Hierarchical decision processes that operate over distinct timescales underlie choice and changes in strategy

Braden A. Purcell[a] and Roozbeh Kiani[a,1]

[a]Center for Neural Science, New York University, New York, NY 10003

Edited by Charles R Gallistel, Rutgers University, Piscataway, NJ, and approved June 10, 2016 (received for review December 17, 2015)

Decision-making in a natural environment depends on a hierarchy of interacting decision processes. A high-level strategy guides ongoing choices, and the outcomes of those choices determine whether or not the strategy should change. When the right decision strategy is uncertain, as in most natural settings, feedback becomes ambiguous because negative outcomes may be due to limited information or bad strategy. Disambiguating the cause of feedback requires active inference and is key to updating the strategy. We hypothesize that the expected accuracy of a choice plays a crucial rule in this inference, and setting the strategy depends on integration of outcome and expectations across choices. We test this hypothesis with a task in which subjects report the net direction of random dot kinematograms with varying difficulty while the correct stimulus–response association undergoes invisible and unpredictable switches every few trials. We show that subjects treat negative feedback as evidence for a switch but weigh it with their expected accuracy. Subjects accumulate switch evidence (in units of log-likelihood ratio) across trials and update their response strategy when accumulated evidence reaches a bound. A computational framework based on these principles quantitatively explains all aspects of the behavior, providing a plausible neural mechanism for the implementation of hierarchical multiscale decision processes. We suggest that a similar neural computation—bounded accumulation of evidence—underlies both the choice and switches in the strategy that govern the choice, and that expected accuracy of a choice represents a key link between the levels of the decision-making hierarchy.

hierarchical decision-making | adaptive behavior | perceptual decision-making | executive control | confidence

Goal-directed behavior in natural settings depends on a hierarchy of decision processes. Higher-level decision strategies establish potential actions and expected outcomes for lower-level choices about incoming stimuli (1, 2). However, the correct strategy is rarely known a priori, and must be inferred from the outcome of decisions. As a result, the cause of negative outcomes is often ambiguous. An error could be due to a poor decision strategy, in which case the strategy should be promptly revised, or an error may be due to limited information, in which case the underlying strategy may still be sound. This ambiguity is particularly problematic because the environment can change without warning, altering the true associations between choices and outcomes and rendering a previously good strategy ineffective. Resolving the cause of negative feedback requires one to make inferences about strategy over multiple choices, but the mechanisms by which lower-level choices interact with higher-level decisions about strategy are poorly understood.

Expected accuracy in our choices (i.e., choice confidence) can be an important source of information for disambiguating negative feedback. If choices begin to yield negative outcomes despite strong positive expectations, then this provides strong evidence that the strategy must change. For example, consider a physician treating a patient based on an initial diagnosis. If the doctor knows that a treatment is highly effective for this ailment, but the patient's health still declines, then this provides strong evidence that the diagnosis should be reconsidered. Alternatively, if the treatment is known to be unreliable, then the doctor may persist with other treatment options before reconsidering the diagnosis. At the core of this example and many similar hierarchical decisions is the use of confidence to set the decision strategy and guide future behavior.

Recent behavioral, computational, and neurophysiological studies have provided key insights into the mechanisms by which choice confidence is computed and represented (3–11). However, these studies do not shed light on how this representation supports higher-level decisions about strategy. Conversely, although various models have been developed to explain revisions of strategy in dynamic environments (12–21), these models rarely explore the form of interactions with lower-level decision processes.

We developed a novel task and computational framework to understand how interactions across a hierarchy of decision processes support adaptive regulation of behavior in a dynamically changing environment. Subjects made decisions about the net direction of a random dot motion stimulus, and the environment determined the subset of eye movement targets that they should use to report their choice. The environment was not cued, and it changed without any warning after several trials, requiring subjects to determine when their decision strategy should switch from persisting in the old environment to exploring a new one. Subjects' environment choices revealed a long-term influence of both outcomes and confidence of their perceptual decisions. These observations motivated a computational framework that simultaneously explains both lower- and higher-level choices based on three key principles: (*i*) lower-level choices are based on integration of sensory evidence within trials, (*ii*) lower-level choices are associated with a subjective confidence that reflects the expected likelihood of success, and (*iii*) higher-level choices are based on integration of outcomes and choice confidence across multiple trials. We show that these three principles can be

Significance

Decisions are guided by available information and strategies that link information to action. Following a bad outcome, two potential sources of error—flawed strategy and poor information—must be distinguished to improve future performance. In a direction discrimination task where subjects decide by accumulating sensory evidence to a bound, we show that humans disambiguate sources of error by integrating expected accuracy and outcome over multiple choices. The strategy switches when the integral reaches a threshold. A hierarchy of decision processes in which lower levels integrate sensory evidence over short timescales, and higher levels interact with lower levels over longer timescales, quantitatively explains the behavior. Expected accuracy links these two levels and enables adaptive changes of decision strategy.

Author contributions: B.A.P. and R.K. designed research; B.A.P. and R.K. performed research; B.A.P. analyzed data; and B.A.P. and R.K. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. Email: roozbeh@nyu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1524685113/-/DCSupplemental.

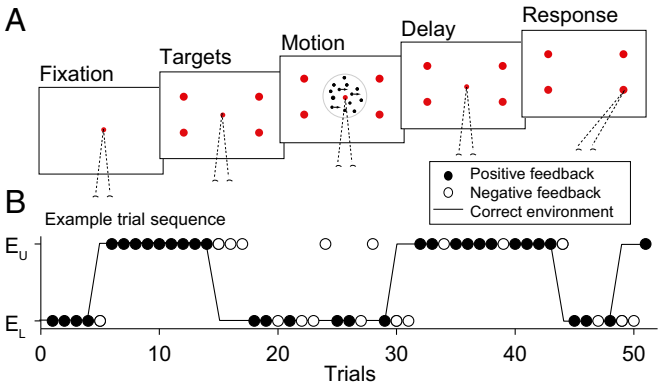

Fig. 1. Changing environment task. (*A*) Task design. The pairs of targets above and below the FP represented two environments. The right and left targets in each environment represented the two possible directions of motion. Subjects received positive feedback for choosing the target that corresponded to both the correct environment and correct motion direction. The motion direction, motion strength (percentage of coherently moving dots, %Coh), and duration varied randomly from trial to trial. The rewarding environment stayed fixed for a variable number of trials (2–15, truncated geometric distribution) and then changed without explicit cue. Subjects had to discover the correct environment based on the history of feedback, choice, and choice certainty. (*B*) Example sequence of trials from one experimental session. On each trial, the subject chose a target in the upper (E_U) or lower (E_L) environment (circles). They received positive feedback (filled circles) if the chosen target matched both the correct environment (black line) and motion direction, and negative feedback (open circles) if either was incorrect.

understood as a neurally plausible implementation of the Bayes optimal solution to the task. The framework demonstrates that adaptive behavior in dynamic environments can be understood as a hierarchy in which both lower- and higher-level decision processes integrate information over distinctly different timescales, and choice confidence is a key connection across these levels. We use our task and framework to establish key properties of these integration processes and shed light on mechanisms of adaptive decision-making.

Results

Six human subjects performed a task in which they adapted their decision strategy in response to unpredictable changes in the environment. In this "changing environment" task (Fig. 1*A*), subjects viewed a patch of stochastic moving dots (22) and reported the net motion direction (right or left) with a saccadic eye movement to a corresponding peripheral target. However, unlike conventional direction discrimination tasks, subjects were provided with two rightward and two leftward targets. The targets were arranged in two right–left pairs above and below the dot patch and represented two distinct environments. On each trial, only one environment was correct. The correct environment stayed fixed for several trials and then changed without an explicit signal to subjects (Fig. 1*B*). In addition to changes in the environment, we controlled the difficulty of the motion direction discrimination by randomly varying motion strength and duration across trials. Subjects received positive feedback only if their chosen target corresponded to both the correct motion direction and the correct environment. Negative feedback, however, could arise from choosing the wrong environment or direction target. Therefore, to determine when to shift their decision strategy from persisting in the old environment to exploring the new one, subjects needed to resolve ambiguous negative feedbacks based on the history of the experienced sensory evidence, choices, and feedbacks.

Before engaging in the changing environment task, subjects were introduced to a simple direction discrimination task with two targets that corresponded to the motion directions (22). Motion direction discrimination training continued until subjects achieved a high level of performance as indicated by low psychophysical thresholds (<17.0% for all subjects, pooled threshold = 13.1 ± 1.45%). This training extended to the changing environment task. Subjects maintained a high level of direction discrimination accuracy and low psychophysical thresholds for motion direction choices, irrespective of the reported environment (pooled threshold = 13.3 ± 0.24%). Similarly, all subjects exhibited improved motion choice accuracy for higher motion strength (Fig. 2*A* and Fig. S1*A*; Eq. **1**, $\beta_1 = 10.1 \pm 0.26$, $P < 10^{-10}$) and duration (Fig. S2*A*; Eq. **1**, $\beta_2 = 0.4 \pm 0.09$, $P = 3.8 \times 10^{-7}$), consistent with previous studies (22, 23). Thus, a subject's ability to perform the direction discrimination was not compromised by the increased complexity of the changing environment task.

A crucial feature of the task design is that it explicitly dissociates choices about motion direction ("direction choices," left versus right choice targets) and choices about the environment ("environment choices," upper versus lower choice targets). This dissociation enabled us to directly measure when subjects switched environments and to assess the factors that shaped subjects' decisions. Below, we report experimental results that elucidate those factors. Then, we explore the underlying computational mechanisms and provide a model that offers a quantitative explanation for the motion direction and environment choices based on within-trial accumulation of sensory evidence and across-trial integration of expected accuracy and feedback.

Environment Choices Were Shaped by Integration of Feedback and Uncertainty About Motion Direction Across Trials. Subjects rarely switched environments following positive feedback [$P(switch) = 0.005$; Fig. 2*B* and Fig. S1*B*], indicating that they understood the relative stability of the environments. In contrast, subjects switched environments frequently following negative feedback [$P(switch) = 0.39$], and more so when negative feedback was given on trials with higher motion strength (Fig. 2*B* and Fig. S1*B*; Eq. **2**, $\beta_1 = 6.2 \pm 0.23$, $P < 10^{-10}$) and duration (Fig. S2*B*; Eq. **2**, $\beta_2 = 0.3 \pm 0.15$, $P = 0.03$), that is, the trials in which they were more likely to have accurate direction responses (Fig. 2*A* and Figs. S1*A* and S2*A*). Indeed, feedback and expected direction choice accuracy seemed to be the critical factors in determining whether subjects switched. The probability of switching environments after negative feedback increased monotonically with subjects' accuracy (Fig. S3), and different combinations of motion strength and duration that produced the same expected accuracy also produced a similar probability of switching. In fact, the expected accuracy on a trial with negative feedback explained 90.7% of the variance in

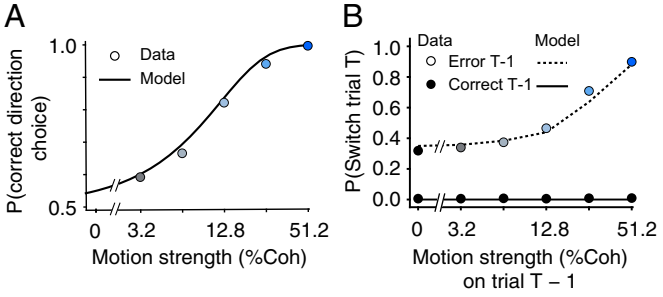

Fig. 2. The motion stimulus of the current trial informed direction choices, and feedback and expected accuracy of previous trials informed environment choices. (*A*) Motion direction discrimination accuracy increased with motion strength. Data points show the accuracy of direction choices disregarding environment choices. (*B*) The proportion of environment switches increased following negative feedback on trials with stronger motion (colored points) and was consistently low following positive feedback (black points). The circles in both panels are data, and the lines show model fits. Data and model fits in both panels are pooled across subjects (see Fig. S1 for individual subjects, and see Table S1 for parameter values). Error bars are SE.

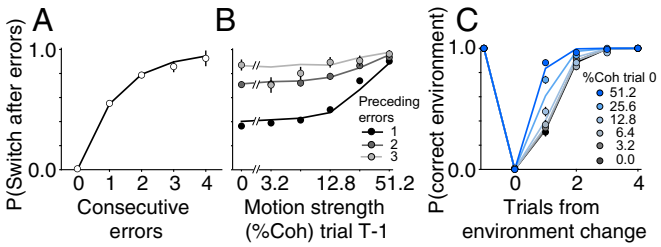

Fig. 3. Environment choices were shaped by integration of feedback and expected motion direction accuracy across multiple trials. (*A*) Consecutive negative feedbacks increased the probability of switching environment choices. In all panels, lines show model fits and circles show data points pooled across subjects. (*B*) The probability of switching increased with motion strength on the previous trial. Different shades of gray show the number of preceding consecutive errors. (*C*) Subjects recognized environment changes faster when they received negative feedback with higher expected direction choice accuracy. Data points show the proportion of correct environment choices as a function of the number of trials relative to an uncued environment change. Trials are divided by motion strength (%Coh) on the change trial (trial = 0). Error bars are SE. See Fig. S4 for data and fits from individual subjects.

switch probabilities on the next trial (Eq. **9**, $R^2 = 0.907$), whereas additional knowledge about the motion strength and duration explained only an additional 1% (Eq. **10**, $R^2 = 0.917$). Thus, subjects' environment switches seemed to be primarily informed by the feedback and expected direction choice accuracy.

