Short title: Confidence drives value-based learning in the absence of feedback

The value of confidence: Confidence prediction errors drive value-based learning in the absence of external feedback

Lena Esther Ptasczynski[1,2*], Isa Steinecker[1,3], Philipp Sterzer[1], Matthias Guggenmos[1]

[1]Department of Psychiatry and Psychotherapy, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany

[2]Berlin School of Mind and Brain, Humboldt-Universität zu Berlin, Berlin, Germany

[3]Bernstein Center for Computational Neuroscience, corporate member of Humboldt-Universität zu Berlin, Berlin, Germany

* Corresponding author

E-mail: lena-esther.ptasczynski@charite.de (LEP)

Abstract

Reinforcement learning algorithms have a long-standing success story in explaining the dynamics of instrumental conditioning in humans and other species. While normative reinforcement learning models are critically dependent on external feedback, recent findings in the field of perceptual learning point to a crucial role of internally-generated reinforcement signals based on subjective confidence, when external feedback is not available. Here, we investigated the existence of such confidence-based learning signals in a key domain of reinforcement-based learning: instrumental conditioning. We conducted a value-based decision making experiment which included phases with and without external feedback and in which participants reported their confidence in addition to choices. Behaviorally, we found signatures of self-reinforcement in phases without feedback, reflected in an increase of subjective confidence and choice consistency. To clarify the mechanistic role of confidence in value-based learning, we compared a family of confidence-based learning models with more standard models predicting either no change in value estimates or a devaluation over time when no external reward is provided. We found that confidence-based models indeed outperformed these reference models, whereby the learning signal of the winning model was based on the prediction error between current confidence and a stimulus-unspecific average of previous confidence levels. Interestingly, individuals with more volatile reward-based value updates in the presence of feedback also showed more volatile confidence-based value updates when feedback was not available. Together, our results provide evidence that confidence-based learning signals affect instrumentally learned subjective values in the absence of external feedback.

Author summary

Reinforcement learning models successfully simulate value-based learning processes (e.g., "How worthwhile is it to choose the same option again?") when external reward feedback is provided (e.g., drops of sweet liquids or money). But does learning stagnate if such feedback is no longer provided? Recently, a number of studies have shown that subjective confidence can likewise act as an internal reward signal, when external feedback is not available. These results are in line with the intuitive experience that being confident about choices and actions comes with a satisfying feeling of accomplishment. To better understand the role of confidence in value-based learning, we designed a study in which participants had to learn the value of choice options in phases with and without external feedback. Behaviorally, we found signatures of self-reinforcement, such as increased confidence and choice consistency, in phases without feedback. To examine the underlying mechanisms, we compared computational models, in which learning was guided by confidence signals, with more standard reinforcement learning models. A statistical comparison of these models showed that a confidence-based model in which generic confidence prediction errors (e.g., "Am I as confident as expected?") guide learning indeed outperformed the standard models.

Introduction

The reinforcement learning principle, according to which learning is controlled by action-contingent feedback, explains fundamental forms of learning across many modalities and species (1). Yet, there are important instances of learning that occur in the absence of external feedback, and which thus challenge the generality of this model class.

A prominent example is perceptual learning, for which behavioral improvements are frequently found through training or mere exposure and without any external feedback (2–6). Moreover, the (subjective) sense of accomplishment in an unrelated task likewise induces perceptual learning, even in the absence of stimulus awareness (7,8). Together, these findings have led to the notion of a 'diffuse internal reward signal' (9), i.e. a reinforcement signal that is triggered based on some form of internal feedback.

More recently, such internal feedback signals have been investigated by means of fMRI, operationalized in the form of confidence reports (10–13). The consistent finding of these studies was that confidence-based learning signals engaged a network of brain regions that has previously been identified for the coding of *reward* prediction errors (14), including the ventral striatum (a dopaminergic target region) and the ventral tegmental area (a dopaminergic source region). In line with these neurobiological observations, a recent study has shown that having confidence in one's own actions is associated with a feeling of increased pleasantness and satisfaction (15). Together, these findings suggest that learning based on external and internal feedback operates on a shared neural mechanism.

In the present study, we aimed to examine the generality of such putative confidence-based learning signals. We hypothesized that, if confidence in actions indeed takes the form of a diffuse internal reward signal, it should also affect the subjective values of these actions,

similar to instances of external reinforcement. In fact, the notion that subjective values are malleable even in the absence of external feedback is not new. The most prominent example is the cognitive dissonance theory of Festinger (16), which posits that values of chosen options are reinforced to reduce cognitive dissonance between the chosen and the unchosen option. Although early evidence for the theory by Brehm (17) has been challenged on methodological grounds (18,19), more recent studies have provided new support (20–27). In a very recent study, Luettgau and colleagues (28) have shown that such choice-induced preference changes can also be observed for classically conditioned stimuli.

In the present work, we designed an instrumental conditioning task in which observers learned about the monetary values of a set of conditioned stimuli (CS). Crucially, after an initial training phase with monetary feedback, subjects entered a second phase in which action-contingent feedback was omitted. Subjects were told that they would eventually receive the rewards for their actions at the end of a block, but they did not get trial-by-trial feedback on their choices. We reasoned that, in the absence of external feedback, value representations would still be shaped by a subject's confidence in their choices.

While our main analytic approach was model-based (see below), we also tested three direct behavioral hypotheses. Specifically, we reasoned that if the degree of confidence in a value-based choice reinforces the value of this very choice, the result is a self-reinforcing cycle in which subjective values for more preferred choices are further strengthened and less preferred choices are further devalued. Over time, the absence of external feedback in instrumental conditioning should thus lead to an augmentation of preferences for available choice options. We therefore hypothesized that the absence of feedback would lead to 1) an increase of choice confidence, 2) an increase of choice consistency and 3) an augmentation of initial preferences ("the rich get richer and the poor get poorer").

To better understand the dynamics of value changes in the absence of feedback – and a potential role of confidence therein – we devised a family of computational models in which confidence guides learning when no external feedback is available. In terms of a confidence-based learning signal, we adopted the notion of confidence prediction errors: the difference between expected confidence and actual confidence (12). We have previously shown that confidence prediction errors constitute a sensible computational learning signal in the context of perceptual learning and that a ventral striatal correlate of this signal was predictive of perceptual learning success (12,29).

Results

Behavioral results

The experimental paradigm was structured in the logic of a standard value-based decision making task in which participants had to learn about the values of initially neutral conditioned stimuli (CS). The experiment consisted of 11 blocks in each of which participants had to learn about the value of 5 new CS with different objective values. Trial-wise feedback was provided in the first and third phase (Phases 1 and 3) of a block, but critically, was omitted for a varying number of trials in between (Phase 2) (Figure 1A). In each individual trial, participants had to make a choice between two CS and subsequently indicated their choice confidence on a scale from 0 to 10 (Figure 1B).

