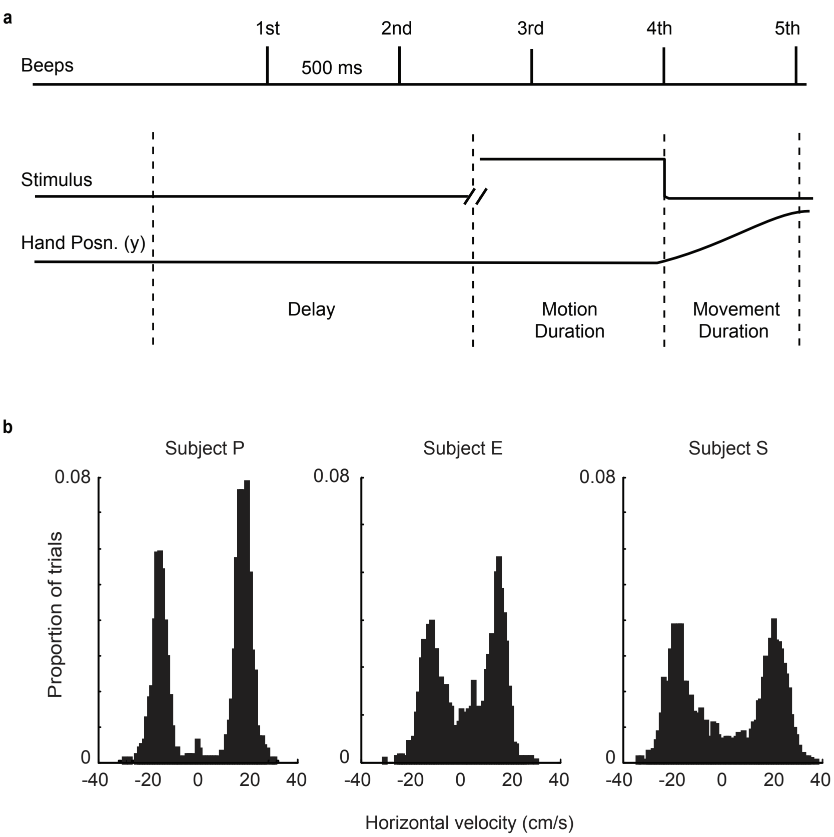