The effect of feedback and motion strength on future environment switches extended for multiple trials. Because of subjects' uncertainty about the correct motion direction, they did not always switch environment choices immediately after one negative feedback. When the environment changed, subjects frequently continued to choose the previous (incorrect) environment for two to four trials (43.9% of all environment changes). However, subjects were also more likely to switch as the number of consecutive negative feedbacks mounted (Fig. 3*A* and Fig. S4*A*; Eq. **3**, $\beta_3 = 1.5 \pm 0.06$, $P < 10^{-10}$), suggesting that the effect of negative feedback lasted for multiple trials (13, 24). Importantly, this persistence was also dependent on motion strength, ruling out the possibility that subjects simply counted the number of errors to decide when to switch. The presence of a trial with low motion strength in the sequence of negative feedbacks reduced the likelihood of switching both on the next trial (Fig. 3*B* and Fig. S4*B*; Eq. **3**, $\beta_1 = 5.9 \pm 0.26$, $P < 10^{-10}$) and on the subsequent ones (Fig. 3*C* and Fig. S4*C*; Eq. **5**, $P < 10^{-6}$ for β_{1-3}). Conversely, negative feedback on trials with higher motion strength was more likely to trigger a switch in subjects' environment choice and terminate the sequence of consecutive errors (Fig. 3*C* and Fig. S4*C*). The decision to switch environment choices, therefore, depended on integration of feedback and expected direction choice accuracy across multiple trials.

For a more formal test of the properties of the multitrial integration process, we focused on sequences of two consecutive errors in the same environment, because they were the most frequent type of consecutive errors (80.6%). We asked how the motion strength on those trials influenced subjects' decisions to switch or persist on the subsequent trial. We found that the probability of switching was significantly influenced by the motion strength of both the first (Eq. **6**, $\beta_2 = 1.9 \pm 0.76$, $P = 0.01$) and second ($\beta_1 = 4.9 \pm 0.58$, $P < 10^{-10}$) negative feedback. In addition, we found that the motion strength of the more recent negative feedback exerted a stronger influence on the probability of switching (Eq. **7**, $\beta_2 = -1.8 \pm 0.37$, $P = 1.5 \times 10^{-6}$). The stronger influence of more recent trials could be indicative of two possible mechanisms. Switch evidence may be leaky (22, 25, 26), in which case newer information more strongly influences the decision to switch or stay. Alternatively, subjects could be switching environment choices after accumulated evidence reaches a bound, in which case the latest samples of evidence are more likely to exceed the bound, if bound crossing has not occurred thus far. We will evaluate these possibilities in the following sections.

An Uncertainty Accumulation Model Explained Motion Direction and Environment Choices. We developed a computational framework to understand how expected direction choice accuracy and feedback support adaptive changes in the environment choice (Fig. 4). The model is based on the following three key principles: (*i*) Direction choices result from the accumulation of sensory evidence within trials (10, 22, 26, 27), (*ii*) subjects compute expected accuracy of their direction choices (3–11), and (*iii*) environment choices (to switch or not) result from the integration of expected direction choice accuracy and feedback across trials (14–17, 20, 21). Thus, the model provides a unified framework to explain perceptual decisions, the confidence associated with those decisions, and the mechanisms by which confidence supports adaptive behavior.

We modeled subjects' direction choices using a bounded accumulation model. Integration of sensory evidence toward a decision bound (Fig. 4*A*) (22) accurately explains choices, response times, confidence, and several other aspects of decision-making, including speed–accuracy tradeoff, across a broad range of perceptual tasks (3, 4, 22, 27–32). In addition, neurophysiological recordings from parietal cortex, frontal cortex, basal ganglia, and superior colliculus of animals engaged in basic motion direction discrimination tasks exhibit dynamics consistent with integration of evidence over time to a bound (26, 30–35). We used a simplified variant of the bounded accumulation model known as the drift–diffusion model to account for direction choices. In this model, noisy sensory evidence is integrated over time in a domain bounded by two absorbing decision thresholds that represent the two choices. The direction choice is determined when the accumulated sensory evidence (the "sensory decision variable") reaches one of the two thresholds. If a threshold is not reached by the end of the motion stimulus, then the sign of the sensory decision variable dictates the choice (22, 36, 37).

The same bounded accumulation model can also explain the confidence associated with direction choices (3, 4, 38, 39). The crucial point is that both the magnitude of accumulated evidence and elapsed time provide information about the probability of being correct. To illustrate this mapping, Fig. 4*B* shows the probability distribution of accumulated sensory evidence at each possible decision time for a rightward stimulus with a particular motion strength (6.4% coherence). By applying the decision rule described above, we can compute the probability that the decision variable at a particular magnitude and time would result in a correct response given the full set of stimuli experienced by the subject (Fig. 4*C*). By learning this mapping through experience, subjects could estimate their expected direction choice accuracy based directly on accumulated sensory evidence and elapsed time.

This expected direction choice accuracy can guide environment choices. From a normative perspective, each instance of negative feedback provides some evidence that the current environment has changed ("switch evidence," *SI Text*). The cause of negative feedback is ambiguous, and the magnitude of switch evidence for a single negative feedback is a function of the expected direction choice accuracy for that trial (Fig. 4*D*). Subjects would maximize the accuracy of environment choices by switching when the posterior probability of a new environment exceeds that of the current environment given the history of feedback, expected direction choice accuracy, and belief about the probability of an environment change given that one has not yet occurred (i.e., subjective hazard rate). Such a Bayes optimal solution can be formulated as an accumulation of switch evidence over trials to a bound (*SI Text*). The optimal switch evidence is in units of log-likelihood ratio of a negative feedback under the two environments (Fig. 4*D*): $\log[p(F^-|E_n, C, \tau)/p(F^-|E_o, C, \tau)] = \log[1/(1-A)]$ (Eq. **S6**), where A is the expected direction choice accuracy derived from the sensory decision variable and elapsed time, F^- is negative

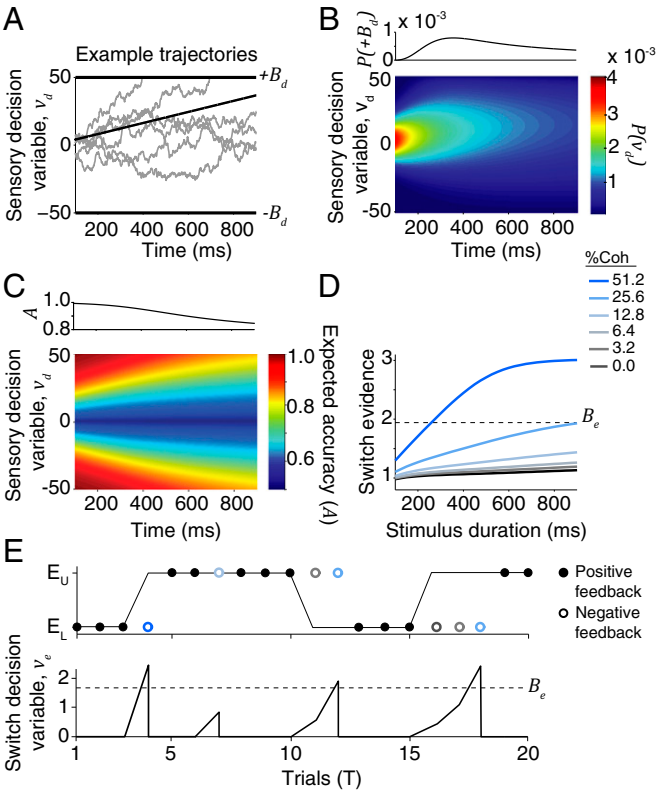

Fig. 4. The Uncertainty Accumulation model. (*A*) Direction choices result from accumulation of sensory evidence within trials (gray lines are example single trial trajectories, and black line shows mean accumulation rate, $\mu_d = kC$). A direction choice is made when accumulated evidence (sensory decision variable, v_d) reaches a bound ($\pm B_d$) or by the sign of v_d when the motion stimulus ends. (*B*) (*Lower*) The probability density of v_d for a rightward motion strength (6.4% coherence). (*Upper*) The probability of reaching the upper (rightward) bound, $P(+B_d)$, over time. (*C*) (*Lower*) Expected direction choice accuracy, *A*, for different v_d and decision time given the decision rule and stimulus set. (*Upper*) Expected accuracy as a function of decision time when the positive bound is crossed. (*D*) Switch evidence of a negative feedback [$log[1/(1-\hat{A})]$; *Materials and Methods*] for different motion strength and duration. Switch evidence grows with motion strength and stimulus duration due to gradual drift of v_d away from 0. The dashed line indicates a fixed switch bound, B_e. (*E*) Example trial sequence and how accumulated switch evidence (switch decision variable, v_e) drives switches in environment choice. (*Upper*) The sequence of environments (lines) and subject's choices (circles) resulting in positive (filled) or negative (open) feedback. Color indicates motion strength. (*Lower*) Changes in v_e across trials. Subjects switch when v_e exceeds the switch bound. For simplicity, we illustrate a fixed B_e (but see text for switch urgency).

feedback, E_n and E_o are the new and current (old) environments, respectively, and C and τ are motion strength and duration, respectively. In other words, just as integration of sensory evidence, in units of log-likelihood ratio of sensory signals, is the optimal computation for two-alternative perceptual choices (28, 40–42), integration of switch evidence over multiple trials of negative feedback to construct a "switch decision variable" is the optimal computation for environment choices.

We hypothesized that subjects approximate the normative computation to decide when to switch environment choices (*Materials and Methods*). We tested this hypothesis by fitting subjects' direction and environment choices simultaneously with a model based on the principles outlined above: integration of sensory evidence within a trial to explain direction choices, computation of direction choice confidence, and, finally, integration of confidence and feedback across trials to a dynamic bound (Fig. 4*E*). The model used only five free parameters (Table S1 and *Materials and Methods*), and, despite its low degrees of freedom, it provided a quantitative explanation for all key aspects of subjects' direction and environment choices (Figs. 2, 3, and 5 and Figs. S1, S2, and S4, lines), including (*i*) changes of direction choice accuracy with motion strength and duration; (*ii*) increased likelihood of switching with negative feedback for stronger and longer motion stimuli; and (*iii*) the long-term, multitrial influence of feedback and motion strength through integration of switch evidence over trials. Altogether, the close match between the model and data strongly suggests that the model captures the computations that guided subjects' behavior.

Accumulation of Switch Evidence Across Trials Is Not Leaky but It Resets After Positive Feedback and Reflects Across-Trial Urgency. Subjects' patterns of environment choices revealed key properties of the switch evidence accumulation across trials. We highlight three of these properties. First, subjects were likelier to switch environment choices after negative feedback when they stayed in an environment for more trials (Fig. 5 *A* and *B*; Eq. **8**, $\beta_3 = 0.3 \pm 0.01$, $P < 10^{-10}$). This increased switch rate was not due to the increased chance of consecutive errors for longer environment durations, because similar results were obtained when we confined the analysis to sequences with only one error (Eq. **8**, $P < 10^{-10}$). Instead, it likely reflects a growing urgency to switch environments. This growing urgency to switch environment choices is akin to the urgency to respond observed in perceptual decision-making tasks (29, 30), except that it happens at much longer timescales (over trials, not within single trials).

Our model provided further support for this urgency signal and its necessity to explain behavior. The optimal form of the switch bound, B_e, should collapse over trials as a function of the subjective hazard rate and the number of consecutive negative feedbacks (Eq. S6). A lower switch bound promotes switches with less accumulated evidence, increasing the likelihood of switches over trials. Our

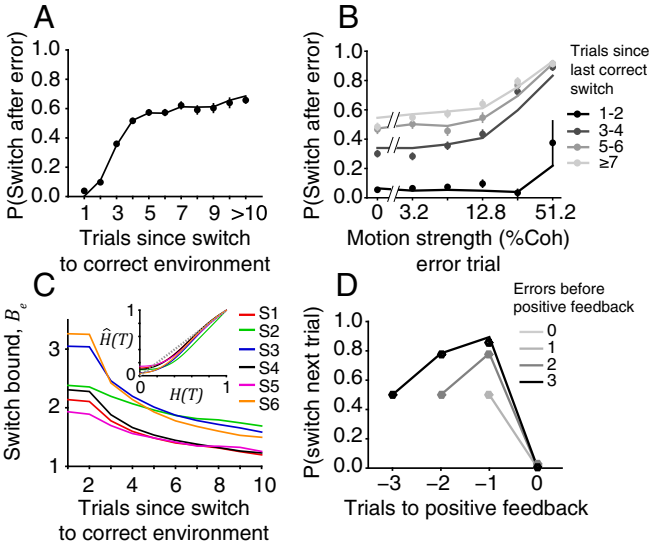

Fig. 5. Switch evidence reflects across-trial urgency and resets after positive feedback. (*A*) The proportion of environment switches after negative feedback increased as a function of the number of trials since the last correct switch. In all panels, circles are data and lines are model fits. (*B*) The probability of switching after negative feedback increased with motion strength and the number of trials in the current environment. (*C*) Mean switch bound resulting from the best-fitting probability weighting functions (*Inset*) relating the experienced hazard rate, $H(T)$, to subjective hazard rate, $\hat{H}(T)$ (*Materials and Methods*). Color indicates different subjects. (*D*) The probability of switching increased with consecutive errors, but dropped to almost 0 after just one positive feedback (trial 0). Switch probabilities before the positive feedback were calculated for an increasing number of consecutive errors within each sequence. Error bars are SE.

model implemented this bound collapse based on two assumptions. First, subjects could estimate the hazard rate of environment changes based on the negative feedbacks they experienced in the task (*Materials and Methods*). Second, subjective hazard rates were related to the experienced hazard rates based on a probability weighting function (43) (Eq. **16** and Fig. 5*C*, *Inset*) that slightly distorted subjective probabilities, as reported previously (44). The best fitting probability weighting function and the optimal switch bound based on the subjective hazard functions are shown in Fig. 5*C*. For all subjects, the quality of fit was consistently worse when this collapsing switch bound was replaced with the best-fitting static bound for the data (likelihood ratio test, $P < 0.01$ for all subjects). Therefore, the integration mechanism that underlies environment switches seems to be susceptible to evidence-independent urgency signals that modulate the termination criterion for the switch decisions. The dynamics of the urgency signal could accommodate various statistics of environment duration, for example, larger urgency for more short-lived environments, providing a basis for adaptive adjustment of behavior (Fig. S5). We revisit this point in *Discussion*.