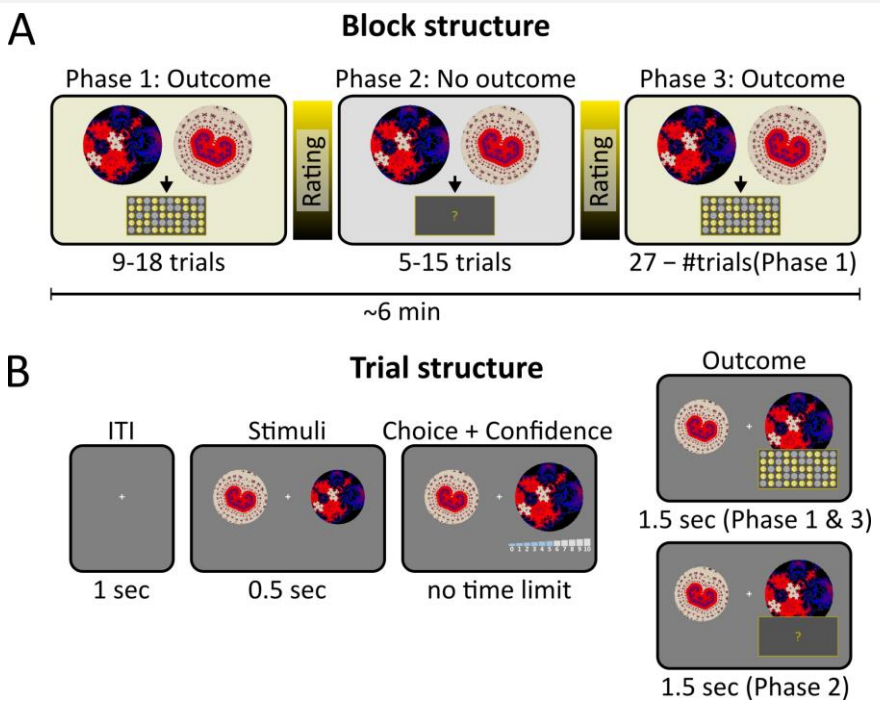

Figure 1. Experimental design. (A) Block structure. In each block, participants had to learn about the values of five CS based on the feedback in phases 1 and 3. The critical phase was a period of 5-15 trials in-between phases 1 and 3 in which participants did not receive feedback (phase 2). Before and after phase 2, participants rated the values of each CS on a continuous scale. **(B)** Trial structure. In each trial, participants chose between two CS and indicated their confidence on a scale from 0 to 10. In phases 1 and 3, the reward outcome for the chosen option was presented in the form of a scratch card with 50 fields, each of which could contain a 1 EUR coin or a blank. In phase 2, the scratch card was not revealed; however, participants were instructed that they would receive the hidden reward on the scratch card at the end of the experiment.

We first ensured that participants successfully learned the task. For all analyses involving behavioral learning effects, we used either generalized linear (GLMM; for the correctness of choices) or linear (LMM; for confidence) mixed effects models. We found that participants improved their choice performance (proportion correct) by learning from trial-wise feedback, as indicated by a main effect of trial number across the feedback phases 1 and 3 (GLMM: $z = 11.72$, $p < 0.001$; Figure 2A and Supplementary Table S1). In addition, this was reflected in a concurrent increase of subjective confidence across trials (LMM: $z = 68.20$, $p < 0.001$; Figure 2B and Supplementary Table S2). Overall, participants' performance increased from 0.63 ± 0.01 (s.e.m.) in phase 1 to 0.77 ± 0.01 (s.e.m.) in phase 3 (paired t-test: $t_{63} = 13.32$, $p < 0.001$)

and their confidence increased from 3.27 ± 0.25 (s.e.m.) in phase 1 to 6.06 ± 0.27 (s.e.m.) in phase 3 ($t_{63} = 17.26$, $p < 0.001$).

The primary focus of our investigation was on the behavioral dynamics in phase 2, in which no feedback was provided. Specifically, we were interested whether behavioral changes across phase 2 in terms of choice consistency (see below), confidence ratings and subjective value ratings showed signatures of self-reinforced learning.

Across trials in phase 2, performance did not change significantly (0.75 ± 0.003 [s.e.m.]), as shown by a non-significant main effect of trial number (GLMM: $z = -0.35$, $p = 0.726$; Figure 2A and Supplementary Table S3). By contrast, confidence increased across phase 2 (5.75 ± 0.04 [s.e.m.]; LMM: $z = 3.12$, $p = 0.002$; Figure 2B and Supplementary Table S4) despite the absence of any new information. The confidence increase in phase 2 was still measurable in phase 3: confidence in phase 3 was higher in blocks including phase 2 (5.8 ± 0.3 [s.e.m.]) compared to control blocks in which phase 2 was omitted (5.5 ± 0.3 [s.e.m.]; $t_{63} = 1.9$, $p = 0.032$).

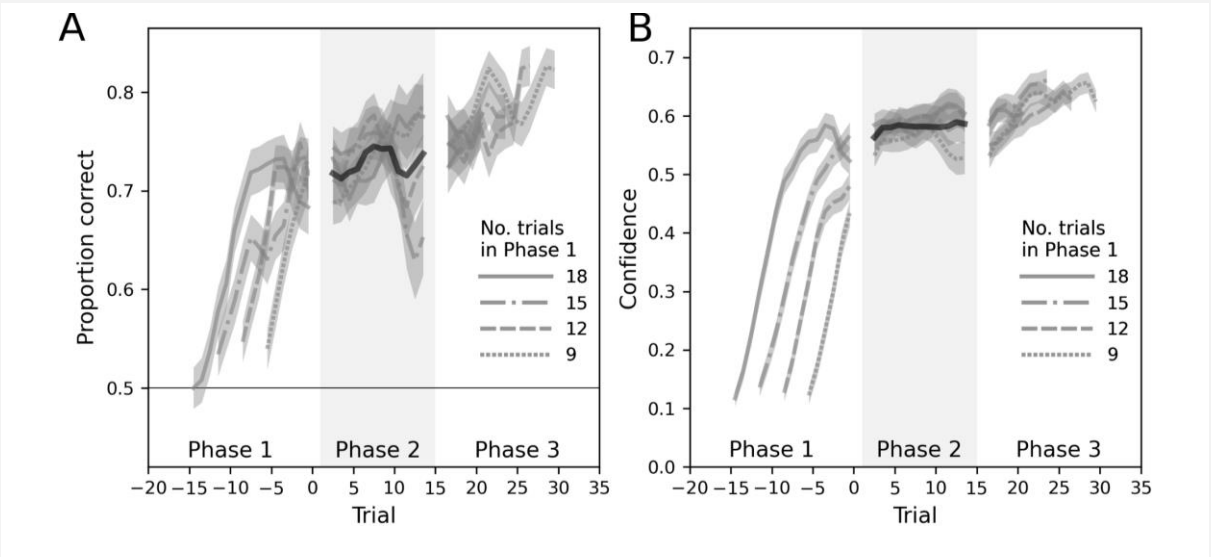

Figure 2. Performance and confidence. Block-averaged time courses are separated according to the duration of phase 1 (9-18 trials) and aligned to the beginning of phase 2. Shaded areas indicate standard error of the mean. **(A)** Value-based learning. The accuracy of choices gradually increased across the phases with feedback (phases 1 and 3), indicating that participants successfully learned the task. **(B)** Confidence. Reported confidence (normalized to [0; 1]) likewise increases across the course of a block.

A second signature of self-reinforced learning is an increase of choice consistency, such that participants become more consistent in their choices when repeatedly being faced with the same pair of CS. Indeed, we found that choice consistency increased in the course of phase 2, indicated by a positive effect of CS pair repetition number (GLMM: $z = 1.85$, $p = 0.064$; Supplementary Table S5). While the effect is rather subtle, given our directional hypothesis we interpret the effect as significant. Figure 3A visualizes the increase in choice consistency by showing the average choice consistency of participants between the first and second occurrence of a choice pair (blue), as well as between the second and the third occurrence (orange). In particular, the proportion of participants showing perfect choice consistency increases from 19% at the second occurrence to 64% at the third occurrence.

