Figures and figure supplements

Confidence predicts speed-accuracy tradeoff for subsequent decisions

Kobe Desender et al
Figure 1. Schematic of drift diffusion model (DDM) with varying drift rates (A) and varying decision bounds (B). Noisy sensory evidence is accumulated over time, until the decision variable reaches one of two bounds (a or -a), corresponding to correct and incorrect choices. The efficiency of information accumulation is given by v (mean drift rate). The time for sensory encoding and response execution is given by $T_{er}$. By increasing the separation between decision bounds, the probability of being correct increases, at the expense of prolonged reaction times. RT distributions (upper bounds) and error distributions (lower bounds) are depicted for different levels of drift rate (A) and decision bound (B).

DOI: https://doi.org/10.7554/eLife.43499.002
Figure 2. Experimental task and results from Experiment 1. (A) Sequence of events in a trial from Experiment 1. Participants decided, as fast and accurately as possible, whether the majority of dots were moving left or right. After their response and a 1 s blank or 1 s of continued motion, they...
indicated the degree of confidence in their decision using a six-point confidence scale (ranging from certainly correct to certainly wrong). (B) Mean reaction time on correct trials (top), accuracy (middle) and estimated mean drift rate (bottom) as a function of coherence. Green crosses show fits from the DDM. (C) Confidence as a function of coherence level and error category. (D) Accuracy as a function of decision confidence (dot size reflects percentage of trials per confidence label, separately for each participant). (E) Accuracy as a function of reaction time. Data are pooled across the blank and the continued motion condition. Gray dots: individual participants, black dots: group averages.
Figure 2—figure supplement 1. Similar confidence judgments for blocks with and without post-decisional evidence. (A-B) Confidence as a function of coherence level, separately for blocks with and without post-decisional evidence (A and B, respectively). The interaction between coherence level and confidence judgments is illustrated in Figures 2A and 2B. (C-D) Accuracy as a function of confidence, with error bars indicating standard error. The accuracy is shown for four different coherence levels: 0.5%, 40%, and 80% confidence.
Figure 2—figure supplement 1 continued

Choice accuracy in predicting confidence was highly similar and significant in both (post-decisional evidence: F(4,3387.2) = 88.32, p<0.001; post-decisional blank: F(4,3450.0) = 71.36, p<0.001) (C–D) Accuracy as a function of decision confidence, separately for blocks with and without post-decisional evidence (C and D, respectively). Pairwise comparisons did not reveal any differences between the mean accuracy at any of the confidence levels, all |ts| < 1.959, all ps > 0.060. Dot size reflects the percentage of trials per confidence label, separately for each participant.

DOI: https://doi.org/10.7554/eLife.43499.004
Figure 3. The influence of decision confidence on subsequent decision bound. (A) Model-free measures of response caution on trial n+one as function of confidence on trial n: mean RT, accuracy, and their product. Inset, distribution of empirical and fitted RTs. (B) Model-based estimate of decision bound and drift rates on trial n+one as function of confidence on trial n. Distributions show the group posteriors over parameter estimates. In all panels, as well as in all subsequent figures, ‘delta’ on y-axis refers to deviation of dependent variable from its value in the high-confidence condition (i.e., centered on zero for high-confidence). For model fits, trials from the high-confidence condition served as reference so that parameter estimates reflected deviations from their value on high-confidence. Statistical significance is reflected in overlap between posterior distributions over parameter estimates (Materials and methods).

DOI: https://doi.org/10.7554/eLife.43499.005
Figure 3—figure supplement 1. The influence of decision confidence on subsequent decision bounds and drift rates, separately for blocks with post-decisional evidence presentation (A–B) and a post-decisional blank (C–D). Due to a lack of data in one of the cells, these analyses were performed on the data of 27 and 24 participants, respectively. Same conventions as in Figure 3.

DOI: https://doi.org/10.7554/eLife.43499.006
Figure 3—figure supplement 2. Simple effects of confidence on trial \(_n\) and confidence on trial \(_{n+2}\) on the product of subsequent RTs and accuracy as a model-free measure of decision bound (Experiment 1). Same conventions as in Figure 3.
DOI: https://doi.org/10.7554/eLife.43499.007
Figure 3—figure supplement 3. Simple effects of confidence on trial\textsubscript{n} and confidence on trial\textsubscript{n+2} on decision bound (A) and drift rate (B) on trial\textsubscript{n+1} (Experiment 1). Same conventions as in Figure 3.
DOI: https://doi.org/10.7554/eLife.43499.008
Figure 3—figure supplement 4. Complementary approach controlling for slow drifts in performance (Experiment 1). Separate models were fitted estimating the decision bound and drift on trial \( n \), and i) we compared low confidence on trial \( n \) to high confidence on trial \( n \) for which trial \( n+2 \) was a low confidence trial, and ii) we compared perceived errors on trial \( n \) to high confidence trials on trial \( n \) for which trial \( n+2 \) was a perceived error. Note that low confidence and perceived errors cannot be compared directly because they stem from different models. These results remained unchanged when only including correct trials (there were not enough trials when only including errors).