The optimal form of switch bound predicts that the urgency signal can be divided into two components: one that depends on the subjective hazard rate for an initial negative feedback and a second that increases with subsequent negative feedbacks (Eq. **S6**). The preceding analyses establish the necessity of the first urgency component to explain increased switching with time spent in an environment. To test the necessity of the second form of urgency, we added a weight, ω, on the magnitude of bound collapse with additional negative feedbacks and fitted the value as a free parameter (Eq. **15**). Three subjects collapsed their bound after subsequent negative feedbacks as evidenced by a significant positive weight (S3: $\omega = 1.8 \pm 0.48$, $P = 5.3 \times 10^{-5}$; S4: $\omega = 5.6 \pm 1.68$, $P = 4.1 \times 10^{-4}$; S5: $\omega = 19.3 \pm 6.07$, $P = 7.3 \times 10^{-4}$). The remaining three subjects showed negligible bound collapse after the initial negative feedback (S1: $\omega = 0.05 \pm 0.17$, $P = 0.38$; S2: $\omega = 0.004 \pm 0.007$, $P = 0.30$; S6: $\omega = 0.03 \pm 0.17$, $P = 0.43$). The short environment durations in our experiment promoted switching after few consecutive negative feedbacks, and may have reduced the cost of ignoring this urgency component.

Second, the data suggest a reset of accumulated switch evidence (v_e) to zero after positive feedbacks. Although repeated negative feedbacks increased the subject's likelihood of switching, a single positive feedback immediately dropped the likelihood of switching to almost zero regardless of the number of preceding errors and the magnitude of accumulated switch evidence (Fig. 5*D*; Eq. **4**, $\beta_1 = 0.5 \pm 0.69$, $P = 0.49$), indicating that the switch evidence accumulated before the positive feedback is entirely eliminated. Further support for this conclusion comes from our model, where we allowed the change of v_e after positive feedback to be a free parameter (q). This extended model provides the possibility that positive feedback is treated merely as partial evidence against an environment change. However, for all subjects (6/6), the model fits indicated that the reduction of switch evidence after positive feedback significantly exceeded the maximal switch bound (bootstrap, $P < 0.001$ for all subjects), large enough to enforce a complete reset in accumulated switch evidence. This reset is appropriate for our task because the probability of a change is always minimal immediately after a positive feedback.

Lastly, the integration of switch evidence is unsusceptible to leakage—passive decay—across trials. A leaky integration hypothesis has been suggested previously (45) and is widely assumed to account for sequential learning phenomena in the reinforcement learning literature. The long timescale for the integration of switch evidence and potential biophysical limitations of integration circuits make leaky integration a plausible hypothesis. Therefore, we extended the model to include a free parameter for the leakage of v_e (λ, Eq. **14**). Like the model above, we also allowed the change of v_e following positive feedback to be a free parameter to ensure that estimation of the leakage parameter is not disrupted by forced resets of accumulated switch evidence. The model did not support the leaky integration hypothesis. For all but one subject (5/6), the value of leakage was indistinguishable from zero (S1: $\lambda = 0.0880 \pm 0.2148$, $P = 0.34$; S2: $\lambda = 0.0698 \pm 0.1462$, $P = 0.32$; S3: $\lambda = 0.0095 \pm 0.0114$, $P = 0.20$; S4: $\lambda = 0.0010 \pm 0.0035$, $P = 0.39$; S5: $\lambda = 0.0000 \pm 0.0005$, $P = 0.50$; S6: $\lambda = 0.1582 \pm 0.0253$, $P = 1.9 \times 10^{-10}$), indicating that leakage of switch evidence after negative feedback is not necessary to explain subjects' behavior. Further, eliminating switch noise from the model produced significantly worse fits for all subjects, even when leakage was included (likelihood ratio test, all $P < 10^{-10}$), indicating that leakage is not a replacement for switch noise. Finally, compatible with the previous models, the v_e change after positive feedback was greater than the maximum switch bound for all subjects (bootstrap, all $P < 0.001$), indicating complete evidence resets even when leaky integration was allowed in the model.

Altogether, these results confirm that adaptive decision-making in the dynamic environment of our task depended both on a growing urgency to switch environment choices over trials and on perfect resetting of accumulated evidence following a single positive feedback. In contrast, leakage of switch evidence across trials was minimal and did not play a major role in shaping behavior.

Discussion

Our task is a simplified instance of the hierarchical decisions commonly made in complex environments. To obtain one's goals, one must adopt an appropriate decision strategy and also make wise choices using that strategy. Failing to detect changes in the environment or adopt the right strategy for a new environment is a major source of error in natural settings. Identifying such errors is often nontrivial, because changes in the environment are rarely cued. Rather, decision makers must infer the changes, often from feedback for their own past choices. Inferring the changes creates a hierarchical multiscale decision-making process in which outcomes of lower-level choices inform revisions of decision strategy at higher levels. Our task makes such hierarchical decision processes accessible in a well-controlled experimental setting. By doing so, it enables us to study neural mechanisms that underlie (*i*) resolution of ambiguous feedback (e.g., perceptual or environment errors), (*ii*) interactions of lower- and higher-level decision processes, (*iii*) simultaneous integration of evidence over multiple timescales, and (*iv*) commitment to a new decision strategy.

Detailed analysis of subjects' behavior demonstrated that expected accuracy in our perceptual choices resolves ambiguity about negative feedback by providing evidence that the environment has changed (switch evidence). Each negative feedback represents a sample of switch evidence that is weighted by expected accuracy. Negative feedback conferred stronger evidence for a switch when expected accuracy was high, and less evidence for a switch when expected accuracy was low. We found that the optimal solution to the task was to accumulate switch evidence over multiple trials (in units of log-likelihood ratio) and commit to a new environment when the accumulated switch evidence reached a bound that collapses dynamically over trials (40, 41). A computational framework based on these principles quantitatively explained all aspects of the behavior, providing a plausible mechanism for hierarchical, multiscale decision-making.

The observation that expected accuracy of the perceptual choice contributes to computation of switch evidence sheds light on why confidence is so prevalently computed and accompanies our choices. In a hierarchical decision process, a choice is not merely a commitment at a particular point in time; it is also part of a sequence that feeds into a higher level of the decision hierarchy for choices about strategy. Confidence is the subjective belief, before feedback, that a decision is correct (4, 27, 46–48). The match between this subjective expectation and the actual feedback can be used for learning about the environment (48) by serving as input for decisions about updating the current strategy. Several other functions have been attributed to the computation of confidence, including optimal

cue combination (49), arbitration among multiple systems that compete for a behavioral choice (50), and guidance of sequential decisions when immediate feedback is unavailable (3). We suggest that confidence is also the critical link that connects different levels of the decision hierarchy. Automatic computation of confidence, even in experimental settings that do not demand it (14, 51, 52), suggests that decision hierarchies are an indispensible and ingrained component of our behavioral repertoire.

In our model, subjective expected accuracy was derived directly from accumulated sensory evidence and elapsed time. This framework can explain choices, reaction times, and certainty judgments during motion discrimination tasks (4), and it predicts the dynamics of parietal neurons (3). However, it is likely that alternative models of confidence based on the state of the perceptual decision-making process would also be successful in our framework (6–11), so long as they explain the variation of confidence with motion strength and duration. The key point is that the perceptual decision process drives computation of choice confidence that, in turn, drives decisions about strategy, establishing a mechanistic link across levels of the decision-making hierarchy.

This mechanistic link stems from the utility of confidence for disentangling two potential sources of error—flawed strategy or poor information. We directly tested the role of confidence in a follow-up study in which subjects reported confidence in their motion direction choice during the changing environment task (*Materials and Methods* and *SI Text*). In this experiment, a single saccadic eye movement simultaneously indicated the environment and motion direction choices together with the confidence associated with the direction choice (Fig. S6*A*) (4). As predicted by our model and results of the main task, subjects were more likely to switch environments following trials in which a choice associated with higher confidence produced negative feedback (Fig. S6*B*). Importantly, the effect of confidence on switch behavior was not explained away by physical characteristics of the motion stimulus (coherence and duration; Fig. S6*C*), demonstrating that subjective confidence, and not objective stimulus strength, are key to interpreting negative feedback.

More elaborate versions of our follow-up study with direct measurements of both the motion direction and environment confidence have the potential to shed light on another aspect of the model. To explain subjects' behavior, our model requires a term for switch noise. This noise reflects two quantities that we cannot separate in the current experiment: (*i*) fluctuations in subjective expected accuracy and (*ii*) potential noise in integration of switch evidence. Direct measurement of confidence will remove the first source of variability, enabling us to better characterize integration of switch evidence across trials. Recall that, due to the absence of direct confidence measurements, we had to use an estimate of expected accuracy generated by marginalization over different decision times and sensory decision variables compatible with the subject's direction choice on each trial (*Materials and Methods*). We suspect that a substantial part of switch noise is due to the difference between subjective expected accuracy and the marginalized values we plugged into the model. Therefore, we predict that, by removing this measurement noise, future work will demonstrate more accurate across-trial integration than shown here and will provide even better quantitative fits for the switch behavior.

The breadth of our framework allows it to connect with a broad number of existing models for perception, learning and decision-making, but, also, several critical aspects of our study distinguish it from previous studies. We briefly mention three commonly used classes of model in this paragraph. Model-free reinforcement learning can use choice certainty to improve perception and categorization in a stable environment (53–55). These powerful models, however, say little about how subjects decide that the environment statistics have changed. Our framework also connects with a broad class of hierarchical control models (2, 56), but many of those models lack the clear bridge between perceptual decision-making and decisions about changes in strategy that our model provides. The hierarchical control models thus far have focused on a different, but equally important, aspect of guiding behavior: how task complexity can be reduced by grouping sequences of actions related to a common goal (i.e., temporal abstraction) (2). Combining temporal abstraction with our framework for adaptive hierarchical decisions will be a fruitful endeavor. The third class of models that should be mentioned here is the predictive coding framework, which is also hierarchical in structure (57). These models have recently been extended to perceptual decision tasks (58) by assuming that the precision or reliability of sensory encoding influences the weighting of sensory evidence for decisions. However, the predictive coding framework has not yet been extended to decisions about when to revise a strategy. Unlike standard predictive coding, a prediction generated at a higher level in our framework is not used to explain away lower-level representations, but is used to guide the lower-level decisions. Our framework goes beyond existing models by explaining how adaptive behavior in dynamic environments depends on a specific form of interactions between lower- and higher-level decision processes. Our model quantitatively explains details of behavior with remarkable accuracy, connects to the normative solution to the task, and is built upon neurally plausible mechanisms that can be directly tested through neurophysiological experiments.

By accumulating switch evidence to a bound, our model establishes a simple termination rule by which an evolving belief about the environment can be translated into a concrete decision strategy. This approach also distinguishes our framework from a broad family of learning models based on delta update rules (13–15, 18, 45, 59). These models explain learning through sequential updates in a probabilistic belief about the current environment. In real-world environments, however, it is often necessary to explicitly commit to a strategy to effectively guide future choices (12, 13, 16, 60), for example, when alternative strategies are incompatible. Bounded accumulation of switch evidence offers a powerful method to select among alternative hypotheses about the true state of the world before committing to a strategy. This approach quantitatively captured subjects' switching behavior in our task, and similar mechanisms have been shown to explain switching behavior when environment statistics change along a continuum (16).

The success of this approach suggests that the brain uses the same bounded accumulation mechanism over different timescales to carry out both perceptual decisions and higher-level decisions about strategy (59). Neural mechanisms of accumulation of sensory evidence for perceptual choices are relatively well understood (36), but far less is known about the neural mechanisms underlying integration of switch evidence over multiple trials. Neural responses in parietal and prefrontal cortexes are influenced by past rewards (24) and modulate their responses when subjects shift their decision strategy (13, 59, 60). Neural responses in the medial prefrontal cortex also exhibit peri-saccadic bursts that reach a fixed firing rate peak immediately before switches in a dynamic foraging task (24), suggestive of a mechanism similar to a switch bound. Human imaging data also support the role of prefrontal and parietal cortexes in updating belief about a changing environment (12, 17). Our task provides a framework to study whether and how these areas could support long-term integration of switch evidence, perhaps via interactions with identified neural representations of perceptual confidence (3, 5, 6).

