 upwards-moving dots as catch trials. Any subject who failed 2 or more such catch trials was excluded. Additionally, we excluded trials with incomplete data and trials where subjects pressed both 'Y' and 'N' response keys. Any subject with more than four such excluded trials was excluded completely.

### *Analysis*

As in Experiment 1, we used a linear mixed-effects model to estimate the relationship between response duration and confidence, with response duration as a fixed effect and subjects as random intercepts. The analysis revealed a significant negative correlation between response duration and confidence: Subjects held the key down longer on trials where they subsequently reported lower confidence in their response ( $\beta = -0.000835$ ,  $SE = 0.000227$ ,  $t = -3.67$ ,  $P < 0.0005$ ).

To determine whether this relationship between response duration and confidence was dependent on response initiation time, we fitted an additional mixed-effects model including both response duration and response initiation time as fixed effects, with subjects as random intercepts. This model isolates the variance in confidence uniquely predicted by response duration, by accounting for response initiation time. This model revealed that both response duration and response initiation time significantly correlated with confidence (response initiation time:  $\beta = -0.000053$ ,  $SE = 0.00000140$ ,  $t = -37.85$ ,  $P < 0.0001$ ; response duration:  $\beta = -0.000899$ ,  $SE = 0.000220$ ,  $t = -4.08$ ,  $P < 0.0001$ ). However, response duration remained a significant predictor of confidence even when response initiation time was included in the model, suggesting that response duration captured variance in confidence not explained by response initiation time.

Moreover, we computed a within-subject Pearson correlation between response initiation time and response duration across trials to verify the independence between response duration and response initiation time found by Pfister et al. (2023). The mean correlation was not significantly different from zero (mean Pearson  $r = -0.0283$ ,  $SE = 0.0152$ ,  $t(185) = -1.90$ ,  $P = 0.059$ ), suggesting in line with previous work that response duration and response initiation time are independent of one another.

Additionally, we tested whether adding response duration as a regressor improves overall prediction of confidence relative to response initiation time alone. To do this, we compared two linear mixed-effects models: one with response initiation time as the sole predictor of confidence, and the other including both response initiation time and

response duration. An ANOVA revealed that the model including both predictors provides a significantly better fit to reported confidence than the model including only response initiation time ( $\chi^2(1) = 16.64, P < 0.0001$ ).

Finally, we examined the relationship between response duration and confidence while controlling for accuracy. We fit an additional mixed-effects model that predicts confidence with both response duration and trial-by-trial accuracy as fixed effects, with subjects as random intercepts. This model revealed that both response duration and accuracy significantly correlated with confidence (Accuracy:  $\beta = 0.363960, SE = 0.01361284, t = 26.737, P < 0.0001$ ; Response duration:  $\beta = -0.000858, SE = 0.000224, t = -5.56, P = 0.0001$ ), suggesting that response duration captures variance in confidence that is not explained by accuracy.

Together, these results suggest that the correlation between response duration and confidence is not task-specific, but generalizes to other perceptual tasks.

### **Experiment 3**

Experiment 3 tested whether the link between response duration and reported confidence generalizes to non-perceptual, cognitive tasks, such as answering trivia questions.

#### *Participants*

We recruited 200 subjects from the online platform Prolific; data from all subjects were successfully received by the server. 2 subjects were excluded in accordance with our pre-registered exclusion criteria: 1 for failing at least two catch trials, and 1 for having at least 1 trial with RD > 1000ms. An additional 180 trials were excluded for having an RT > 15000ms. 198 subjects remained after exclusion.

#### *Stimuli and Procedure*

On each trial, subjects saw the names of two American cities displayed on the two sides of their screen. They were asked to judge which city—the one on the left or the one on the right—has a larger population by pressing the Left or Right arrow key on their keyboard. There was a 1000 ms delay between trial onset and city presentation, and we recorded the duration between city presentation and the first KeyDown event as the response initiation time, and the duration between KeyDown and KeyUp events as the response duration. After responding, the chosen city turned gray, and subjects were asked to indicate their confidence on a 1-5 scale, where 1 means low confidence and 5 means high confidence.