DOI: https://doi.org/10.7554/eLife.43499.009
Figure 4. Confidence-dependent modulation of decision bound remains qualitatively similar when modeling only correct (A) or only error trials (B). Same conventions as in Figure 3. Note that, due to a lack of trials in one of the cells, results are based on 25 (A) and 24 (B) participants.

DOI: https://doi.org/10.7554/eLife.43499.010
Figure 5. Confidence-dependent modulation of decision bounds generalizes to different experimental tasks. (A) In Experiment 2, participants need to decide as quickly as possible which of two boxes has more dots. In Experiment 3, participants needed to decide, as fast and accurately as possible, whether the average color of the eight elements was more red or blue. (B) Subsequent RT, accuracy and their product as a model-free measure of decision bound as a function of confidence. Green crosses show fits from the DDM. (C) Subsequent decision bounds and subsequent drift rates as a function of confidence. Inset in B shows the distribution of empirical and fitted RTs.

DOI: https://doi.org/10.7554/eLife.43499.011
Behavioral results of Experiment 2. (A) Accuracy as a function of decision confidence. Confidence ratings were closely linked to choice accuracy, even on a trial-by-trial basis (logistic regression of confidence on accuracy: significant positive slopes for all 16 observers, with all ps < 0.001). Dot size reflects the percentage of trials per confidence label, separately for each participant. (B) Accuracy as a function of reaction time. Although there was a significant linear relationship between RTs and accuracy, b = 0.022, t(15) = 4.19, p<0.001, a polynomial regression confirmed the apparent inverted U-shape seen in the data, t(15) = –7.52, p<0.001.

DOI: https://doi.org/10.7554/eLife.43499.012
Figure 5—figure supplement 2. Simple effects of confidence on trial_\textit{n} and confidence on trial_\textit{n+2} on the product of subsequent RTs and accuracy as a model-free measure of decision bound for Experiment 1 (A) and Experiment 2 (B). Same conventions as in Figure 3.
DOI: https://doi.org/10.7554/eLife.43499.013
Figure 5—figure supplement 3. Simple effects of confidence on trial, and confidence on trial, on decision bound (A) and drift rate (B) on trial, (Experiment 2). Trials with high confidence are treated as reference category.

DOI: https://doi.org/10.7554/eLife.43499.014
Figure 5—figure supplement 4. Complementary approach controlling for slow fluctuations (Experiment 2). Same procedure as in Figure 3—figure supplement 3. These results remained unchanged when only including correct trials (there were not enough trials when only including errors).

DOI: https://doi.org/10.7554/eLife.43499.015
Figure 5—figure supplement 5. Behavioral results of Experiment 3. (A) Mean reaction time on correct trials (top), accuracy (upper middle) estimated mean drift rate (lower middle), and confidence (bottom) as a function of coherence level. RTs on correct trials and choice accuracy scaled with SNR in both experiments (Exp 3A: RT, F(4, 3012.3)=21.60, p<0.001, error rates, $\chi^2$(4)=170.74, p<0.001; Exp 3B: RT, F(3, 2173.1)=3.07, p=0.027, error rates, $\chi^2$(3) =83.311, p<0.001). Correspondingly, drift rates estimated from fits of a hierarchical drift diffusion model also increased monotonically with SNR in both experiments (Exp 3A: Friedman $\chi^2$(4)=41.02, p<0.001; Exp 2B: Friedman $\chi^2$(3)=31.61, p<0.001). In these model fits, decision bound was not allowed to vary as a function of SNR, and its average was 1.33 (SD = 0.17) and 1.33 (SD = 0.20) in Experiment 3A and 3b, respectively. Similarly, non-decision time was estimated independently from SNR, and its average was 0.28 (SD = 0.08) and 0.22 (SD = 0.05). Green crosses show the fits from the DDM. (B) Accuracy as a function of decision confidence. Confidence ratings were closely linked to choice accuracy, even on a trial-by-trial basis while including SNR as a covariate (logistic regression of confidence on accuracy: significant positive slopes for 21 observers, with all ps < 0.001, and non-significant positive slopes for two observers, with both ps > 0.354), showing that confidence predicts accuracy over and above SNR. Dot size reflects the percentage of trials per confidence label, separately for each participant. (C) Accuracy as a function of reaction time. There was no linear relationship between RTs and accuracy, b = −0.001, p>0.545. A polynomial regression did, however, confirm the inverted U-shape seen in the data, $t$(29.2) = −3.19, p=0.003. (D) Confidence as a function of coherence level, separately for corrects and errors. Confidence judgments only scaled with SNR in Experiment 3A (Exp 3A: F(4,3953.9) = 37.34, p<0.001; Exp 3B: F(3,2867.1) < 1, p>0.141), but there was a highly significant interaction between SNR and accuracy in both versions of the experiment (Exp 3A: F(4,3955.1) = 87.59, p<0.001; Exp 3B: F(3,2867.8) = 48.16, p<0.001): confidence increased with SNR for corrects (linear contrast: Exp 3A: p<0.001, Exp 3B: p<0.001) and decreased with SNR for errors (linear contrast: Exp 3A: p<0.001, Exp 3B: p<0.001).