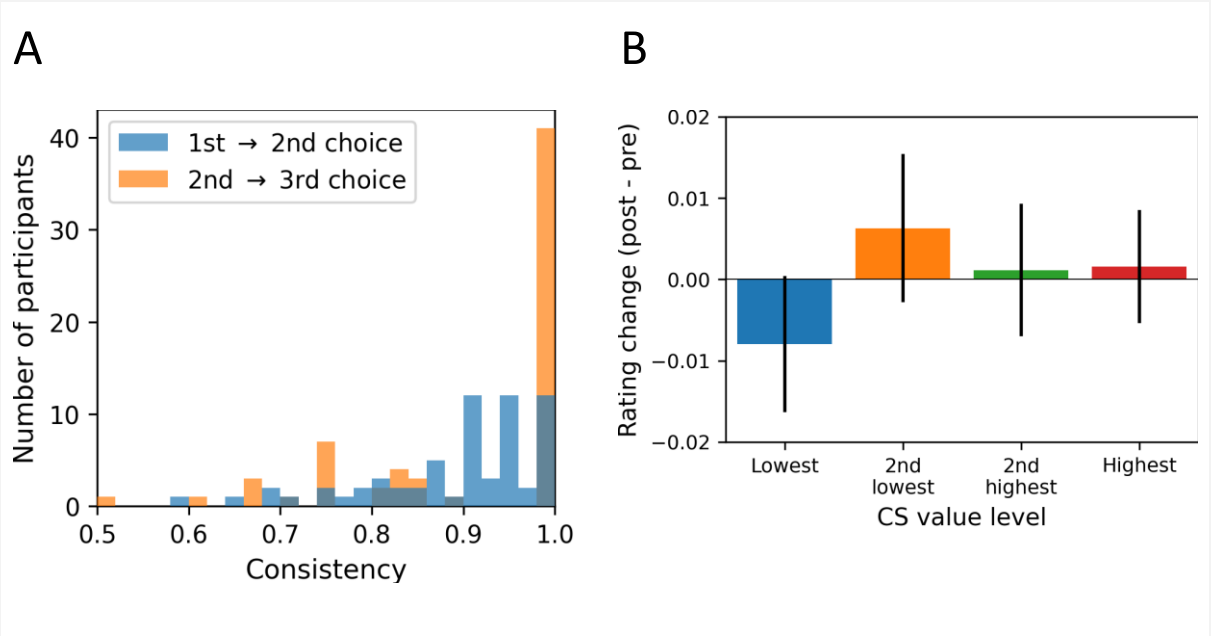

Figure 3. Changes in choice consistency and subjective value ratings in phase 2. (A) Choice consistency between first and second (in blue), as well as between second and third choice (in orange) for identical CS pairs in phase 2. **(B)** Subjective value ratings. Depicted are the changes of the subjective value ratings (post-phase-2 minus pre-phase-2), separately for each of the four CS value levels within a block.

Finally, we tested whether subjective value ratings before and after phase 2 would likewise show a self-reinforcing effect, such that CS with higher objective value would gain subjective value relative to CS with lower objective value. We performed a mixed linear regression

analysis with rating change (post- minus pre-phase-2) as a dependent variable and objective stimulus value as our main independent variable of interest. While the effect was in the expected direction, the effect was far from being significant (LMM: $z = 0.64$, $p = 0.522$; Supplementary Table S6).

Figure 3B visualizes the rating change as a function of CS value, aggregated by the relative CS value order for simplicity reasons (note that while there were 5 CS per block, they were assigned to 4 distinct value levels). Although the absence of an interaction is apparent, the lowest-ranking CS (here displayed in blue) showed an overall rating decrease, while the higher-ranking CS showed numeric increases. As we will elaborate in the discussion, ceiling effects or regression to the mean effects may have masked a potential interaction. Yet, even in this case, the effect is likely a weak one. In an exploratory analysis, we found that the value dependency significantly increased over the course of phase 2 (LMM interaction effect: $z = 2.72$, $p = 0.006$; Supplementary Table S7 and Supplementary Figure S1). This suggests that longer phases without feedback lead to a stronger effect of value on rating changes.

Computational models of value-based learning in the absence of feedback

In line with the neurocomputational similarities between reward- and confidence-based learning (11,12,30), we assume two basic feedback modes. In reward mode, observers maintain a running estimate of *expected values* $\bar{v}_i$ of each stimulus *i* that is updated by means of a conventional Rescorla-Wagner learning rule. Learning is based on reward prediction errors, i.e., the difference between the reward *r* that was obtained in a given trial and the expected value $\bar{v}_i$ of the chosen stimulus:

$$\bar{v}_i \rightarrow \bar{v}_i + \alpha_r \, \Delta v \qquad (1)$$

$$\Delta v = r - \bar{v}_i \quad (2)$$

The speed of learning is controlled by a *reward learning rate* α_r.

Analogously, we assume that observers maintain a running average of the confidence $\bar{c}_{(i)}$ they experienced in past choices of stimuli *i*.

$$\bar{c}_{(i)} \rightarrow \bar{c}_{(i)} + \alpha_c \, \Delta c \quad (3)$$

$$\Delta c = c - \bar{c}_{(i)} \quad (4)$$

Thus, expected confidence $\bar{c}_{(i)}$ is likewise learned and updated by a prediction error signal – in this case the difference between current confidence *c* and the preceding estimate of expected confidence (confidence prediction error). Crucially, current confidence is a behavioral measure obtained through subjective reports in a given trial. The update speed is controlled by a distinct *confidence learning rate* α_c. Note that we put the index (*i*) in brackets to anticipate that we will distinguish between models that update expected confidence in either a stimulus-specific or -unspecific manner. Stimulus-specific models maintain a running estimate of expected confidence for each stimulus separately, whereas stimulus-unspecific models maintain a single stimulus-independent estimate of expected confidence.

Our key hypothesis is that, in the absence of external feedback, value estimates are affected by confidence prediction errors. For instance, when making a choice in which we are very confident, and which thus will typically elicit a positive confidence prediction error, the value of the chosen option is increased. This mechanism is controlled by the *confidence transfer parameter* γ:

$$\bar{v}_i \rightarrow \bar{v}_i + \gamma \, \Delta c \quad (5)$$

Thus, the value of the chosen option (as predicted by the model) is updated in proportion to confidence prediction errors. Note that while expected confidence is tracked throughout the experiment, we assume that confidence-based value updates only apply when no external feedback is available.

As a control, we additionally test a model in which the mere act of a choice − without a modulation by confidence prediction errors − leads to a reinforcement of the associated stimulus:

$$\bar{v}_i \rightarrow \bar{v}_i + \gamma \quad (6)$$

This model is reminiscent of the idea of choice-induced preferences changes (16), which posits that values of chosen options are reinforced to reduce cognitive dissonance between the chosen and the unchosen option.