Our model also revealed several fundamental properties of the higher-level decision process. Leakage or gradual loss of accumulated switch evidence was negligible for the timescales tested in this experiment and unnecessary to explain environment switches. We cannot rule out more complex scenarios that involve leak rates that vary with task parameters. Variable leakage can be advantageous in some scenarios (25, 37, 61), but perfect integration over consecutive negative feedbacks is optimal for this task (*SI Text*). Also predicted by the normative solution, we found that accumulated switch evidence resets after positive feedback, consistent with abrupt behavioral and neural changes observed in other learning tasks (19, 60). Further,

subjects showed a higher propensity to switch the longer they stayed in an environment, indicating a gradual drop in their switch bound (i.e., urgency), most likely due to a growing subjective hazard rate or the prior odds that the environment has changed (16, 21). The optimal form of switch bound incorporates this growing subjective hazard rate and provided an excellent account of behavior, suggesting a normative basis for this growing urgency signal. Changes of urgency can augment the behavioral flexibility achieved by the static shifts of the switch bound. These static and dynamic changes of switch bound enable adjustment of switch rate based on the volatility of the environment, which can be learned through experience (14, 16, 61). Indeed, increasing the average duration of the environment in our experiment reduced the switch rate largely by increasing the switch bound—subjects accumulated more evidence before a switch (Fig. S5).

To summarize, we showed how expected accuracy in perceptual choices disambiguates negative feedback and bridges levels of the decision-making hierarchy by furnishing evidence for changes of strategy. Both perceptual and higher-level decision processes use similar bounded accumulation mechanisms that operate concurrently at different timescales. We showed how this framework uses neurally plausible mechanisms to implement the optimal solution to the task. Our task is simple enough to be performed by nonhuman primates, laying the groundwork for critical experiments to determine the neural implementation of these mechanisms.

Materials and Methods

Six human subjects (five male and one female) participated in the main experiment. Observers had normal or corrected-to-normal vision. All subjects were naïve to the purpose of the experiment and provided informed written consent before participation. All procedures were approved by the Institutional Review Board at New York University.

Behavioral Tasks. Here, we summarize the behavioral tasks; details are provided in *SI Text*. Subjects were first trained to perform a direction discrimination task. Subjects initiated a trial by shifting gaze to a central fixation point (FP). After a short delay, two targets appeared on opposite sides of the screen, followed by a random dot motion stimulus. The subjects' task was to determine the net direction of motion (left or right). The percentage of coherently moving dots (motion strength) and the duration of stimulus presentation varied from trial to trial and determined the difficulty of the motion direction discrimination. After a second short delay, FP turned off, signaling subjects to report the perceived direction of motion by shifting gaze to the left or right choice target. Distinct auditory tones delivered positive or negative feedback if the choice was correct or wrong.

Subjects were introduced to the changing environment task (Fig. 1*A*) following motion direction discrimination training. The experimental setup, motion stimulus, and timing of events were unchanged from training. However, instead of one pair of choice targets, subjects were presented with two pairs of choice targets, one pair above and one pair below the FP (four total), corresponding to the two environments. The right and left targets in each environment represented the two possible motion directions. Subjects received positive feedback for choosing the target that corresponded to both the correct environment and the correct motion direction. We refer to the choice of left versus right targets as the "direction choice" and the choice of upper versus lower targets as the "environment choice." The active environment stayed fixed for a variable number of trials (for the main experiment, 2–15 trials, mean = 6, truncated geometric distribution) and then changed without explicit cue (Fig. 1*B*).

We conducted two follow-up experiments that further tested the mechanisms underlying revisions of decision strategy. In the first experiment, the active environment persisted longer (3–20 trials, mean = 10) to test the influence of changes in environment stability on behavior. In the second follow-up experiment, subjects simultaneously reported their direction choice confidence along with their direction and environment choices using a single saccadic eye movement to an elongated bar (4). See *SI Text* for details.

Behavioral Analyses. We assessed the effects of motion strength and duration on direction choices independent of environment choices using the following logistic regression:

$$\text{Logit}\left[P_T(\textit{correct dir})\right] = \beta_1 C_T + \beta_2 \tau_T, \quad [1]$$

where $\text{Logit}(p) = log(\frac{p}{1-p})$, and $P_T(\textit{correct dir})$ is the probability of a correct motion direction choice on trial T. C_T and τ_T are the motion strength and duration on the same trial, respectively. The β_i are regression coefficients. β_1 tests for a main effect of motion strength on the proportion of correct motion direction choices, β_2 tests for a main effect of stimulus duration. Regression coefficients in Eq. **1** and all subsequent logistic regressions were calculated using maximum likelihood fitting and are summarized in Table S2. In Eq. **1** and other logistic regressions in this paper, the probabilities on the left-hand side of the equations are conditional on the factors listed on the right-hand side. For simplicity and to keep the equations short, we do not list these factors in the conditional probabilities.

To quantify the effect of the last trial on the decision to switch environment choices, we used the following logistic regression:

$$\text{Logit}\left[P_{T,F}(\textit{switch})\right] = \beta_0 + \beta_1 C_{T-1} + \beta_2 \tau_{T-1}, \quad [2]$$

where $P_{T,F}(\textit{switch})$ is the probability that the environment choice on trial T does not match the environment choice on trial $T-1$ (i.e., the subject switched environment choices) given positive (F^+) or negative (F^-) feedback on trial $T-1$. C_{T-1} and τ_{T-1} indicate the motion strength and duration on the previous trial, $T-1$. This regression was performed separately for trials in which feedback was positive or negative on trial $T-1$.

We tested for the effect of consecutive negative feedbacks on environment choices using the following equation:

$$\text{Logit}\left[P_{T,F^-}(\textit{switch})\right] = \beta_0 + \beta_1 C_{T-1} + \beta_2 \tau_{T-1} + \beta_3 N, \quad [3]$$

where N indicates the number of consecutive negative feedbacks that preceded trial T . The null hypothesis is that subjects did not take feedback history into account beyond the last trial and that their decisions to switch environment choices were mainly influenced by the last trial ($H_0 : \beta_3 = 0$).

We tested whether the history of negative feedback was negated by a single positive feedback using the following equation:

$$\text{Logit}\left[P_{T,F^+}(\textit{switch})\right] = \beta_0 + \beta_1 K, \quad [4]$$

where K indicates the number of consecutive negative feedbacks followed by a single positive feedback before trial T. β_1 tests whether the influence of repeated negative feedbacks remains following a single positive feedback.

We used several additional analyses to investigate the effects of consecutive errors on environment choices. First, we evaluated how motion strength on the trial in which the environment changed (and the subject received negative feedback) influenced the accuracy of future environment choices. The following logistic regression was used:

$$\text{Logit}[P_{T+i}(\textit{correct env})] = \beta_0 + \sum_{j=1}^{4} \beta_j C_T \delta_{ij}, \quad [5]$$

where $P_{T+i}(\textit{correct env})$ indicates the probability of choosing the correct environment i trials after the environment change on trial T. Here δ_{ij} is a Dirac delta function, which is 1 when i equals j and 0 otherwise. Up to four trials after the environment change are considered for this analysis. Subjects almost always received negative feedback on trial T because they were unaware of the environment change and chose the previous environment. Each β_j tests the hypothesis that the motion strength of the change trial influences environment accuracy j trials into the future.

Second, we analyzed sequences of two consecutive errors to test how the motion strength for each error trial influenced the subsequent environment choice. We used the following regression equation:

$$\text{Logit}\left[P_{T,F^-}(\textit{switch})\right] = \beta_0 + \beta_1 C_{T-1} + \beta_2 C_{T-2} + \beta_3 C_{T-1} C_{T-2}, \quad [6]$$

where C_{T-2} and C_{T-1} are the motion strengths of the first and second trials in the sequence, respectively. The β_1 coefficient tests for a main effect of motion strength on the decision to switch environment choices on the next trial, the β_2 coefficient tests for a main effect of motion strength two trials in the future, and the β_3 coefficient tests for an interaction of the two preceding trials. We focused on sequences of two errors, because of their abundance in the dataset. Similar trends were obtained for longer error sequences.

Third, we used the following equation to test whether the influence of consecutive errors on the decision to switch environment choices depended on the ordering of the trials:

$$\text{Logit}\left[P_{T,F^-}(switch)\right] = \beta_0 + \beta_1(C_{T-2} + C_{T-1}) + \beta_2(C_{T-2} - C_{T-1}). \quad [7]$$

The null hypothesis is no effect of ordering ($H_0: \beta_2 = 0$). If $\beta_2 > 0$, then the probability of switching is greater when the motion strength for the first error (trial $T-2$) is greater than the motion strength for the second error (trial $T-1$). The opposite is true if $\beta_2 < 0$. We obtained identical results when we instead defined a set of contrasts for C_{T-2} and C_{T-1} in Eq. **6**, but we report coefficients from Eq. **7**, for simplicity.

Finally, we used the following regression to assess whether subjects' decisions to switch environment choices were influenced by the number of trials spent in the current environment:

$$\text{Logit}\left[P_{T,F^-}(switch)\right] = \beta_0 + \beta_1 C_{T-1} + \beta_2 \tau_{T-1} + \beta_3 L_{T-1}, \quad [8]$$

where L_{T-1} is the number of trials since the subject switched into the new environment. In the actual task design, the environment duration was sampled from a truncated geometric distribution with a relatively flat hazard rate between trials 3 and 10 and an increasing hazard rate closer to the point of truncation. Subjects could inform their switching behavior by learning the statistics of environment duration [either the hazard rate or the prior odds (16) that the environment has changed since the last switch] through experience. The null hypothesis is that subjects were not influenced by the number of trials in the current environment ($H_0: \beta_3 = 0$). We obtained identical results when we confined the analysis to the trials preceded by only a single error ($T-2$ was correct), indicating that the increased tendency to switch with the experienced duration of the current environment is not explained by the increased likelihood of multiple preceding errors.

We sought further support for the role of expected accuracy in determining environment switches by computing the variance explained (R^2) for two nested logistic regressions. The first regression predicted switching after negative feedback based only on the expected accuracy of the preceding trial,

$$\text{Logit}\left[P_{T,F^-}(switch)\right] = \beta_0 + \beta_1 \log\left(\frac{1}{1 - A_{T-1}}\right). \quad [9]$$

The second regression included additional terms for both motion strength and duration,

$$\text{Logit}\left[P_{T,F^-}(switch)\right] = \beta_0 + \beta_1 \log\left(\frac{1}{1 - A_{T-1}}\right) + \beta_2 C_{T-1} + \beta_3 \tau_{T-1}. \quad [10]$$

Only trials following negative feedback (F^-) were included in this analysis. Variance explained was computed based on a comparison of model fits to observed probability of switches for different motion strengths and durations. If expected accuracy is the primary factor in explaining switching behavior, then it should explain the majority of variance in the probability of switching without the inclusion of the motion strength and duration terms (i.e., the variance explained by the second regression should not greatly exceed the variance explained by the first).

Uncertainty Accumulation Model. Perceptual and environment choices in our task can be explained through three core mechanisms. Direction choices are produced by integrating momentary sensory evidence within trials (22, 26, 27, 34). Direction choice confidence (expected accuracy) is derived from accumulated sensory evidence and elapsed time (3, 4). Environment choices are guided by integrating feedback and expected accuracy across trials (14, 16, 17, 21). We developed a model that combines these core mechanisms to simultaneously explain subjects' direction and environment choices across a session. We further show how this framework implements the Bayes optimal computation to maximize environment choice accuracy given the experienced trial sequence. Below, we outline the details of this model.

Direction choices are produced by accumulating noisy sensory evidence to a threshold (decision bound) (36). A simplified version of this process can be formulated by a drift–diffusion model, which has been shown to successfully explain choice, reaction time, and confidence judgments for a broad range of cognitive and perceptual decision-making tasks, including motion direction discrimination (4, 22, 27–30). According to this model, the decision terminates when accumulated sensory evidence reaches a positive or negative bound or when the incoming sensory evidence stops (i.e., the motion stimulus ceases). The choice is determined by the bound that is reached (upper or lower) or, if a bound is not reached, by the sign of the accumulated sensory evidence (positive or negative) at the end of the stimulus duration.

The accumulated sensory evidence undergoes drift plus diffusion according to the following stochastic differential equation (Fig. 4 *A* and *B*):

$$dv_d = dt\mu_d + \sqrt{dt}\xi_d, \quad v_d(0) = 0, \quad [11]$$

where v_d is the state of the accumulated sensory evidence (i.e., the sensory decision variable for motion direction) (36), t is time in milliseconds, μ_d is the mean of momentary sensory evidence, and ξ_d is a Wiener process with unit SD. The distribution of momentary sensory evidence is stationary over time, and its mean is linearly related to the motion strength; $\mu_d = kC$, where k is a sensitivity parameter and C is motion strength (3); v_d starts at zero on each trial. A sensory decision bound parameter, B_d, defines positive and negative absorbing bounds for the direction choices.

There is a unique mapping between the magnitude of accumulated evidence, decision time, and the probability that the direction choice is correct (3, 4). Given the set of motion strengths in the experiment, one can calculate the expected direction choice accuracy, A, for all possible values of accumulated sensory evidence and decision times (Fig. 4*C*),

$$A = p(D_1|v_d, t_d) = \frac{\sum_i p(v_d, t_d|D_1, C_i)p(C_i)}{\sum_j \sum_i p(v_d, t_d|D_j, C_i)p(C_i)}, \quad [12]$$

where t_d is the decision time, v_d is accumulated evidence at decision time, and D_1 and D_2 are the correct and incorrect motion direction choices, respectively. $P(D_1|v_d, t_d)$ is the probability that the chosen direction will turn out to be correct for a particular sensory decision variable and time. C_i and $p(C_i)$ are the set of motion strengths and their probabilities in the experiment. The summation term on the right-hand numerator implements marginalization over motion strength, and the summation terms in the denominator implement marginalization over motion strength and direction. As shown in Fig. 4*C*, this equation implies that expected accuracy depends on both the magnitude of accumulated sensory evidence (i.e., greater accumulated evidence is associated with greater confidence) and also elapsed time (longer decision times are associated with lower confidence), a relationship that has been experimentally verified (3, 4). By learning this mapping through experience, subjects could gauge the expected accuracy of their direction choices in individual trials based on their accumulated sensory evidence and decision time.