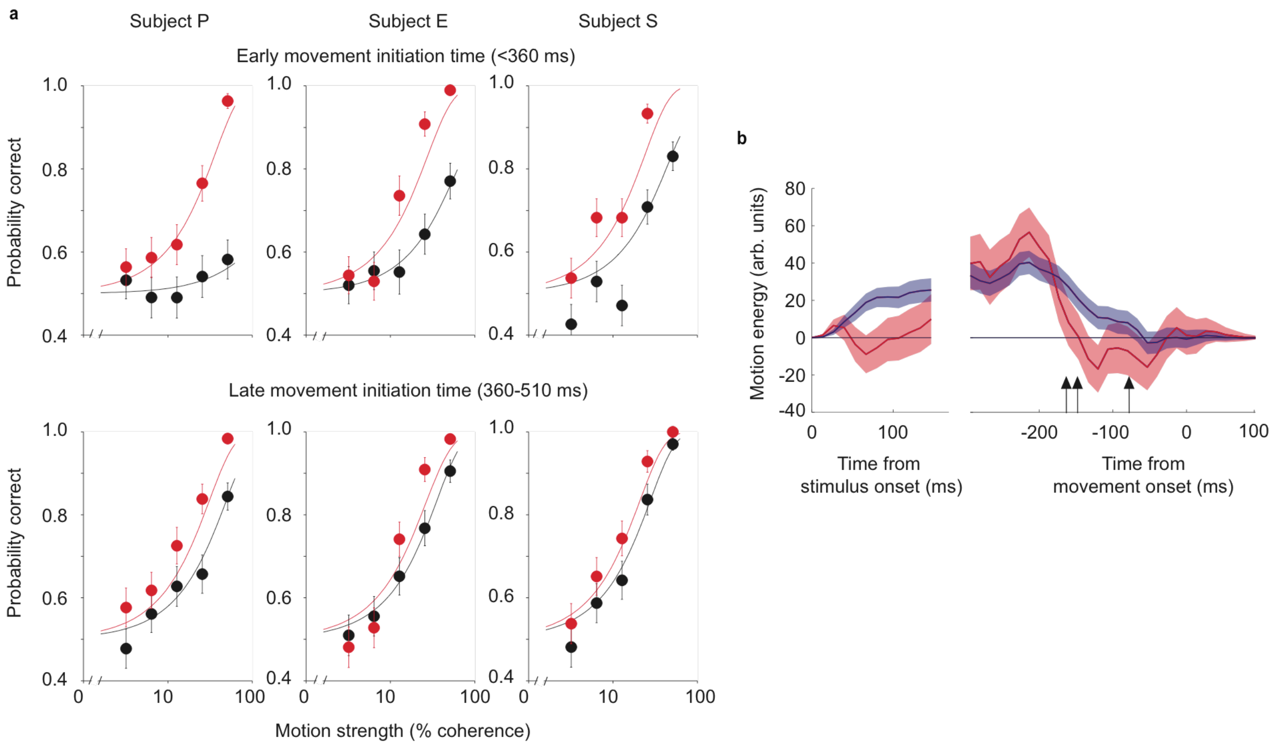
Supplementary material