Finally, we consider the possibility that, in the absence of external feedback, stimuli are subject to devaluation. Although subjects are aware that they will receive the rewards associated with all choices at the end of the experiment, the omission of a choice-contingent reward display might nevertheless cause a devaluation of choice options. This third mechanism is implemented in a way that subjects perceive the absence of trial-by-trial reward feedback as if they received an effective reward of zero. The reward prediction error thus becomes $0 - \bar{v}_i$:

$$\bar{v}_i \rightarrow \bar{v}_i + \alpha_d\,(0 - \bar{v}_i) = (1 - \alpha_d)\,\bar{v}_i \quad (7)$$

The speed of devaluation is controlled by a separate devaluation learning rate α_d.

In sum, we therefore consider models in which values are either unaffected in the absence of feedback, affected by devaluation, affected by the mere act of a choice or affected by confidence prediction errors (stimulus-specific or -unspecific). Table 1 provides an overview of the models under consideration.

Table 1. Models.

Name	**Dynamics in the absence of external feedback**
Static	Values are unchanged / static
Deval	Values of chosen options are subject to devaluation
Choice	Values of chosen options are reinforced irrespective of confidence
ConfSpec	Values of chosen options are updated in proportion to stimulus-specific confidence prediction errors
ConfUnspec	Values of chosen options are updated in proportion to stimulus-unspecific confidence prediction errors

Model comparison: unspecific confidence prediction errors guide value-based learning in the absence of external reward feedback

While the behavioral analyses provides partial evidence for self-reinforcing effects in the absence of external feedback, they are agnostic about the underlying mechanism. To differentiate between different possible mechanisms, and in particular the role of confidence therein, we statistically compared the models introduced before. Three main research questions were associated with this comparison. First, we aimed to clarify whether a confidence-based learning signal interacts with subjective values and thereby partially explains the dynamics of choices in the absence of external feedback. Second, in the context of confidence-based learning models we were specifically interested in whether the

computation of confidence prediction errors relies on a running estimate of expected confidence that is computed in a stimulus-specific (*ConfSpec* model) or stimulus-unspecific (*ConfUnspec* model) manner. And third, we tested whether two simpler models may account for the behavior in phase 2: the *Choice* model, in which subjective values are influenced by the mere act of a choice without a modulation by confidence; and the *Deval* model, in which stimuli are subject to devaluation in the absence of feedback.

We computed the model evidence by means of the Akaike information criterion (AIC; 31) in order to account for the varying complexity of models. As shown in Figure 4, we found that the *ConfUnspec* model best accounted for the choice dynamics in phase 2. The model evidence of the *ConfUnspec* model was significantly better compared to the evidence for the second-best model, the *ConfSpec* model (paired t-test: $t_{63} = 4.14$, $p < 0.001$), and compared to the *Static* model ($t_{63} = 7.55$, $p < 0.001$).

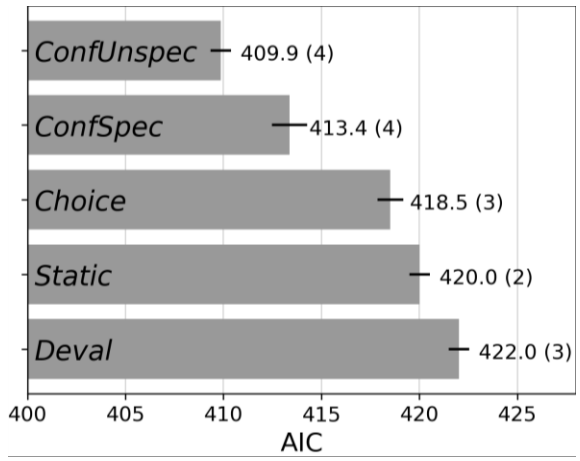

Figure 4. Model evidence and N of free parameters. Akaike information criterion and number of parameters (in parentheses) for all computational models considered and ordered by model fit.

Overall, this comparison thus supports our hypothesis that choice dynamics in value-based decision making are partially driven by confidence-prediction-error-based learning signals. Confidence prediction errors are likely computed in reference to a stimulus-unspecific baseline, i.e. only a single estimate of expected confidence is maintained. By contrast, a model in which the mere act of a choice affects subjective values regardless of confidence performed

239 better than an entirely static model, but was clearly inferior to the confidence models. This
240 suggests that choice confidence may be a key variable to consider when examining the effects
241 of choice-preference changes also in contexts other than the present value-based decision
242 making paradigm.

243 Finally, it is worth pointing out that the evidence against a simple devaluation model was
244 striking. Not only did this model perform worse than the *Static* model, an inspection of
245 devaluation learning rates α_d also revealed that for 96.88% of the participants the best fit for
246 α_d was exactly zero.

247 **Latent dynamics of the winning model: expected value, expected confidence and**
248 **confidence prediction error**

249 To get a better picture of the inner workings of the *ConfUnspec* model, we inspected the time
250 courses of the latent variables *expected value*, *expected confidence* and *confidence prediction*
251 *error*. The time course of the model's expected value shows how value estimates become
252 more distinct over time and become arranged in the order of objective CS values (Figure 5A).
253 This pattern reflects the fact that, on average, participants successfully learned the task (cf.
254 Figure 2A). Expected confidence likewise increased over time, in line with the concurrent
255 increase of confidence ratings (Figure 5B; cf. Figure 2B). For expected confidence, the
256 differentiation with respect to the objective CS values is also evident, although less
257 pronounced than in the case of expected value. It is noteworthy that confidence prediction
258 errors, on average, are positive in phase 2 for all but the lowest-value CS (Figure 5C). One
259 reason is that the learning rates for expected confidence (α_c) often are quite small, such that
260 expected confidence reflects the increase of behavioral confidence only with a certain delay.
261 A likely second reason is that the confidence value transfer (γ) of positive CPEs itself triggers

a self-reinforcing cycle: positive CPEs increase the value of the chosen CS and thus the confidence in future choices of this CS, which in turn increases the probability of positive CPEs.

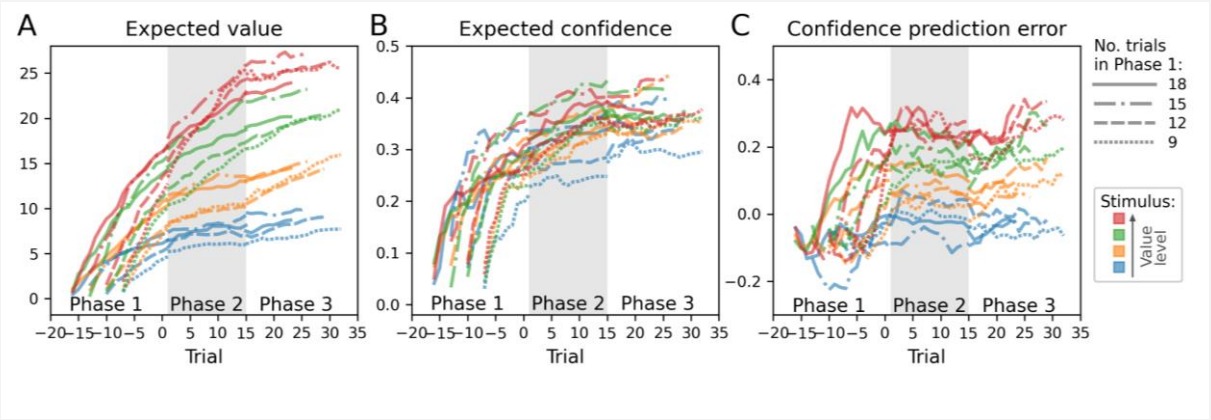

Figure 5. Latent variables of model *ConfUnspec*. All time courses represent averages across blocks and subjects, split according to the duration of phase 1 (line styles) and the four CS value levels within a block (colors). **(A)** Expected value. **(B)** Expected confidence. **(C)** Confidence prediction errors.