We estimated the expected direction choice accuracy on individual trials, $\hat{A}$, by marginalizing over possible accumulated evidence and decision times associated with each choice and stimulus,

$$\hat{A} = \hat{p}(D_1|C, R, \tau) = \frac{1}{\psi}\int\int p(D_1|v_d, t_d)p(v_d, t_d, |C, R, \tau)g(v_d, t_d, R)dv_d dt_d, \quad [13]$$

where C is the motion strength, τ is the motion duration, R is the direction choice, and ψ is the normalization factor; $g(v_d, t_d, R)$ is an indicator function that implements the decision rule explained above,

$$g(v_d, t_d, R) = \begin{cases} 1 & \text{if } v_d \text{ and } t_d \text{ terminate the process and lead to } R, \\ 0 & \text{otherwise.} \end{cases}$$

The marginalization in Eq. **13** reflects the fact that, in this experiment, subjects could have committed to a direction choice before the Go signal. The expected probability of an erroneous direction choice would be

$$\hat{p}(D_2|C, R, \tau) = 1 - \hat{p}(D_1|C, R, \tau).$$

A key feature of the changing environment task design is that both feedback and expected direction choice accuracy furnish evidence bearing on the decision to switch or repeat the previous environment choice. Positive feedback always minimizes the probability that the environment will change on the next trial. Negative feedback, on the other hand, is always ambiguous, but expected accuracy of the direction choice can resolve this ambiguity. Higher expected accuracy translates to a larger probability that the environment has changed. That offers a principle that subjects must take into account to optimize their environment choices. Hereafter, we use the term "switch evidence" to refer to the combined evidence that feedback and expected accuracy of direction choices provide about the probability that the environment has changed.

A Bayesian decision maker would switch environment when the posterior probability of a new environment exceeds the old one given the history of feedback, expected direction choice accuracy, and trials spent in the old environment (*SI Text*). This Bayes optimal solution can be formulated as integration of switch evidence over trials, where switch evidence is defined as the log-likelihood ratio of an error feedback for the new and old environments: $log[1/(1-A)]$ (Eq. **S6**), where A is the expected direction choice accuracy (Eq. **12**). The intuition for the formulation of switch evidence is as follows. The probability of negative feedback for staying in the old environment following an environment change is 1. However, if the old environment is still effective, the probability of negative feedback for staying in the environment

is $1-A$. The log of the ratio of these two likelihoods constitutes evidence for a change in the environment. Because we do not know the decision time on each trial, we use the expected direction choice accuracy, $\hat{A}$ (Eq. **13**), as a substitute for A. Subjects should switch environments when integrated switch evidence exceeds a bound dictated by the hazard rate and the number of consecutive negative feedbacks (Eq. **S6**). Overall, just as accumulating sensory evidence to a bound is the optimal computation for making a decision based on incoming sensory evidence (28, 40, 41), integrating switch evidence over trials to a bound is the optimal solution to decide when to switch environment in our task.

The optimal model replicates all of the major trends in subjects' behavior (Fig. S7). A quantitative fit to the data, however, requires knowledge about subjective hazard rates and properties of the accumulation process. We developed a series of plausible models to explore these properties. In our models, switch evidence is integrated across trials according to the following nonstationary diffusion process:

$$dv_e = dT\left[\mu_{e,T} - \lambda v_e\right] + \sqrt{dT}\xi_e, \quad v_e(0) = 0, \qquad [14]$$

where, v_e is the accumulated switch evidence (i.e., switch decision variable), T is time in units of trials, $\mu_{e,T}$ is the switch evidence on trial T, ξ_e is a Wiener process with SD σ_e ("switch noise"), and λ is a leakage term that discounts past evidence. The leakage parameter controls the time constant of integration across trials and ranges from 0 (perfect integration) to 1 (no integration). The switch noise reflects fluctuations of subjective expected accuracy and potential noise in the accumulation of switch evidence. Following negative feedback on trial T, μ_e is the switch evidence as explained above. Following positive feedback, μ_e is a negative constant q. Thus, negative feedback increases switch evidence according to the expected accuracy, and positive feedback decreases it. The process is nonstationary because μ_e depends on the specific sequence of feedbacks, motion strengths, motion durations, and choices across trials. The model predicts that a switch in the current environment choice is initiated when a switch bound, B_e, is exceeded. Following the switch, v_e is reset back to zero. The model also assumes a lower reflecting bound at 0 that prevents v_e from becoming negative. The fact that the probability of another environment change can only grow after a correct switch justifies such a lower bound. The exact location of this reflecting lower bound is not critical for our conclusions, and so we did not make it a free parameter in the model.

We explored alternative nested models that made different assumptions about the presence or absence of leakage, noise, and the influence of positive feedback on accumulated switch evidence. For our main model, we fixed $q=-\infty$ to impose a complete reset of switch evidence following positive feedbacks. We also fixed $\lambda=0$, assuming that integration of switch evidence does not suffer from leakage. This formulation is consistent with the optimal solution (Eq. **S6**), but we also evaluated several plausible alternatives. In a second model, we relaxed the constraint on q and allowed it to be a free parameter, to formally test whether a reset of accumulated switch evidence is a warranted assumption. In a third model, we also allowed λ to be a free parameter, to estimate the amount of leak in the integration process. The results of these three models are explained in *Results*. In a fourth model, we tested the necessity of switch noise by forcing it to zero and comparing the fits with the main model. The results verified that switch noise was necessary to explain behavior (likelihood ratio test, $P < 10^{-10}$ for all subjects). Finally, in a fifth model, we forced switch noise to zero and allowed λ and q to be free parameters to assess whether leakage can compensate for switch noise. The fits were generally inferior to our main model or the third model above (likelihood ratio test with the third model, $P < 10^{-10}$ for all subjects).

The switch bound, B_e, is informed by the subjective hazard rate of environment changes and the consecutive negative feedbacks experienced before the current trial (*SI Text*). Because environment durations were sampled from a truncated geometric distribution, the true hazard rate gradually increased as the number of trials in an environment approached the truncation point. Subjects were not told about the distribution of environment duration but could develop a subjective estimate by experience. The increasing hazard rate created an urgency to switch by collapsing B_e over trials. According to the Bayesian optimal solution, consecutive errors would further accelerate this bound collapse. We tested the influence of consecutive errors using a modified version of Eq. **S6**,

$$B_e(T) = -\log\left[\hat{H}(T-n)\right] + \log\left[1-\hat{H}(T-n)\right] + \omega\sum_{i=n-1}^{0}\log\left[1-\hat{H}(T-i)\right], \qquad [15]$$

where $B_e(T)$ is the switch bound and $\hat{H}(T)$ is the subjective hazard rate on trial T; n is the number of consecutive negative feedbacks preceding the negative feedback on trial T (total number of consecutive errors in the sequence is $n+1$). The first and second terms in Eq. **15** establish a positive baseline for the first negative feedback that is modulated by the subsequent negative feedbacks up to the current trial (third term in Eq. **15**). Because $\hat{H}(T)$ is bounded between zero and 1, $\log[1-\hat{H}(T-i)]$ are negative and decrease the bound from the baseline; ω is a weighting parameter that scales the magnitude of the bound collapse due to subsequent negative feedbacks. In our main model, we fix ω to 1 to implement the optimal bound, but we also evaluated alternative models in which ω was a free parameter to test whether modulation by subsequent negative feedbacks is a necessary form of switch urgency (*Results*).

We estimated subjective hazard rates based on experienced environment changes and well-known distortions in perception of objective probabilities. First, we measured the experienced hazard rate, $H(T)$, for the trial sequences in the task. To do so, we calculated the likelihood that the subject received a negative feedback on trial i within an environment and subtracted a baseline likelihood for negative feedback due to motion direction errors. These experienced hazard rates were similar across subjects and matched expectations based on the truncated geometric distribution of environment durations. Second, we allowed for the possibility that subjective hazard rates, $\hat{H}(T)$, may deviate systematically from the experienced hazard rates, because subjects tend to overweight lower probabilities and underweight higher probabilities (44). To account for individual differences in subjective hazard rates, we implemented a probability weighting function following (43)

$$\hat{H}(T) = B_0 + \left(\frac{H(T)^{\gamma}}{\left\{H(T)^{\gamma} + [1-H(T)]^{\gamma}\right\}^{1/\gamma}}\right)(1-B_0), \qquad [16]$$

where B_0 and γ are free parameters that determine the mapping between actual and subjective probabilities. Because the form of $H(T)$ was derived directly from the data and Eq. **15** directly relates $\hat{H}(T)$ to $B_e(T)$, these are the only free parameters needed to describe the switch bound. The form of the optimal bound resulting from the best fitting probability weighting functions are shown in Fig. 5*C*. To test the necessity of the bound collapse, we also evaluated a model in which the bound was static over all trials using a single parameter (*Results*).

Model Fitting. In total, our nested models have between two and seven free parameters, depending on whether leakage (λ), negative switch evidence (q), switch noise (σ_e), and switch bound parameters (ω, B_0, and γ) are fixed or free to change. Motion direction choices depend on the stimulus sensitivity, k, and sensory decision bound, B_d. These parameters, along with q, determine the mean switch evidence, μ_e, on each trial. In addition to μ_e, environment choices depend on a switch noise parameter, σ_e, leakage, λ, and the switch bound parameters, B_0, γ, and ω (Eqs. **15** and **16**). The main model in the paper has five free parameters (k, B_d, σ_e, B_0, and γ; Table S1, Figs. 2, 3, and 5, and Figs. S1, S2, S4, and S5).

For each model, the parameters were simultaneously fit to individual subjects' data by maximizing the joint likelihood of direction choices (correct or error) and environment choices (switch or nonswitch) across trials. We calculated the likelihood of each direction choice using numerical solutions to the Fokker–Planck formulation of Eq. **11** (3). We calculated the likelihood of each environment choice using Monte Carlo simulations (15,000 iterations) to solve Eq. **14**. The exact sequences of motion strengths, motion durations, and choices experienced by the subjects were used for these calculations. Further, to ensure that the model used a trial history that matched the subject's experience, we reset accumulated switch evidence to zero on trials after subjects switched environments. The trial following a switch error was excluded from fitting, because subjects were explicitly told when they incorrectly switched environments (see *Behavioral Tasks* and *SI Text*).

We estimated the SE of best-fitting parameter values using a bootstrap procedure. Typical bootstrapping involves randomly sampling individual trials, but our model predictions depend on trial history. Therefore, we instead sampled, with replacement, consecutive runs of trials between environment switches. This preserves the effective trial history, because evidence accumulation always resets following the switches. The total number of runs sampled was equal to the total number of runs in each data set. We repeated this process 100 times and identified the parameters that maximized the likelihood of the sampled data in each iteration. The SD of the resulting parameter distribution provided an estimate of the SE of model parameters.

ACKNOWLEDGMENTS. We thank Bill Newsome, Mike Shadlen, Josh Gold, and Valerio Mante for useful discussions and Saleh Esteki for assistance with data collection. This study was supported by a Simons Collaboration on the Global Brain postdoctoral fellowship (to B.A.P.) and National Institutes of Health Grant R01 MH109180-01, a Whitehall Foundation research grant, and a NARSAD Young Investigator Award (to R.K.).