Changing your mind: a computational mechanism of vacillation

Arbora Resulaj, Roozbeh Kiani, Daniel M. Wolpert & Michael N. Shadlen

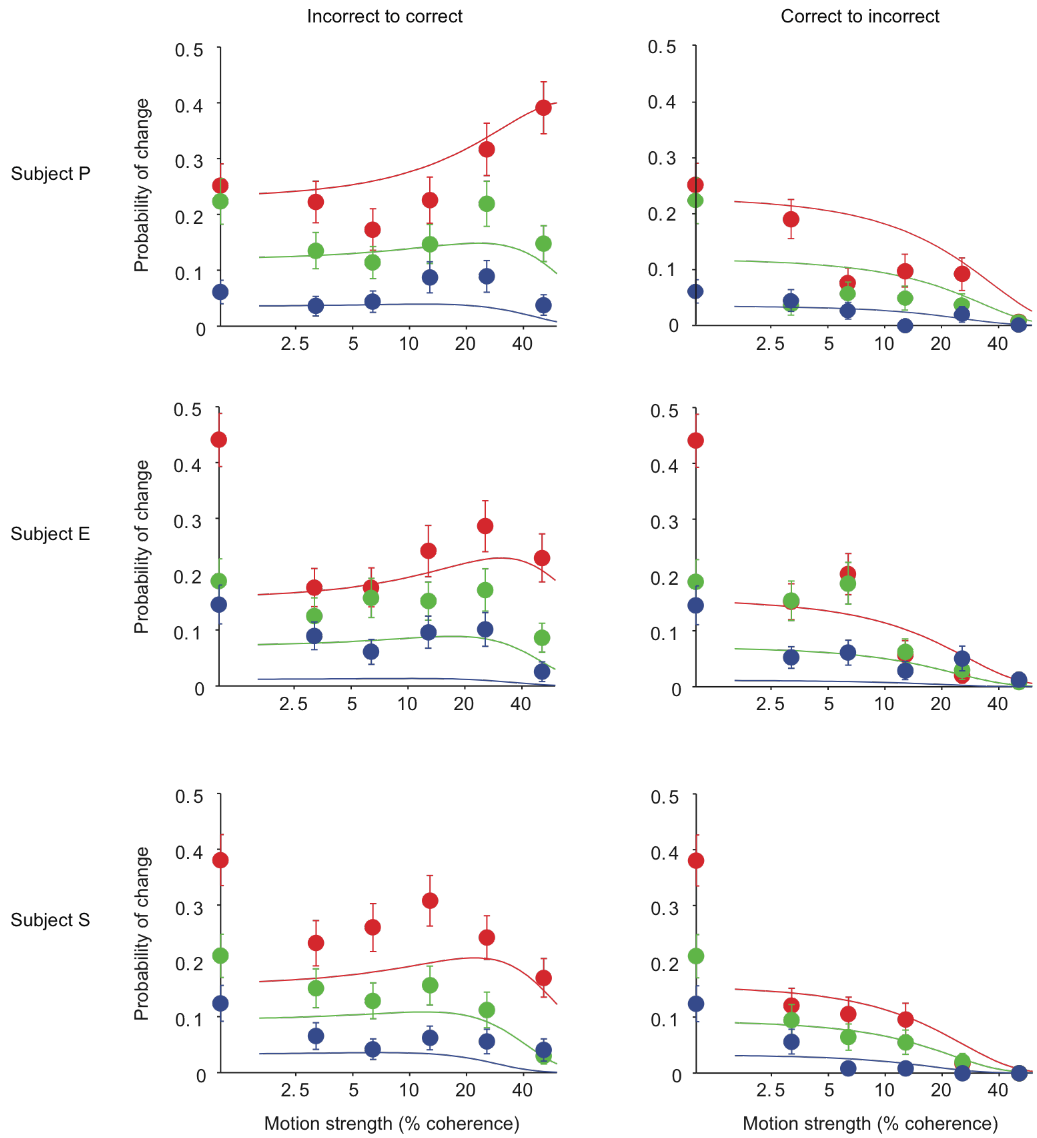


Supplementary Figure 1. Cued movement task. **a**, Behavioral task. As in the main experiment, subjects identified the direction of motion of dynamic random dot stimuli. Instead of responding when ready, subjects were trained to time the initiation of their movement so that it coincided with the 4^{th} of a series of 5 beeps and to bring the cursor to the choice target on the 5^{th} beep. The stimulus motion began at a random time 200-2000 ms (mean 440 ms) before the 4^{th} beep and ended at the beep or at movement initiation, whichever occurred first (see Methods for additional details). **b**, Distribution of initial hand velocity in the cued condition for the 3 subjects. Histograms of lateral (horizontal) hand velocity when it had moved 2 cm outward from the home position. Hartigan's test of unimodality[37] shows that the distributions are bimodal ($p<0.001$ for all subjects). Note that the classification of initial choices was not based on these measurements but instead on analysis of the hand trajectories (see Methods). Subjects clearly indicated a choice at initiation by moving in the direction of one of the targets. On many trials, however, subjects reversed this initial choice during the movement (see Supplementary Figure 2).