Relationship between reward-based and confidence-based learning

We reasoned that, if learning with and without external feedback is based on a similar mechanism, interindividual differences in reward-based learning may be predictive of interindividual differences in confidence-based learning. While reward-based learning is characterized by the reward learning rate α_r, the impact of confidence on subjective values is captured by the confidence transfer parameter γ. Figure 6 shows the distributions of parameters for α_r and γ, as well as for the two remaining parameters of the winning model, namely decision noise β and confidence learning rate α_c.

We indeed found a strong correlation between the reward learning rate α_r and the confidence transfer parameter γ in our winning model ($r = 0.52$, $p < .001$). As a control analysis, and to ensure that both estimates are independent of one another, we correlated the reward learning rate of the *Static* model to the confidence transfer parameter. Here, again, the effect holds, with $r = 0.56$, $p < .001$ (Figure 6E). Thus, observers who show more volatile reward-

based updating of their value-based beliefs also show higher volatility for learning based on confidence prediction errors, when feedback is no longer provided. Of note, the reward learning rate α_r was not correlated to the speed with which observers updated their estimates of expected confidence, characterized by the confidence learning rate α_c ($r = -0.09$, $p = 0.498$; control analysis with *Static* model: $r = -0.03$, $p = 0.842$).

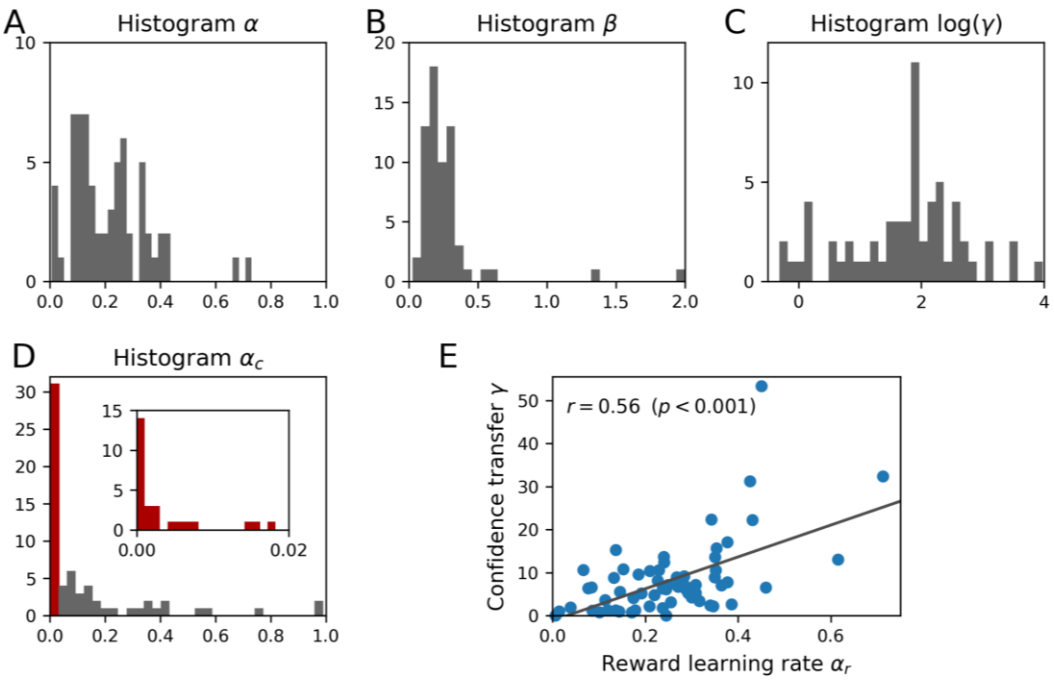

Figure 6. Model parameters for the winning model *ConfUnspec*. (A) Histogram of the reward learning rate α_r (0.21 ± 0.14 [s.d.]). **(B)** Histogram of the inverse decision noise parameter β (0.27 ± 0.28 [s.d.]). **(C)** Histogram of the confidence transfer parameter γ (8.31 ± 8.68 [s.d.]). **(D)** Histogram of confidence learning rate α_c (0.14 ± 0.22 [s.d.]). **(E)** Scatter plot between reward learning rate α_r and confidence transfer γ. The black line indicates a linear fit to the data; the correlation coefficient is based on a Pearson correlation.

Discussion

We investigated the role of confidence-based learning signals in value-based learning and decision-making when external feedback is not available. Consistent with our hypothesis, we found behavioral evidence for signatures of confidence-based self-reinforcement: an increase of subjective confidence, increased choice consistency and a tendency towards self-

reinforcement of subjective values. A model-based analysis showed that a model which considered confidence-based learning signals in phases without external feedback outperformed a static model, as well as a model that predicted devaluation over time.

Overall, our findings thus corroborate the notion that confidence reflects an internal reinforcement learning signal, connatural to reinforcement signals induced through external reward or cognitive feedback. The general mechanistic idea therein is that the brain triggers global reward signals when actions or percepts yield higher confidence than expected, thereby reinforcing underlying neural circuits that gave rise to these actions or percepts. For instance, when practising an instrument, internal reinforcement signals may be triggered when the musician is more confident in a particular performance than expected on the basis of previous attempts. In the context of perceptual learning, such signals may reinforce specific sensory processing pathways that happen to generate percepts associated with above-average confidence.

While the advantage of confidence learning signals is intuitive in these examples, the adaptive advantage of confidence-based learning is less clear in the context of value-based decision making: why should subjective values change at all in the absence of new information?

One possibility is that confidence effects observed in value-based decision making are an accidental side effect − an epiphenomenon − of a mechanism that otherwise proves advantageous in the majority of learning scenarios. In this case, one may seek in vain for the benefits of confidence-based learning in the specific case of value-based decision making. However, another possibility is that self-reinforcement of subjective values may be a pragmatic strategy in the face of a possible memory leakage when feedback is omitted. A classic example for such leakage is retrieval-induced forgetting, i.e. the observation that our memories for items become imprecise merely due to the mnemonic retrieval of these items

311 (32–34). In line with this notion, a recent study has shown that the mere act of a choice
312 between CS induces changes to hippocampal representations of stimulus-outcome
313 associations (28). Thus, without external feedback subjective values of stimuli may become
314 noisy and thus less reliable, at least when observers continue to interact with these stimuli.

315 In this latter view, confidence-based self-reinforcement of subjective values could be a
316 counter strategy for memory loss, trading a more black-and-white estimate of the value
317 landscape (a result of self-reinforcement) with the alternative of an overall flattened
318 landscape in which choices become entirely indifferent. In other words, while it may seem
319 irrational when choice options are transformed into a simplified categorical scheme of either
320 good or bad options, such a scheme may actually be more robust towards mnemonic
321 deterioration. The behavioral effects regarding an increase of confidence and an
322 accompanying increase in choice consistency are in line with such a view.