1. Logan GD, Gordon RD (2001) Executive control of visual attention in dual-task situations. *Psychol Rev* 108(2):393–434.
2. Botvinick MM, Niv Y, Barto AC (2009) Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective. *Cognition* 113(3):262–280.
3. Kiani R, Shadlen MN (2009) Representation of confidence associated with a decision by neurons in the parietal cortex. *Science* 324(5928):759–764.
4. Kiani R, Corthell L, Shadlen MN (2014) Choice certainty is informed by both evidence and decision time. *Neuron* 84(6):1329–1342.
5. Middlebrooks PG, Sommer MA (2012) Neuronal correlates of metacognition in primate frontal cortex. *Neuron* 75(3):517–530.
6. Kepecs A, Uchida N, Zariwala HA, Mainen ZF (2008) Neural correlates, computation and behavioural impact of decision confidence. *Nature* 455(7210):227–231.
7. Fleming SM, Lau HC (2014) How to measure metacognition. *Front Hum Neurosci* 8: 443.
8. Pleskac TJ, Busemeyer JR (2010) Two-stage dynamic signal detection: A theory of choice, decision time, and confidence. *Psychol Rev* 117(3):864–901.
9. Moran R, Teodorescu AR, Usher M (2015) Post choice information integration as a causal determinant of confidence: Novel data and a computational account. *Cognit Psychol* 78:99–147.
10. Ratcliff R, Starns JJ (2013) Modeling confidence judgments, response times, and multiple choices in decision making: Recognition memory and motion discrimination. *Psychol Rev* 120(3):697–719.
11. Drugowitsch J, Moreno-Bote R, Pouget A (2014) Relation between belief and performance in perceptual decision making. *PLoS One* 9(5):e96511.
12. Donoso M, Collins AG, Koechlin E (2014) Human cognition. Foundations of human reasoning in the prefrontal cortex. *Science* 344(6191):1481–1486.
13. Seo H, Cai X, Donahue CH, Lee D (2014) Neural correlates of strategic reasoning during competitive games. *Science* 346(6207):340–343.
14. Nassar MR, et al. (2012) Rational regulation of learning dynamics by pupil-linked arousal systems. *Nat Neurosci* 15(7):1040–1046.
15. Courville AC, Daw ND (2007) The rat as particle filter. *Adv Neural Inf Process Syst* 20: 369–376.
16. Gallistel CR, Krishan M, Liu Y, Miller R, Latham PE (2014) The perception of probability. *Psychol Rev* 121(1):96–123.
17. Behrens TE, Woolrich MW, Walton ME, Rushworth MF (2007) Learning the value of information in an uncertain world. *Nat Neurosci* 10(9):1214–1221.
18. Yu AJ, Dayan P (2005) Uncertainty, neuromodulation, and attention. *Neuron* 46(4): 681–692.
19. Costa VD, Tran VL, Turchi J, Averbeck BB (2015) Reversal learning and dopamine: A Bayesian perspective. *J Neurosci* 35(6):2407–2416.
20. Meyniel F, Schlunegger D, Dehaene S (2015) The sense of confidence during probabilistic learning: A normative account. *PLOS Comput Biol* 11(6):e1004305.
21. Brown SD, Steyvers M (2009) Detecting and predicting changes. *Cognit Psychol* 58(1): 49–67.
22. Kiani R, Hanks TD, Shadlen MN (2008) Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment. *J Neurosci* 28(12): 3017–3029.
23. Shadlen MN, Newsome WT (2001) Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *J Neurophysiol* 86(4):1916–1936.
24. Hayden BY, Pearson JM, Platt ML (2011) Neuronal basis of sequential foraging decisions in a patchy environment. *Nat Neurosci* 14(7):933–939.
25. Ossmy O, et al. (2013) The timescale of perceptual evidence integration can be adapted to the environment. *Curr Biol* 23(11):981–986.
26. Purcell BA, et al. (2010) Neurally constrained modeling of perceptual decision making. *Psychol Rev* 117(4):1113–1143.
27. Link SW (1992) *The Wave Theory of Difference and Similarity* (Lawrence Erlbaum Assoc, Hillsdale, NJ).
28. Bogacz R, Brown E, Moehlis J, Holmes P, Cohen JD (2006) The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychol Rev* 113(4):700–765.
29. Ditterich J (2006) Evidence for time-variant decision making. *Eur J Neurosci* 24(12): 3628–3641.
30. Churchland AK, Kiani R, Shadlen MN (2008) Decision-making with multiple alternatives. *Nat Neurosci* 11(6):693–702.
31. Purcell BA, Kiani R (2016) Neural mechanisms of post-error adjustments of decision policy in parietal cortex. *Neuron* 89(3):658–671.
32. Hanks TD, et al. (2015) Distinct relationships of parietal and prefrontal cortices to evidence accumulation. *Nature* 520(7546):220–223.
33. Hanks T, Kiani R, Shadlen MN (2014) A neural mechanism of speed-accuracy tradeoff in macaque area LIP. *eLife* 3:3.
34. Ding L, Gold JI (2010) Caudate encodes multiple computations for perceptual decisions. *J Neurosci* 30(47):15747–15759.
35. Ratcliff R, Cherian A, Segraves M (2003) A comparison of macaque behavior and superior colliculus neuronal activity to predictions from models of two-choice decisions. *J Neurophysiol* 90(3):1392–1407.
36. Shadlen MN, Kiani R (2013) Decision making as a window on cognition. *Neuron* 80(3): 791–806.
37. Tsetsos K, Gao J, McClelland JL, Usher M (2012) Using time-varying evidence to test models of decision dynamics: Bounded diffusion vs. the leaky competing accumulator model. *Front Neurosci* 6:79.
38. Yu S, Pleskac TJ, Zeigenfuse MD (2015) Dynamics of postdecisional processing of confidence. *J Exp Psychol Gen* 144(2):489–510.
39. Zylberberg A, Barttfeld P, Sigman M (2012) The construction of confidence in a perceptual decision. *Front Integr Nuerosci* 6:79.
40. Wald A, Wolfowitz J (1948) Optimum character of the sequential probability ratio test. *Ann Math Stat* 19(3):326–339.
41. Gold JI, Shadlen MN (2001) Neural computations that underlie decisions about sensory stimuli. *Trends Cogn Sci* 5(1):10–16.
42. Kira S, Yang T, Shadlen MN (2015) A neural implementation of Wald's sequential probability ratio test. *Neuron* 85(4):861–873.
43. Glaser C, Trommershäuser J, Mamassian P, Maloney LT (2012) Comparison of the distortion of probability information in decision under risk and an equivalent visual task. *Psychol Sci* 23(4):419–426.
44. Kahneman D, Tversky A (1979) Prospect theory: An analysis of decision under risk. *Econometrica* 47(2):263–291.
45. Corrado GS, Sugrue LP, Seung HS, Newsome WT (2005) Linear-nonlinear-Poisson models of primate choice dynamics. *J Exp Anal Behav* 84(3):581–617.
46. Peirce CS, Jastrow J (1885) *On Small Differences of Sensation* (US Gov Print Off, Washington, DC).
47. Henmon VAC (1911) The relation of the time of a judgment to its accuracy. *Psychol Rev* 18(3):186–201.
48. Vickers D (1979) *Decision Processes in Visual Perception* (Academic, New York).
49. Fetsch CR, Pouget A, DeAngelis GC, Angelaki DE (2011) Neural correlates of reliability-based cue weighting during multisensory integration. *Nat Neurosci* 15(1):146–154.
50. Daw ND, Niv Y, Dayan P (2005) Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat Neurosci* 8(12): 1704–1711.
51. Friedman D, Hakerem G, Sutton S, Fleiss JL (1973) Effect of stimulus uncertainty on the pupillary dilation response and the vertex evoked potential. *Electroencephalogr Clin Neurophysiol* 34(5):475–484.
52. Crone EA, Somsen RJ, Van Beek B, Van Der Molen MW (2004) Heart rate and skin conductance analysis of antecendents and consequences of decision making. *Psychophysiology* 41(4):531–540.
53. Law CT, Gold JI (2009) Reinforcement learning can account for associative and perceptual learning on a visual-decision task. *Nat Neurosci* 12(5):655–663.
54. Daniel R, Pollmann S (2012) Striatal activations signal prediction errors on confidence in the absence of external feedback. *Neuroimage* 59(4):3457–3467.
55. Guggenmos M, Wilbertz G, Hebart MN, Sterzer P (2016) Mesolimbic confidence signals guide perceptual learning in the absence of external feedback. *eLife* 5:e13388.
56. Lashley KS (1951) The problem of serial order in behavior. Cerebral Mechanisms in Behavior: The Hixon Symposium, ed Jeffress LA (Wiley, Oxford), pp 112–146.
57. Friston K (2008) Hierarchical models in the brain. *PLOS Comput Biol* 4(11):e1000211.
58. FitzGerald TH, Moran RJ, Friston KJ, Dolan RJ (2015) Precision and neuronal dynamics in the human posterior parietal cortex during evidence accumulation. *Neuroimage* 107:219–228.
59. Pearson JM, Heilbronner SR, Barack DL, Hayden BY, Platt ML (2011) Posterior cingulate cortex: Adapting behavior to a changing world. *Trends Cogn Sci* 15(4):143–151.
60. Karlsson MP, Tervo DG, Karpova AY (2012) Network resets in medial prefrontal cortex mark the onset of behavioral uncertainty. *Science* 338(6103):135–139.
61. Glaze CM, Kable JW, Gold JI (2015) Normative evidence accumulation in unpredictable environments. *eLife* 4:308825.
62. Brainard DH (1997) The Psychophysics Toolbox. *Spat Vis* 10(4):433–436.

Supporting Information

Purcell and Kiani 10.1073/pnas.1524685113

SI Text

Motion Direction Discrimination Training. Throughout the experiment, subjects were seated in an adjustable chair in a semidark room with chin and forehead supported before a Cathode Ray Tube (CRT) display monitor (20", EIZO FlexScan T966; refresh rate 75 Hz, screen resolution 1,600 × 1,200; viewing distance 53 cm). Stimulus presentation was controlled with Psychophysics Toolbox (62) and Matlab. Eye movements were monitored using a high-speed infrared camera (Eyelink; SR-Research). Gaze positions were recorded at 1 kHz.

The trial started when the subject looked at a small red fixation point (FP, 0.3° diameter circle) at the center of the screen. Following a variable delay (200–500 ms; truncated exponential), two red targets (0.5°) appeared on opposite sides of the screen equidistant from the FP (8° eccentricity). Following another random delay (200–500 ms; truncated exponential), a dynamic random dots stimulus appeared within a 5° circular aperture centered on the FP. The dots were white 4 × 4 pixel squares (0.096° × 0.096°) on black background (dot density, 16.7 dots per square degree per second). The stimulus consisted of three independent sets of moving dots shown in consecutive frames. Each set of dots was shown for one video frame and then replotted three video frames later ($\Delta t = 40$ ms). When replotted, a subset of dots were offset coherently from their original location to create apparent motion (speed, 5° per second) while the remaining dots were placed randomly within the aperture. Following the offset of the motion stimulus, a delay period (400–1,000 ms; truncated exponential) was imposed before the Go signal (FP offset). Subjects were instructed to maintain gaze on the FP throughout the trial until the Go signal. If the gaze deviated more than 2° from the FP, the trial was aborted. Following the Go signal, subjects reported their perceived direction of motion by shifting gaze to the choice target in the direction of motion and maintaining the gaze within 3° of the target for 200 ms. Subjects received distinct auditory feedback for correct (positive feedback) and error (negative feedback) responses. Aborted trials had a neutral, uninformative auditory feedback and were excluded from the analyses. Training on the basic motion discrimination task continued until subjects achieved high performance as indicated by psychophysical thresholds <17% (*Results*).

We manipulated the difficulty of the motion direction discrimination in two ways. First, the motion stimulus duration on each trial was randomly sampled from a truncated exponential distribution (100–900 ms, mean = 330 ms). Second, the motion strength varied randomly across trials. The motion strength was determined by the percentage of coherently displaced dots: 0%, 3.2%, 6.4%, 12.8%, 25.6%, and 51.2%. On trials with 0% coherence, positive feedback was randomly delivered for half of the trials, and negative feedback was delivered on the other half. Training on the basic motion discrimination task continued until subjects achieved high performance, as indicated by psychophysical thresholds of <17% (*Results*).

Changing Environment Task. Subjects were introduced to the changing environment task (Fig. 1*A*) following motion direction discrimination training. The experimental setup, motion stimulus, and timing of events were unchanged from training. However, instead of one pair of choice targets, subjects were presented with two pairs of choice targets (four targets total), one pair above and one pair below the FP (10° eccentricity; ±3.5° above/below FP; ±9.4° left/right of FP). The right and left targets in each pair corresponded to the right and left motion directions, respectively. We refer to the upper and lower pairs of choice targets as two environments. On any given trial, only one environment was correct. Subjects were instructed to choose the target that corresponded to the correct motion direction and correct environment. We refer to the choice of left versus right targets as the "direction choice" and the choice of upper versus lower targets as the "environment choice." An environment remained stable for several trials according to a truncated geometric distribution (range 2–15 trials, mean 6) and then changed (Fig. 1*B*). Subjects were not explicitly cued about the correct environment or when it changed—they had to discover it. They received positive feedback only when both the environment and direction choice were correct. Negative feedback, however, was ambiguous; it occurred when either the environment or the direction choice were incorrect. Subjects had to resolve this ambiguity based on feedback and their expected accuracy in past choices. The auditory tones corresponding to positive and negative feedback were identical to those used for direction discrimination training. During training, subjects were told that their goal was to maximize the proportion of correct trials and that, to do this, they should try to identify environment changes as accurately and as soon as possible.

We adjusted the changing environment task design to simplify the interpretation of experimental results. First, to eliminate mistakes due to misremembering of the previously chosen environment, the targets for the environment chosen in the last trial were slightly brighter. In other words, choosing a brighter target always corresponded to staying in the same environment, whereas choosing a dimmer target always corresponded to an environment switch. This task design reduced the burden on subjects' working memory, helping them fully focus on the decision about motion direction and environment on the current trial. In a second modification, trials in which subjects incorrectly switched environment (i.e., switch errors) were followed by presentation of the text, "Switch error, Go back!" This prevented prolonged confusion following incorrect switches and simplified the interpretation of results. Neither of these modifications was critical for our results—very similar results were obtained in earlier versions of the task without these modifications.

Each subject contributed several sessions of data across days. In each session, subjects performed three or four blocks of 100–200 trials (mean trials per subject = 2,958 trials; range = 2,359–3,485; total trials across subjects = 17,749).

To test for an influence of environment statistics on subjects' switching behavior, we conducted a follow-up experiment in which five subjects performed the task with longer environments. The training procedure, experimental setup, stimulus, instructions, and timing of events within trials were identical to those described above. The only difference was that we increased the mean and range of environment durations experienced by the subjects (truncated geometric distribution, range 3–20 trials, mean 10). This allowed us to assess how subjects' switching behavior changed with longer environment durations and, most importantly, how these changes could be explained by our modeling framework (Fig. S5).