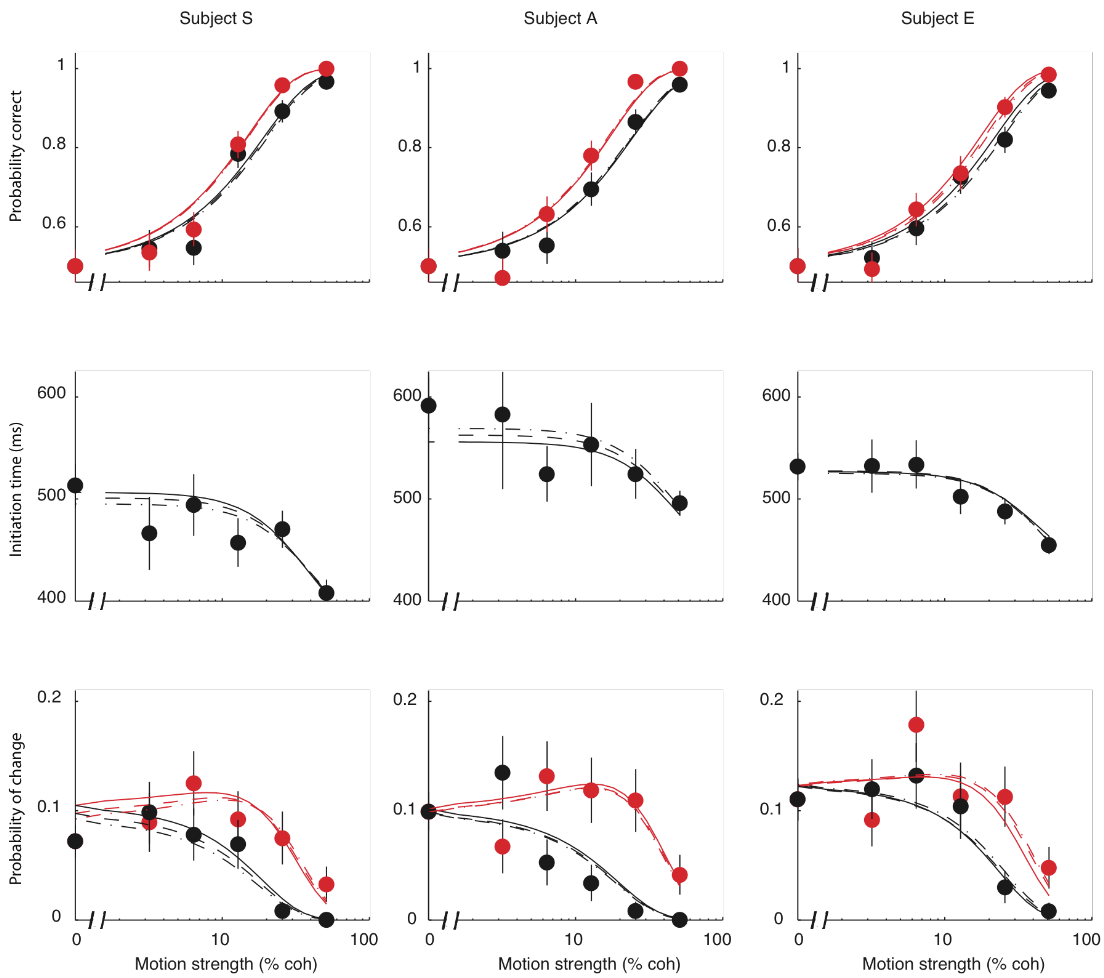


Supplementary Figure 2. Accuracy improves through "changes of mind" in the cued movement task. **a**, Data from three subjects. The top row shows the probability of a correct decision at initiation (black) and at termination (red) for the trials with movement initiation times (with respect to stimulus onset) shorter than 360ms (1st tercile). The bottom row shows trials with later movement initiation times (360-510 ms; 2nd tercile). Smooth curves are model fits; error bars are s.e.m. **b**, Influence of motion energy fluctuations on initial and final decisions. The data are shown for all the trials (blue) and the subset of trials with a change of mind (red) aligned at stimulus onset (left) and movement onset (right). Same conventions as in Figure 3a (main text). Arrows indicate the time from movement onset that the average motion energy fluctuations for each subject falls to within 1 s.e. of zero.