323 Contrary to our expectation, we did not find a significant value by value change interaction for
324 the subjective value ratings before and after the phase without feedback (although the
325 general direction of results is consistent with our hypothesis). We consider two possible
326 effects that may have counteracted a value-dependent increase of subjective values in
327 phase 2. First, participants were instructed to use the continuous rating scale in an intuitive
328 manner. Naturally, subjects therefore tended to select the lowest and highest ratings for the
329 CS they regarded least and most valuable, respectively. However, in many cases, this intuitive
330 usage of the rating scale effectively left little room for even lower or higher post-phase-2
331 ratings. Thus, the hard constraints imposed by the scale may represent a systematic bias in
332 the opposite direction of our hypothesis.

333 Second, the possibility of noisy memory leakage over the course of phase 2 is expected to lead
334 to a regression to the mean for all CS. Although our proposed mechanism is thought to

mitigate this leakage, the regression-to-the-mean effect is likewise in the opposite direction of our hypothesis and thus reduces the sensitivity to find the interaction. Higher statistical power is necessary to clarify whether the observed null effect is real or a consequence of insufficient statistical power. Alternatively, it is possible that participants simply were not aware of the subtle value changes occurring in phase 2 and hence these changes were not reflected in the subjective ratings.

In the logic of the best-fitting computational model (*ConfUnspec*), subjective values of chosen CS are reinforced if, and only if, choice confidence is higher than expected on the basis of previous confidence experiences, i.e. in the case of positive confidence prediction errors. By contrast, chosen CS are devalued if confidence prediction errors are negative. It is noteworthy that the *ConfUnspec* model, i.e. a model with an unspecific reference (*expected confidence*) to which momentary confidence levels are compared, outperformed a model in which expected confidence was CS-specific. We considered this unlikely, a priori, since an unspecific reference deprives the confidence prediction error from its natural convergence property: an unspecific reference maintains the average confidence level across all CS so that prediction errors, in principle, can be persistently positive (for CS judged to be of relatively high value) or negative (for CS judged to be of relatively low value). As shown in our model inspection (Figure 5A) such a model predicts continuous spreading of values over prolonged periods without feedback. While it is highly unlikely that these dynamics continue ad infinitum, they nevertheless best explain the variance of choices within the 5-15 trials of phase 2.

Interestingly, there is evidence for the presence – and the potency – of an unspecific confidence reference also in the domain of perceptual learning. A classic example is the bisection task, in which observers have to judge whether the center line of three parallel lines is closer to the left or right line. When the stimulus is presented with a single difficulty level,

perceptual performance increases systematically with training (35–39). However, when different difficulty levels of the stimulus are presented in an interleaved manner, or when mixed with another type of stimulus, perceptual learning is abolished or at least reduced (36–38). Such impeded learning under 'roving' conditions has been confirmed for other stimulus types as well (40–43).

To explain these results, Herzog and colleagues have proposed that perceptual learning is based on a reward prediction error signal in which the reference reward is based on a running average of previous rewards (44). Rewards may either be based on explicit external feedback or on internal feedback (6). When stimulus conditions are presented in an interleaved manner, the running average reward ceases to be stimulus-specific, thereby corrupting the learning signal. The issue of maintaining a stimulus-specific reference for the computation of prediction errors may thus be a ubiquitous challenge for reinforcement learning signals, whether for internal or external feedback and across different task modalities.

Our notion that observers maintain an estimate of expected confidence is very similar to the idea of global confidence put forward in perceptual metacognition research (45–47). Whereas local confidence refers to the confidence of an observer in a single trial, global confidence refers to the estimate of one's performance over an entire task or session. This line of research found that humans indeed are able to compute an estimate of global confidence and that this estimate is based on an integration of confidence information across multiple decisions (47), as also proposed in this work. Moreover, local subjective confidence ratings were found to predict global self-beliefs over and above objective performance (45), emphasizing that global confidence is based on internal and not external performance evaluation.

Rouault and colleagues proposed a slightly different model of how humans maintain and update an estimate of global confidence, which fitted well with their empirical data. In this

model, global confidence is estimated in the form of a Beta distribution, a probability distribution defined on the interval [0, 1] which naturally represents the uncertainty of probability estimates (as in the case of performance confidence). In the absence of external feedback, the Beta parameters α and β were updated as follows: $\alpha = \alpha + \text{confidence}$, $\beta = \beta + (1 - \text{confidence})$. In this way, the mode $(\alpha - 1) / (\alpha + \beta - 2)$ of the Beta distribution continuously represented the best current estimate of global confidence, whereas the variance $\alpha\beta / [(\alpha + \beta)^2(\alpha + \beta + 1)]$ represented the uncertainty of this estimate. Yet, although the underlying mathematical framework by Rouault is thus different, conceptually the idea is quite similar to our Rescorla-Wagner estimate of expected confidence. Both models avoid storing all previous confidence experiences and instead maintain an efficient and continuously updated estimate of global/expected confidence based on either one (Rescorla-Wagner) or two (Beta distribution) variables.

The main differences are thus that 1) a model of type Rescorla-Wagner introduces a rate of forgetting such that more recent confidence experiences receive stronger weights, and 2) a Beta distribution model represents not only an estimate of the most likely value, but also the uncertainty of this estimate (hence two variables). Indeed, transferred to our model, a parallel estimate of the uncertainty in expected confidence naturally leads to the notion of precision-weighted prediction errors (48) and may be an interesting extension to our model: a confidence-prediction-error-based learning signal would then not only be influenced by the difference between current confidence and expected confidence, but also by the precision of our estimate of expected confidence; when we are yet very unsure about expected confidence, for instance in the beginning of a new task, learning signals would be weaker, thereby taking into account the lower reliability of computed confidence prediction errors.

A key parameter in our winning model *ConfUnspec* is the confidence-transfer parameter γ, which controls the degree to which confidence prediction errors affect subjective values when no external feedback is available. By contrast, in the case of external feedback, the update of subjective values is based on reward prediction errors and governed by the learning rate parameter α_r. Intriguingly, we found that both parameters are strongly correlated, such that participants with more volatile reward-based value learning also showed more volatile confidence-based value learning.

This finding fits well with our motivating hypothesis that learning based on external reward feedback and internal confidence-based feedback share a similar − perhaps the same − underlying mechanisms. The parameters γ and α_r thus may both characterize the tuning of one and the same learning machinery, observed in scenarios with and without external feedback. Together with the observed neurobiological parallel of learning based on internal and external feedback (10–12), the shared algorithmic logic of the respective learning signals (12,49,50), and the shared phenomenology (15), this parametric correspondence adds another piece of evidence to the view that confidence-based learning is based on an internally-triggered reinforcement learning mechanism.

Finally, our model and results may have an interesting implication for one of the most prominent and controversial effects in the decision-making literature − choice-induced preference changes (16,17,19). Here too, changes in subjective values are induced in the absence of external feedback, putatively caused by the mere act of the choice itself. Surprisingly, to our knowledge, almost no study has yet examined the role of choice confidence in choice-induced preference changes (for an exception, see 51). Indeed, taking Festinger's idea of cognitive dissonance as a cause of these preference changes seriously, it would predict a role of confidence that is in opposition to our model.