To verify that subjects used confidence to disambiguate the causes of negative feedback—flawed strategy or poor information—we conducted a follow-up experiment in which six subjects reported their direction choice confidence on each trial (Fig. S6). The task was identical to the main experiment, except that targets were replaced by elongated bars (7° long, 0.75° wide) and the environment duration distribution was matched to our first follow-up experiment (truncated geometric distribution, range 3–20, mean 10). The targets were placed at 7° eccentricity and oriented 45° (upper left and lower right targets) or 135° (upper right and lower left targets) to create a diamond pattern around the FP (Fig. S6*A*). As before, subjects indicated their environment choices by responding to the upper or lower targets

and indicated their direction choices by responding to the left or right targets. In addition, subjects indicated their degree of confidence that their direction choice was correct by varying the landing point of their saccade along the length of the chosen target (4). We report subjects' confidence as the saccade end point in units of degrees along the target in the direction of increasing confidence (min = –4.5°, max = +4.5°, which includes the response window surrounding each target). Each target was colored with a spectrum ranging from red at one end (maximal certainty) to green at the other end (minimal certainty) in 10 discrete steps. Subjects were instructed that their confidence ratings should reflect only direction choice confidence and not environment choice confidence. To test whether motion direction choice confidence predicted subjects' environment choices independent of stimulus properties, we removed the trial-to-trial variability of the motion stimulus for half of the trials by using a fixed seed for the pseudorandom number generator (one per coherence and direction) (3).

Supplemental Behavioral Analyses. Our follow-up experiment allowed us to test whether subjects' direction choice confidence predicted environment choices independent of stimulus properties by analyzing only the subset of trials in which trial-to-trial stimulus variability was removed. For these trials, we computed saccade end point residuals by subtracting the mean saccade end point for each motion strength and duration quantile (20 quantiles; other numbers of quantiles produced similar results). The resulting saccade end point residuals were symmetrically distributed around zero. We obtained identical results using a parametric approach in which we fit saccade end points with a linear regression using motion strength, duration, and their interaction as predictors and then computed residuals by subtracting model predictions.

Optimal Solution for the Changing Environment Task. Switching from an old environment ($E=1$) to a new environment ($E=2$) should happen when the posterior odds of the new environment exceeds 1. The posterior odds are

$$PO = \frac{p[E(T)=2|C(T-n,...,T),F(T-n,...,T)]}{p[E(T)=1|C(T-n,...,T),F(T-n,...,T)]} = \frac{p[E(T)=2,C(T-n,...,T),F(T-n,...,T)]}{p[E(T)=1,C(T-n,...,T),F(T-n,...,T)]}, \quad \textbf{[S1]}$$

where $E(T)$, $C(T)$, and $F(T)$ are the environment, motion strength (coherence and duration), and feedback on trial T, respectively. We use C to refer to both the motion coherence and duration only to shorten the equations—separating the two will not change the final conclusion. The equality is based on Bayes' rule. It can be shown that the posterior odds ratio becomes 0 for positive feedback on trial T. Therefore, we focus only on sequences of consecutive negative feedbacks that result from staying in the old environment from trial $T-n$ to trial T(trial $T-n$ is the first trial with negative feedback in the sequence, and feedback on trial $T-n-1$ is positive). The numerator and denominator on the second line of Eq. **S1** can be calculated as follows. The denominator is

$$\begin{aligned}
&p[E(T)=1,C(T-n,...,T),F(T-n,...,T)]\\
&=p[E(T-n,...,T)=1,C(T-n,...,T),F(T-n,...,T)]\\
&=p[F(T-n,...,T)|E(T-n,...,T)=1,C(T-n,...,T)]\\
&\quad\times p[E(T-n,...,T)=1]p[C(T-n,...,T)]\\
&=p[C(T-n,...,T)]p[E(T-n,...,T)=1]\\
&\quad\times\prod_{i=n}^{0}p[F(T-i)|E(T-i)=1,C(T-i)]\\
&=p[C(T-n,...,T)]\prod_{i=n}^{0}[1-H(T-i)]\prod_{i=n}^{0}[1-A(T-i)], \quad \textbf{[S2]}
\end{aligned}$$

where $A(T)$ is the expected accuracy (confidence) for the direction choice on trial T. The second line in Eq. **S2** results from our task design that ensures an environment change does not revert until the subject switches and samples the new environment. Put in equations,

$$p[E(T-1)=1|E(T)=1]=1,$$

which can be rearranged using Bayes' rule to show

$$p[E(T-1)=1,E(T)=1]=p[E(T)=1].$$

A similar logic applies to trials before $T-1$ in the sequence. The numerator of posterior odds is

$$\begin{aligned}
&p[E(T)=2,C(T-n,...,T),F(T-n,...,T)]\\
&=\sum_s p[E(T)=2,E(T-n,...,T-1)=s,C(T-n,...,T),F(T-n,...,T)]\\
&=p[C(T-n,...,T)]\sum_s\{p[E(T)=2,E(T-n,...,T-1)=s]\\
&\quad\times\prod_{i=n}^{0}p[F(T-i)|C(T-i),E(T-i)]\}\\
&=p[C(T-n,...,T)]\{H(T-n)\\
&\quad+[1-H(T-n)]H(T-n+1)[1-A(T-n)]+...\}, \quad \textbf{[S3]}
\end{aligned}$$

where s denotes plausible combinations of environments in the previous n trials (e.g., a switch on trial $T-n$, $T-n+1$, etc.). Putting Eqs. **S2** and **S3** in Eq. **S1**, we have

$$\begin{aligned}
PO &= \frac{\left[\begin{array}{l}H(T-n)+H(T-n+1)[1-H(T-n)][1-A(T-n)]+\\ H(T-n+2)[1-H(T-n)][1-H(T-n+1)][1-A(T-n)][1-A(T-n+1)]+\\ \ldots+H(T)\prod_{i=n}^{1}\{[1-H(T-i)][1-A(T-i)]\}\end{array}\right]}{\prod_{i=n}^{0}[1-H(T-i)]\prod_{i=n}^{0}[1-A(T-i)]} \quad \textbf{[S4]}\\
&\approx \frac{H(T-n)}{\prod_{i=n}^{0}[1-H(T-i)]\prod_{i=n}^{0}[1-A(T-i)]}
\end{aligned}$$

The approximation in the second line of the equation is justified because the higher terms in the numerator become exponentially smaller. This approximation makes Eqs. **S5** and **S6** deviate slightly from the true optimal solution. For simplicity, however, we use the term "optimal" (instead of nearly optimal) for those equations throughout the paper.

Subjects should switch environment when $PO > 1$, that is, when

$$\frac{1}{\prod_{i=n}^{0}[1-A(T-i)]} > \frac{\prod_{i=n}^{0}[1-H(T-i)]}{H(T-n)} \qquad \textbf{[S5]}$$

or

$$\sum_{i=n}^{0} \log \frac{1}{1-A(T-i)} > -\log[H(T-n)] + \sum_{i=n}^{0} \log[1-H(T-i)]. \qquad \textbf{[S6]}$$

Eq. **S6** suggests that accumulation of switch evidence, represented by $\log\{1/[1-A(T)]\}$, toward a switch bound, represented by $-\log[H(T-n)] + \sum_{i=n}^{0} \log[1-H(T-i)]$, is an optimal solution for this task. The first term in the right-hand side of Eq. **S6**, $-\log[H(T-n)]$, shows that the switch bound depends on the location of the first negative feedback in the sequence of trials within the environment. This dependence contributes to switch urgency if subjective hazard rates grow over time. The second term in the right-hand side of Eq. **S6**, $\log[1-H(T-i)]$, is negative because $H(T)$ is bounded between zero and 1. This bound collapse contributes to the switch urgency as the number of consecutive negative feedbacks increases.

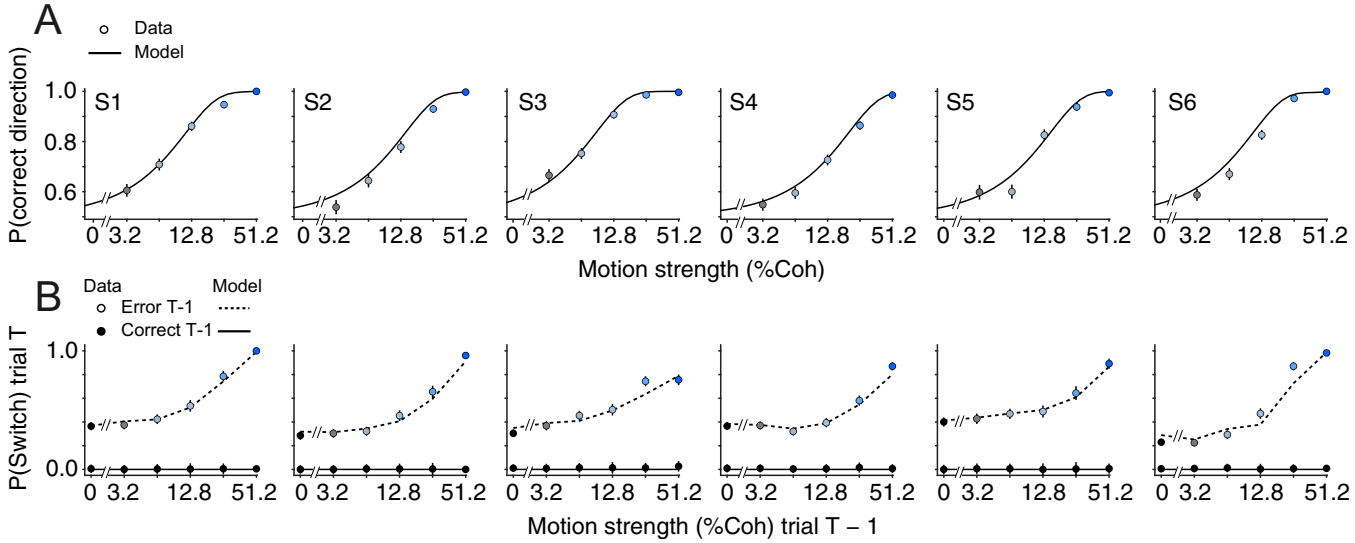

Fig. S1. (*A*) Direction and (*B*) environment choices for individual subjects (S1–S6). Conventions are similar to Fig. 2. In both panels, circles are data, and lines are model fits. All subjects were more likely to switch environment choices following negative feedback on trials with stronger motion. Error bars are SE.

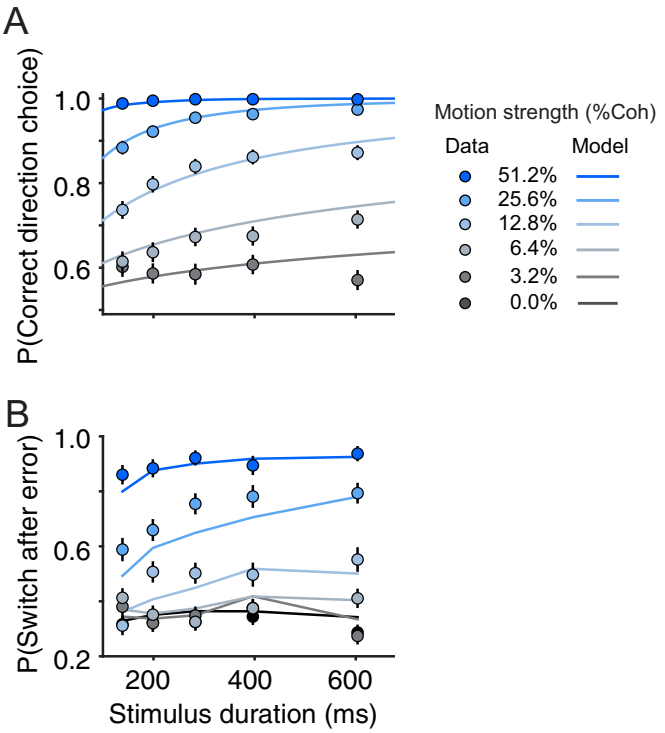

Fig. S2. Motion strength and duration influence both motion direction choices and environment choices. (*A*) The accuracy of motion direction choices increased with motion strength and duration on the current trial. Data points show the accuracy of direction choices disregarding environment choices. Stimulus viewing durations were divided into quintiles. The lines show model fits in both panels. Data and fits are pooled across subjects. (*B*) Environment switches were more likely following negative feedback on trials with stronger motion and longer stimulus durations. Error bars are SE.

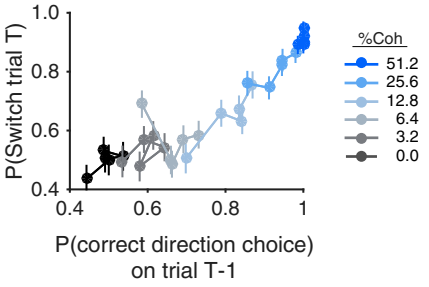

Fig. S3. Expected accuracy is the key variable determining switching. The probability of switching after negative feedback is plotted as a function of direction choice accuracy on the preceding trial for each motion strength (%Coh) and duration quintile. If different motion strengths and durations are associated with similar expected accuracies, they predict a similar probability of switch following a negative feedback.