Notes. For the shorter durations, the changes of mind led to an improvement in decision accuracy ($p<0.003$ for all 3 subjects). In contrast to the previous model, the decision variable might fail to reach a termination bound before initiation, especially for the shorter duration stimuli, whereas it may reach a bound some time prior to initiation, especially for the longer duration stimuli. A small adjustment to the model accommodates these two situations (see Methods). For all subjects, the model explains the degree of improvement in accuracy associated with changes of mind (panel a, solid lines, R^2 = 0.92 to 0.95 across subjects) and the degree of improvement in both the initial and final choices as a function of viewing duration. Moreover, it accounts for the proportion of changes to correct and to erroneous choices as a function of motion strength and stimulus duration (Supplementary Figure 3). The motion energy analysis suggests that information acquired at the beginning of motion viewing affects the initial choice (panel b, left blue curve; over first 150 ms $p<0.0001$), whereas fluctuations near the end of motion viewing are unrelated (right blue curve). In contrast, these late fluctuations do influence the decision changes that occur after initiation. Near the time of initiation, the motion energy on change of mind trials was significantly less than the motion energy on the remaining trials ($p<0.0001$; bootstrap, see Methods). Compared to the reaction time experiment, the motion fluctuations appear to influence the initial choice later into the trial (panel b, arrows), consistent with approximately 100 ms shorter t_{nd} in the model fits (Supp. Table 2). We suspect that this may be a consequence of the external cue to initiation, resembling the conditions that are known to produce 'express' saccadic latencies[38].



Supplementary Figure 3. Probability of change of mind from incorrect to correct (left column) and from correct to incorrect (right column) in the cued movement task varies with motion strength and with the time of movement initiation. Data were divided into 3 quantiles based on the movement initiation time relative to motion onset (red, the shortest quantile; blue, the longest quantile). The curves are the expected values based on the model fits shown in Supplementary Fig. 2a. The model tends to underestimate the frequency of changes of mind because it does not incorporate initial biases or initial guesses. Despite this limitation, the fit captures qualitative differences between subjects and between correct and error vacillations.



Supplementary Figure 4. Cross validation for the reaction time experiment. The graphs correspond to those shown in Figures 2 and 3b of the paper. The symbols are means using a random half of the data for each subject. Solid curves are model predictions, based on the fit to the other half of each subject's data. Dashed and dot-dash curves show the fits to the entire data set and to the half of the data shown here. Error bars are s.e.m. The similarity of the predictions and fits provides reassurance that the model is not over parameterized (see also Supplementary Table 3).

	Subject S	Subject E	Subject A
k	0.30±0.025	0.27±0.03	0.25±0.03
B	13.2±0.6	12.4±0.5	13.0±1.3
t_{nd} *right* (ms)	322±12	368±11	390±28
t_{nd} *left* (ms)	326±13	378±12	395±29
μ_0	0.006±0.002	0.018±0.003	0.013±0.003
y_0	0.0004±0.0002	0.0003±0.00005	-0.0005±0.0002
B_Δ	23.3±0.6	18.4±1.0	25.5±0.7
PIP duration as fraction of t_{nd}	1.0 (CI: >0.8)	0.68 (CI: 0.54 -0.7)	1.0 (CI: >0.95)

Supplementary Table 1. Fitted parameters of the accumulation-to-bound model with post-initiation processing in the reaction time experiment. The six parameters for initiation and two parameters for post-initiation processing are shown with standard errors for each subject (see Methods). The interval for post-initiation processing (PIP) is bounded by t_{nd}; parentheses show the 95% confidence intervals (CI).

	Subject S	Subject E	Subject P
k	0.30±0.03	0.26±0.03	0.19±0.05
B	16.1±2.8	15.1±7.8	22.9±7.5
t_{nd} (ms)	260±18	273±12	273±12

Supplementary Table 2. Fitted parameters of the accumulation-to-bound model with post-initiation processing in the cued experiment. The three parameters are shown with standard errors for each subject.

	S $R^2_{all}, R^2_{half}, R^2_{pred}$	*E* $R^2_{all}, R^2_{half}, R^2_{pred}$	*A* $R^2_{all}, R^2_{half}, R^2_{pred}$
Prob Correct init (Fig 2, top black)	0.956, 0.938, 0.939	0.98, 0.98, 0.96	0.954, 0.976, 0.976
Prob Correct final (Fig 2, top red)	0.976, 0.962, 0.960	0.99, 0.97, 0.96	0.964, 0.938, 0.937
Initiation time (Fig 2, lower)	0.92, 0.69, 0.54	0.87, 0.89, 0.87	0.74, 0.61, 0.56
Prob Ch. to correct (Fig 3b, red)	0.63, 0.67, 0.47	0.85, 0.51, 0.43	0.72, 0.64, 0.62
Prob Ch. to error (Fig 3b, black)	0.76, 0.84, 0.82	0.99, 0.90, 0.89	0.88, 0.769, 0.767
mean	0.85, 0.83, 0.75	0.94, 0.90, 0.82	0.85, 0.79, 0.77
combined	0.91, 0.69, 0.55	0.87, 0.89, 0.88	0.74, 0.61, 0.56