According to Festinger, subjective values are increased for chosen options (and decreased for unchosen options) as a form of post-hoc rationalization, to reduce the dissonance that would arise otherwise when reflecting on the positive attributes of an unchosen option. The larger the dissonance, the stronger the expected preference changes. Since the dissonance will be stronger for choices that are subjectively perceived as harder, those choices should be associated with a lower level of choice confidence. Thus, Festinger's theory predicts that higher choice confidence leads to higher preference changes for the chosen option, whereas our proposed model predicts the opposite (note however, that our model does not consider changes for the unchosen option). It will be an interesting avenue for future research to systematically investigate the interplay of choice confidence and subjective values changes and thereby clarify which prediction best passes the empirical evidence. Our finding of superior model evidence of confidence-based models (*ConfSpec* and *ConfUnspec*) over a choice-only model (*Choice*) certainly suggests that choice confidence is a key variable to consider in this question.

In conclusion, our study shows that confidence-based learning signals can explain significant dynamics of value-based decision making in the absence of external feedback, thereby extending previous findings in the specific domain of perceptual learning to one of the most fundamental forms of human learning: instrumental conditioning. Our results provide additional evidence that a previously suggested conceptual and algorithmic parallel between reward-based feedback and cognitive feedback (e.g., "correct'/'incorrect"; 30) should be extended to internal cognitive feedback – confidence – as well.

Methods

Ethics statement

Ethical approval for this study was granted by the ethics committee of Charité, Universitätsmedizin Berlin. Written informed consent was obtained from all participants prior to the experiment.

Participants

Sixty-six healthy volunteers (age: 29 ± 8.4 [s.d.]; gender: 40 female) were recruited via online advertisement and word of mouth. Participants were 18 or above and had normal or corrected to normal vision. Their participation was remunerated depending on performance (on average 16.25€). Two participants were excluded due to low task performance (<55% correct responses). The sample size calculation was based on a forward simulation. Choices and confidence ratings (based on the choice probability) were sampled from the generative models using the number of blocks and trials of the empirical experiment. We used educated guesses for all parameters ($\alpha_r = 0.1$, $\alpha_c = 0.1$, $\alpha_n = 0.1$, $\beta = 1/3$, $\gamma = 1$). The sample size was determined such that the model evidence (AIC) of all non-static models could be significantly dissociated from the static model with at least 80% probability (using a two-tailed paired t-test).

Mixed effects modeling

All analyses involving behavioral learning effects were performed with mixed effects models as implemented in the Python package *statsmodels* (for linear models; 52) and the *lme4* and *lmerTest* packages in R (for logistic models). *Subject* was a random effect and *block* a nested

472 random effect. Fixed effects were the block-level predictors *block_value_level* (18, 23 and 28,
473 i.e. the overall value level in a block), *block_difficulty* (3 or 6, i.e. the average absolute value
474 difference in a block), *block_stimulus_type* (0 or 1, i.e. stimulus types fractals or Chinese
475 symbols), block_ntrials_phase1 (duration of phase 1) and *block_ntrials_phase2* (duration of
476 phase 2). Trial-level predictors were *trial_number*, *trial_difficulty* (the absolute value
477 difference between the two CS in a trial) and *trial_value_chosen* (i.e., the value of the chosen
478 CS in a trial).

479 **Experimental task and procedure**

480 The instrumental conditioning task consisted of 11 blocks with an identical structure (Figure
481 1A). In each block, participants had to learn about the monetary values of five new
482 conditioned stimuli (CS). Each block started with an initial training phase (phase 1) of variable
483 length (9, 12, 15 or 18 trials) in which feedback was provided. The training phase was followed
484 by a critical second phase (5, 10 or 15 trials) without feedback. In two blocks, phase 2 was
485 omitted as a control condition. At the beginning of phase 2, participants were informed that
486 no feedback would be provided after choices, but also, that they would receive the associated
487 rewards at the end of the experiment. A block was completed by a third phase in which
488 feedback was again provided. The duration of phase 3 was such that, together with phase 1
489 and phase 2, each block comprised exactly 27 trials.

490 In each trial (Figure 1B), participants were presented with a choice between two CS on the left
491 and right of a fixation cross, respectively. To choose e.g. the left CS, participants moved the
492 mouse cursor to the left. The choice movement activated a 11-point confidence scale that
493 appeared under the chosen CS. The confidence scale consisted of 11 bars of increasing height
494 (maximum height for maximum confidence). Each bar was labeled with the respective rating

(0 to 10). In addition, the first and last bar, corresponding to the minimum and maximum confidence rating, were labeled with “Guessing” and “100% sure”. Higher confidence could be indicated by moving the mouse further to the left (or right, when the right CS was chosen), which highlighted all bars up to the respective confidence level. To make the choice/confidence experience more plastic, the CS increased in size proportional to the selected confidence. Participants could still switch their choice during the confidence selection by clicking the right mouse button, although this was rarely the case. When participants were satisfied with their response, they clicked the left mouse button. At this point, the unchosen CS disappeared and the chosen CS remained on the screen for 1000ms.

In phases 1 and 3, participants received monetary rewards for their choices. Rewards were presented in the form of a scratch ticket with 50 initially grey fields. The 50 fields were successively, but quickly, revealed such that each field was either a blank (in which case the field remained grey) or a hit (in which case a 1-Euro coin appeared on the field). We chose this reward presentation style – over a more conventional reward display with explicit numbers – to induce a mere “feeling” for the value of the CS rather than an explicit cognitive representation of rewards. The revealed scratch card remained on the screen for 500ms and then disappeared in an indicated slit below the card. The presentation in phase 2 was similar except that the fields of the scratch card were not revealed. At the end of the experiment, the overall reward was determined by means of 33 draws from an imaginary lottery box, which comprised all 1-Euro coins and blanks (including those from phase 2 which were initially not revealed) collected during the experiment. The average reward was 16.25€ (SEM 1.64€).

To avoid a learning transfer between blocks, different reward schedules were applied. First, each block was assigned one of three different overall average reward levels (18, 23 and 28€ per scratch card). Second, the mean value difference between CS in a block was either 3€ or

6€, which affected the average performance (3€: 68.8% correct; 6€: 77.1% correct). And third, in each block two CS were of identical value. Specifically, there were four different possible values per block to which the five CS were randomly assigned. Rewards were drawn from a truncated normal distribution with the given mean for a CS and a standard deviation of 10€. Since together, these conditions constitute more possible combinations than blocks, the conditions were pseudo-randomly distributed across the blocks. Similar to the variable phase durations, the main purpose was to prevent participants from learning about the task or reward structure and thus to enforce 'learning from scratch' in each block.

In half of the blocks, the CS were multicolor fractals, in the other half monocolor Chinese symbols. There was no meaningful performance difference between the stimulus types (fractals: 72.2% correct; Chinese symbols: 73.6% correct). The size of the CS was between 10.7 and 12.8 degrees of visual angle depending on the confidence level. All CS appeared roughly an equal number of times in each phase of a block.

Before and after phase 2, a rating scale appeared in which participants rated the subjective value of each CS in the current block on a continuous scale. The extremes of the scales were labeled with a scratch card of only blanks (lower end) and only 1-Euro coins (upper end). The scale itself was a horizontal bar with a color gradient from black (lower end) to gold (upper end). To select their rating, participants moved a thin sliding vertical bar across the rating scale (using the computer mouse).

The experiment was programmed in Python using PsychoPy (53). The experiment took place in a moderately lighted laboratory room in front of a computer screen (1920x1080 pixels, 47.7x26.8cm; viewing distance: 60cm). The entire experiment was operated by a computer mouse.