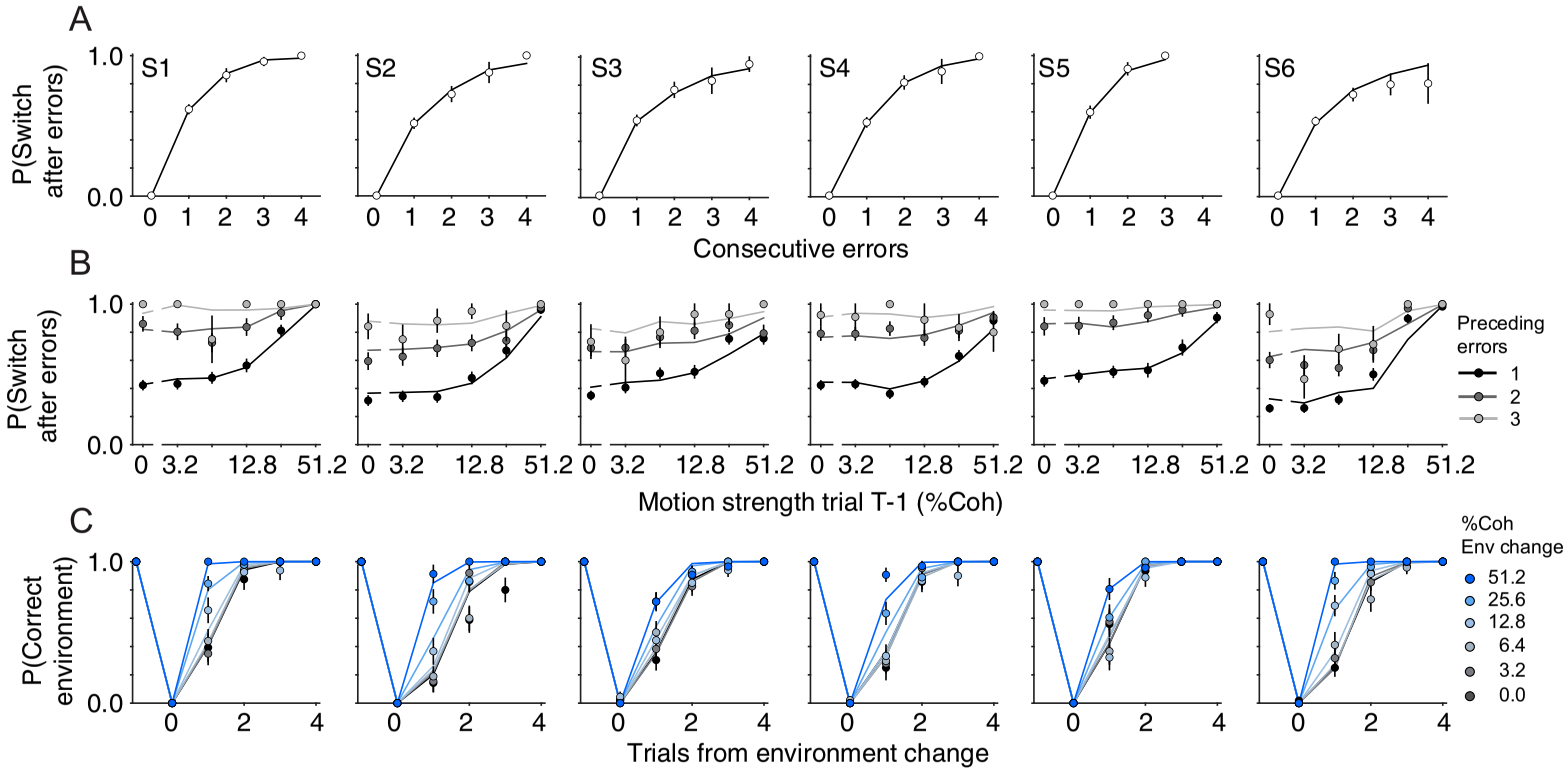

Fig. S4. Environment choices of all subjects were informed by integration of feedback and expected motion direction accuracy across multiple trials. Conventions are similar to Fig. 3. (*A*) Probability of switching environment choices following different numbers of consecutive feedbacks. (*B*) Probability of switching as a function of motion strength on the previous trial following one, two, or three consecutive negative feedbacks. (*C*) Proportion of correct environment choices as a function of the number of trials relative to an uncued environment change. Color indicates motion strength on the trial in which change occurred (trial = 0; see key). In all panels, circles are data, and lines are model fits. Error bars are SE.

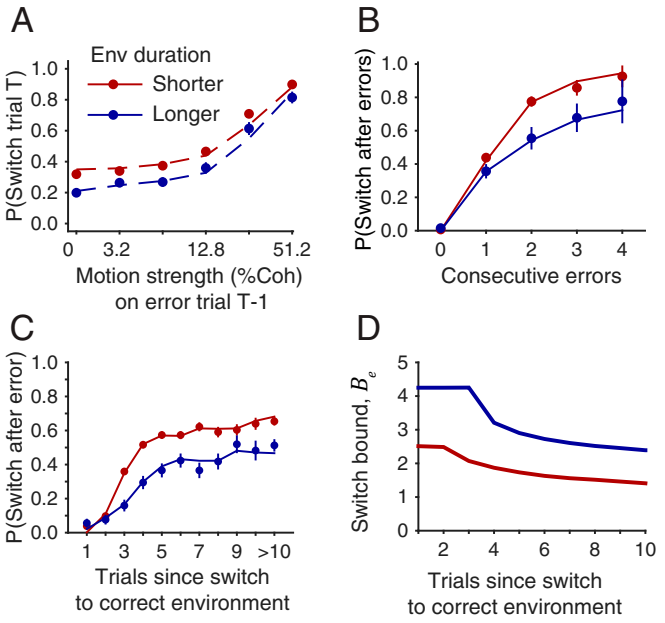

Fig. S5. Increased switch bound explains reduced switch rates for longer environments. Shown are data and model fits from five subjects who performed the task with longer and less volatile environments (blue; environment length, 3–20 trials, mean = 10 trials) compared with the original experiment with shorter environments (red; environment length, 2–15 trials, mean = 6 trials). (*A*) Subjects switched less frequently following negative feedback when environments were longer. Switching still depended on motion strength and feedback on previous trials, but switch rates were lower for all motion strengths. Lines are model fits. The red data points and line are identical to the colored circles and dashed line in Fig. 2*B*. (*B*) Subjects were less likely to switch following runs of consecutive negative feedback when environments were longer. Lines are model fits. The red line and data points are identical to Fig. 3*A*. (*C*) When the environments were longer, switching after errors was less frequent and increased more gradually with number of trials spent in an environment. The red data points and line are identical to Fig. 5*A*. (*D*) Subjects used larger switch bounds for longer environments. The lines show the average collapsing switch bounds for the five new subjects who experienced longer environment durations (blue) and the six subjects who experienced shorter environment durations (red).

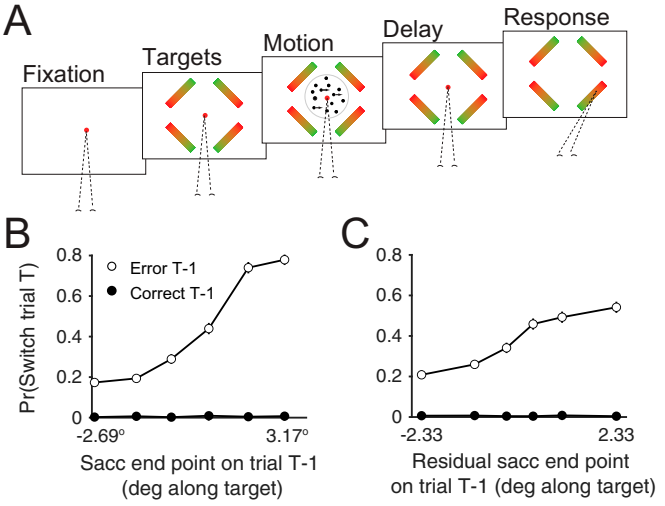

Fig. S6. Subjective confidence informs environment choices. (*A*) Six subjects performed a modified changing environment task with simultaneous report of motion direction confidence. The task structure is identical to our main experiment except that the targets are replaced with elongated bars (length 7°). Subjects varied the end point of their eye movements along the length of the bar to indicate motion direction choice confidence (green, minimal confidence; red, maximal confidence) (4). (*B*) Subjects were more likely to switch environment choices following negative feedback on trials in which they reported higher subjective confidence. Data points show the mean proportion of environment switches as a function of the saccade end point along the length of each target. Saccade end points are divided into six quantiles. (*C*) Subjects were more likely to switch environment choices when subjective confidence was higher for the same motion strength and duration. Probability of switching is plotted as a function of residual variations of saccade end points after subtracting the mean of end points for the motion strength and duration (*SI Text*). Error bars are SE.

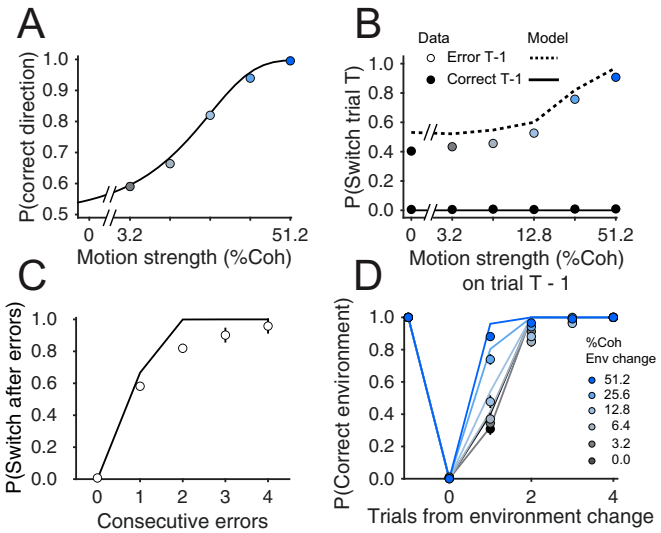

Fig. S7. Comparison of observed switching behavior to ideal switching performance. Conventions are similar to Figs. 2 and 3. We fit the motion direction choices using the sensitivity (k) and decision bound (B_d) on the sensory decision variable. Then, we predicted ideal switch performance based on the optimal form of switch evidence and switch bound given the expected accuracy from the sensory decision process and the experienced hazard rate (*Materials and Methods*). No probability weighting function was applied, and switch noise was excluded. We focused on error sequences that began when the hazard rate was larger than zero (trial three onward). (*A*) Accumulation of sensory evidence to a decision bound explains the proportion of correct motion direction choices. (*B*) The proportion of switches increases after negative feedback for choices associated with greater expected accuracy for both model predictions and subjects, but subjects' overall switch rates are lower. (*C*) The switch rate increases with consecutive negative feedbacks for both model predictions and subjects, but subjects' switch rates increase at a slower rate. (*D*) On the first trial after an environment change, the probability of switching to the correct environment depends on motion strength on the change trial (trial 0) for both model predictions and subjects. However, again, subjects perseverated in the old environment longer than predicted by the optimal model.

Table S1. Best fitting parameters (± SE) of the uncertainty accumulation model

Subject	k	B_d	σ_e	B_0	γ
S1	0.48 ± 0.002	40.91 ± 2.942	0.84 ± 0.009	0.12 ± 0.003	1.89 ± 0.030
S2	0.39 ± 0.002	49.36 ± 5.296	1.00 ± 0.010	0.09 ± 0.002	2.57 ± 0.045
S3	0.65 ± 0.003	18.69 ± 0.218	1.77 ± 0.015	0.05 ± 0.002	2.00 ± 0.022
S4	0.27 ± 0.003	40.16 ± 7.843	0.81 ± 0.034	0.10 ± 0.005	1.84 ± 0.049
S5	0.37 ± 0.003	22.90 ± 0.965	0.82 ± 0.019	0.14 ± 0.005	2.18 ± 0.073
S6	0.51 ± 0.003	61.35 ± 7.027	0.91 ± 0.010	0.04 ± 0.002	1.82 ± 0.019

The main model in our experiment did not include leakage ($\lambda = 0$) and included perfect evidence resets following positive feedback ($q = -\infty$). B_0 and γ determined the switch bound, B_e (Eqs. **15** and **16**, Fig. 5*C*, and *Materials and Methods*), and ω was set to 1 according to the optimal model. These parameters generated the fits shown in Figs. 2, 3, and 5 and Figs. S1, S2, and S4 (lines).

Table S2. Logistic regression coefficients for Eqs. 1–8 (*Materials and Methods*)

Equation	β_0	β_1	β_2	β_3
Eq. **1**	—	10.1 ± 0.26 ($P < 10^{-10}$)	0.4 ± 0.09 ($P = 3.8 \times 10^{-7}$)	—
Eq. **2**	−1.0 ± 0.06 ($P < 10^{-10}$)	6.2 ± 0.23 ($P < 10^{-10}$)	0.3 ± 0.15 ($P = 0.03$)	—
Eq. **3**	−2.9 ± 0.11 ($P < 10^{-10}$)	5.9 ± 0.26 ($P < 10^{-10}$)	0.3 ± 0.18 ($P = 0.06$)	1.5 ± 0.06 ($P < 10^{-10}$)
Eq. **4**	−4.5 ± 0.30 ($P < 10^{-10}$)	0.5 ± 0.69 ($P = 0.49$)	—	—
Eq. **5**	0.5 ± 0.06 ($P < 10^{-10}$)	1.6 ± 0.31 ($P = 6.61 \times 10^{-7}$)	9.9 ± 0.95 ($P < 10^{-10}$)	35.6 ± 3.99 ($P < 10^{-10}$)
Eq. **6**	0.6 ± 0.08 ($P < 10^{-10}$)	4.9 ± 0.58 ($P < 10^{-10}$)	1.9 ± 0.76 ($P = 0.01$)	−12.9 ± 3.32 ($P = 1.0 \times 10^{-4}$)
Eq. **7**	0.8 ± 0.08 ($P < 10^{-10}$)	2.0 ± 0.35 ($P = 2.2 \times 10^{-8}$)	−1.8 ± 0.37 ($P = 1.5 \times 10^{-6}$)	—
Eq. **8**	−2.3 ± 0.08 ($P < 10^{-10}$)	5.9 ± 0.24 ($P < 10^{-10}$)	0.4 ± 0.16 ($P = 0.01$)	0.3 ± 0.01 ($P < 10^{-10}$)

All coefficients were calculated using maximum likelihood fitting. Trials are pooled across subjects.