DOI: https://doi.org/10.7554/eLife.43499.016
Figure 5—figure supplement 6. Simple effects of confidence on trial\textsubscript{n} and confidence on trial\textsubscript{n+2} on decision bound (A) and drift rate (B) on trial\textsubscript{n+1} (Experiment 3). Trials with high confidence are treated as reference category.

DOI: https://doi.org/10.7554/eLife.43499.017
Figure 5—figure supplement 7. Complementary approach controlling for slow drifts in performance (Experiment 3). Same procedure as in Figure 3—figure supplement 3. When only including correct trials, the effect of perceived errors on drift rate turned non-significant, p=0.250, apart from that the results remained unchanged (there were not enough trials when only including errors).

DOI: https://doi.org/10.7554/eLife.43499.018
Figure 6. Confidence-dependent modulation of decision bound remains qualitatively similar when modeling only correct (A) or only error trials (B). Due to a lack of trials in one of the cells, the results shown in A and B are based on 11 participants for Experiment 2, and respectively 22 and 23 participants for Experiment 3.

DOI: https://doi.org/10.7554/eLife.43499.019
Figure 7. Post-decisional ERPs in Experiment 2. (A) Response-locked grand-average ERPs at electrode Pz, separately for the three levels of confidence. Gray bars represent the ERN (−10 ms to 90 ms) and the Pe (250 ms to 350 ms). Inset shows the same ERPs after regressing out the ERN. Topographic plots display amplitude differences between high confidence and perceived errors. (B–C) Average number of trials in each bin of the Pe (C) and the ERN (D), separately for the levels of confidence.

DOI: https://doi.org/10.7554/eLife.43499.020
Figure 8. Subsequent decision bounds and drift rates (A) as a function of Pe and ERN. (B) Pe-dependent variations in subsequent decision computation. (C-D) Modeling results remain qualitatively similar when modeling only correct (B) or only error trials (C). The bin with the lowest Pe
amplitude quantile was always treated as reference category (i.e., fixed to zero). Data were fit using the regression approach, so values reflect coefficients.

DOI: https://doi.org/10.7554/eLife.43499.021
Figure 8—figure supplement 1. Simple effects of rank-ordered Pe and ERN amplitude on trial $n$ (A and C) and trial $n+2$ (B and D) on decision bound (A and B) and drift rate (C and D) on trial $n+1$.

DOI: https://doi.org/10.7554/eLife.43499.022
Figure 8—figure supplement 2. Pe effects without regressing out the ERN. (A-B) Pe-dependent variations in subsequent decision computation without regressing out the ERN. Subsequent decision bounds (C) and drift rates (D). The bin with the lowest Pe amplitude was always treated as reference category (i.e., fixed to zero).

DOI: https://doi.org/10.7554/eLife.43499.023
Simple effects of binned Pe amplitude on trial n (A and C) and binned Pe amplitude trial n+2 (B and D) on decision bound (A and B) and drift rate (C and D) on trial n+1 (Experiment 2). Same conventions as in Figure 8, the bin with the lowest mean amplitude is treated as reference category. Bins were created after first regressing out the ERN from the raw signal.

DOI: https://doi.org/10.7554/eLife.43499.024
Figure 8—figure supplement 4. Complementary approach controlling for slow drifts in performance (EEG data). Same procedure as in Figure 3—figure supplement 3. The results remained unchanged when only including correct trials (there were not enough trials when only including errors).

DOI: https://doi.org/10.7554/eLife.43499.025