There were 14 distinct cities in the experiment, including 7 large cities with populations over 1 million and 7 small cities with populations under 400,000. These cities formed 91 unique pairs: 49 easy pairs that contained a large city and a small city, and 42 difficult pairs that contained two large or two small cities. Six pairs (three easy pairs and three difficult pairs) were used in practice trials, leaving 85 pairs (46 easy pairs and 39 difficult pairs) for experimental trials. A complete list of cities and their populations is available at [https://osf.io/srtf4/overview?view\\_only=78dfd4e7dae84064a5558d41605586d4](https://osf.io/srtf4/overview?view_only=78dfd4e7dae84064a5558d41605586d4); population data were obtained from the 2023 United States Census Bureau.

We used trials containing easy city pairs as catch trials. Any subject who had an accuracy lower than 60% in these easy pairs was excluded. Additionally, we excluded trials with incomplete data and trials where subjects pressed both Left and Right arrow response keys. Any subject with more than four such excluded trials was excluded completely. Moreover, we excluded all trials with a response initiation time longer than 15,000ms. We also excluded any subject with at least one trial with a response duration longer than 1000ms, or with more than 25% trials with a response initiation time longer than 15,000ms.

### *Analysis*

As in Experiments 1 and 2, we used a linear mixed-effects model to estimate the relationship between response duration and confidence, with response duration as a fixed effect and subjects as random intercepts. The analysis revealed a significant negative correlation between response duration and confidence: Subjects held the key down longer on trials where they subsequently reported lower confidence in their response ( $\beta = -0.001279$ ,  $SE = 0.000265$ ,  $t = -4.82$ ,  $P < 0.0001$ ).

To determine whether this relationship between response duration and confidence was dependent on response initiation time, we fitted an additional mixed-effects model including both response duration and response initiation time as fixed effects, with subjects as random intercepts. This model isolates the variance in confidence uniquely predicted by response duration, by accounting for response initiation time. This model revealed that both response duration and response initiation time significantly correlated with confidence (response initiation time:  $\beta = -0.000201$ ,  $SE = 0.00000539$ ,  $t = -37.29$ ,  $P < 0.0001$ ; response duration:  $\beta = -0.001419$ ,  $SE = 0.000255$ ,  $t = -5.56$ ,  $P < 0.0001$ ). However, response duration remained a significant predictor of confidence even when response initiation time was included in the model, suggesting that response duration captured variance in confidence not explained by response initiation time.

Moreover, we computed a within-subject Pearson correlation between response initiation time and response duration across trials to verify the independence between

response duration and response initiation time found by Pfister et al. (2023). The mean correlation was not significantly different from zero (mean Pearson  $r = 0.00659$ ,  $SE = 0.0133$ ,  $t(197) = 0.399$ ,  $P = 0.69$ ), suggesting in line with previous work that response duration and response initiation time are independent of one another.

Additionally, we tested whether adding response duration as a regressor improves the overall prediction of confidence relative to response initiation time alone. To do this, we compared two linear mixed-effects models: one with response initiation time as the sole predictor of confidence, and the other including both response initiation time and response duration. An ANOVA revealed that the model including both predictors provides a significantly better fit to reported confidence than the model including only response initiation time ( $\chi^2(1) = 30.91$ ,  $P < 0.0001$ ).

Finally, we examined the relationship between response duration and confidence while controlling for accuracy. We fit an additional mixed-effects model that predicts confidence with both response duration and trial-by-trial accuracy as fixed effects, with subjects as random intercepts. This model revealed that both response duration and accuracy significantly correlated with confidence (Accuracy:  $\beta = 0.726614$ ,  $SE = 0.01965138$ ,  $t = 36.97520$ ,  $P < 0.0001$ ; Response duration:  $\beta = -0.001089$ ,  $SE = 0.00025511$ ,  $t = -4.26943$ ,  $P < 0.0001$ ), suggesting that response duration captures variance in confidence that is not explained by accuracy.

Together, these results suggest that the link between response duration and confidence extends to non-perceptual, cognitive tasks.