Supplementary Table 3. Fraction of variance for the cross validation in the reaction time experiment. The R^2 terms describe the fraction of variance in data that is explained by the smooth curves (model) in Figures 2, 3b and Supplementary Figure 3. For each of the five predicted functions and for each subject, there are three R^2 terms. The R^2_{all} are computed from the data and model fits shown in the main text figures, using all trials. The R^2_{half} are computed from a random half of the data and model fit to these data (Supplementary Figure 3, dot-dash curve). The R^2_{pred} are computed from the same random half of the data and the model fit to the other half of the data (Supplementary Figure 3, solid line). This last term furnishes the cross validation: a measure of how well the model derived from one set of observations *predicts* a different set of observations. The bottom two rows show (*i*) the means of the R^2 values from the five graphs and (*ii*) the R^2 calculated from the total sum of squares and total residual sum of squares from the data and model predictions combined across the five graphs. With few exceptions, as expected from the lowered number of trials, the fits to the random half of the data are worse than the fits to the entire data set. Critically, the cross validation (prediction) is only slightly worse than the fit to the same data, confirming that the model does not over fit the data. Note that for each subject, one model accounts for all five comparisons.

	Subject S *orig, 1st, 2nd (%Δ)*	Subject E *orig, 1st, 2nd (%Δ)*	Subject A *orig, 1st, 2nd (%Δ)*
k	0.30, 0.30, 0.30 (0.6%)	0.27, 0.26, 0.32 (19.3%)	0.25, 0.26, 0.25 (1.5%)
B	13.2, 12.8, 13.6 (5.3%)	12.4, 12.5, 11.4 (8.8%)	13.0, 13.1, 12.8 (2.6%)
t_{nd} *right* (ms)	322, 325, 319 (1.9%)	368, 363, 391 (7.3%)	390, 393, 388 (1.3%)
t_{nd} *left* (ms)	326, 332, 320 (3.3%)	378, 375, 402 (6.5%)	395, 400, 393 (1.7%)
μ_0	0.006,0.006,0.006 (0.21%)	0.018, 0.019,0.020 (5%)	0.013, 0.014,0.013 (6.5%)
y_0	0.0004, 0.0004,0.0004 (2.2%)	0.0003,0.0004,0.0003(9.5%)	-0.0005,-0.0005,-0.0005 (5.5%)
B_Δ	23.3, 24, 22.3 (7.2%)	18.4, 18.3, 16.2 (11.7%)	25.5, 25.7, 25.0 (2.7%)
PIP duration as fraction of t_{nd}	1, 1, 1 (0%)	0.68, 0.69, 0.50 (27.7%)	1, 1, 1 (0%)

Supplementary Table 4. Model parameter estimates from cross validation for the reaction time experiment. Each column contains the original parameter estimate based on all data (same as Supplementary Table 1), and parameter estimates from fits to random halves of the data. The term in parentheses is the percentage difference between the latter two estimates: the absolute value of the difference, divided by the larger of the two estimates.

Supplementary Notes

[37] Hartigan, J.A. & Hartigan, P.M., The Dip Test of Unimodality. *Annals of Statistics* 13 (1), 70-84 (1985).

[38] Fischer, B. & Boch, R., Saccadic eye movements after extremely short reaction times in the monkey. *Brain. Res.* 260, 21-26 (1983).