Model parameters and model fitting

The model was fitted for each subject individually, using all 11 blocks of the experiment. In the beginning of each block of the fitting procedure, the latent variables *expected value* $\bar{v}_i$ and *expected confidence* $\bar{c}_{(i)}$ were initialized to zero, given that new CS appeared in each block. The choice probability in each trial was computed via a softmax action selection rule (54):

$$p_{right} = \frac{1}{1 + e^{-\beta(\bar{v}_{right} - \bar{v}_{left})}} \quad (8)$$

$$p_{left} = 1 - p_{right} \quad (9)$$

where p_{right} and p_{left} are the choice probabilities for the CS left and right of the center, respectively. The slope β of the logistic function, also referred to as the *inverse decision noise parameter*, accounts for the stochasticity of choices. A value β = 0 implies that agents respond completely at random, whereas higher values of β indicate that agents choose more deterministically the CS associated with the highest expected value.

Importantly, the choice probability in Equation 8 was also used to determine the CS to which the confidence-value transfer (Equation 5) was applied during model fitting (CS_{right} if $p_{right} \geq 0.5$ else CS_{left}). Updating the CS actually chosen by the participants would have not been valid, as in this case the model would have had access to the same information it aims to predict.

Parameters were fitted by minimizing the negative log-likelihood (based on Equations 8 and 9) using the optimize.minimize() function of the Python SciPy package (55) in combination with an initial coarse-grained grid-search. We computed two optimization SciPy routines in parallel – the gradient-based L-BGFS-B algorithm (56) and the conjugate-direction-based

560 Powell algorithm (57) – and chose the parameters of whichever method resulted in a smaller
561 negative log-likelihood.

562 Table 2 provides an overview about the initial values and imposed bounds for all parameters.
563 The initial value was chosen as the average of the lower and upper bound. Note that while the
564 learning rate parameters α_r, α_c and α_d are bound to the range [0; 1], the *confidence transfer*
565 *parameter* γ is not a learning rate and thus has no natural upper bound.

Table 2. Free model parameters.

Parameter	Initial value	Lower bound	Upper bound
α_r	0.5	0	1
α_c	0.5	0	1
α_d	0.5	0	1
γ	50	0	inf
β	1	0	2

566 Note that the two new parameters of the model proposed here, the confidence parameters
567 α_c and γ, were largely uncorrelated (winning model: $r = -0.05$, $p = 0.712$), indicating that
568 neither of them was redundant.

569 # Acknowledgments

570 This study was supported by the grants GU 1845/1-1 and STE 1430/9-1 from the German
571 Research Foundation (DFG), a Clinical Fellowship to PS from the Berlin Institute of Health and
572 a Mind & Brain scholarship to LEP from the Berlin School of Mind & Brain, Humboldt-
573 Universität zu Berlin. We thank Yannick Schmidt for his assistance during the experiments.

Author contributions

LEP conceptualized the computational models, analysed and modelled the data and wrote the manuscript.

IS conceptualized the study and acquired the data.

PS conceptualized the study.

MG conceptualized the study and the computational models, analysed and modelled the data and wrote the manuscript.

All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Data availability

The experimental data, analysis scripts as well as computational models for the current study are available through the ConfLearning GitHub repository: https://github.com/eptas/ConfLearning. We used Zenodo to assign a DOI to this repository (DOI 10.5281/zenodo.5498343).

References

1. Sutton RS, Barto AG. Reinforcement Learning: An Introduction. Cambridge, MA: Bradford Books, MIT Press; 1998.
2. Gibson JJ, Gibson EJ. Perceptual learning; differentiation or enrichment? Psychol Rev. 1955 Jan;62(1):32–41.
3. McKee SP, Westheimer G. Improvement in vernier acuity with practice. Percept Psychophys. 1978;24(3):258–62.

4. Karni A, Sagi D. Where practice makes perfect in texture discrimination: evidence for primary visual cortex plasticity. Proc Natl Acad Sci U S A. 1991 Jun;88(11):4966–70.
5. Herzog MH, Fahle M. The role of feedback in learning a vernier discrimination task. Vision Res. 1997 Aug;37(15):2133–41.
6. Herzog MH, Fahle M. Modeling perceptual learning: difficulties and how they can be overcome. Biol Cybern. 1998;78:107–17.
7. Watanabe T, Náñez JE, Sasaki Y. Perceptual learning without perception. Nature. 2001 Oct;413(6858):844–8.
8. Seitz AR, Watanabe T. Is subliminal learning really passive? Nature. 2003;422(March):2003–2003.
9. Seitz AR, Watanabe T. A unified model for perceptual learning. Trends Cogn Sci. 2005 Jul;9(7):329–34.
10. Daniel R, Pollmann S. Striatal activations signal prediction errors on confidence in the absence of external feedback. NeuroImage. 2012 Mar;59(4):3457–67.
11. Daniel R, Pollmann S. A universal role of the ventral striatum in reward-based learning: Evidence from human studies. Neurobiol Learn Mem. 2014 May;114:90–100.
12. Guggenmos M, Wilbertz G, Hebart MN, Sterzer P. Mesolimbic confidence signals guide perceptual learning in the absence of external feedback. eLife. 2016;5:1–19.
13. Hebart M, Schriever Y, Donner TH, Haynes J-D. The Relationship between Perceptual Decision Variables and Confidence in the Human Brain. Cereb Cortex. 2016;26(1):118–30.
14. Garrison J, Erdeniz B, Done J. Prediction error in reinforcement learning: A meta-analysis of neuroimaging studies. Neurosci Biobehav Rev. 2013;37(7):1297–310.
15. Clos M, Schwarze U, Gluth S, Bunzeck N, Sommer T. Goal- and retrieval-dependent activity in the striatum during memory recognition. Neuropsychologia. 2015 Jun;72:1–11.
16. Festinger L. A Theory of Cognitive Dissonance. Stanford, California: Stanford University Press; 1957.
17. Brehm JW. Postdecision changes in the desirability of alternatives. J Abnorm Soc Psychol. 1956;52(3):384–9.
18. Chen MK. Rationalization and Cognitive Dissonance: Do Choices Affect or Reflect Preferences? Cowles Found Discuss Pap No 1669. 2008;
19. Chen MK, Risen JL. How Choice Affects and Reflects Preferences: Revisiting the Free-Choice Paradigm. J Pers Soc Psychol. 2010;99(4):573–94.
20. Coppin G, Delplanque S, Cayeux I, Porcherot C, Sander D. I'm no longer torn after choice: How explicit choices implicitly shape preferences of odors. Psychol Sci. 2010;21(4):489–93.
21. Coppin G, Delplanque S, Porcherot C, Cayeux I, Sander D. When Flexibility Is Stable: Implicit Long-Term Shaping of Olfactory Preferences. Martinez LM, editor. PLoS ONE. 2012 Jun 21;7(6):e37857.
22. Sharot T, Velasquez CM, Dolan RJ. Do decisions shape preference? Evidence from blind choice. Psychol Sci. 2010;21(9):1231–5.

638 23. Sharot T, Fleming SM, Yu X, Koster R, Dolan RJ. Is Choice-Induced Preference Change
639 Long Lasting? Psychol Sci. 2012;23(10):1123–9.
640 24. Nakamura K, Kawabata H. I Choose, Therefore I Like: Preference for Faces Induced by
641 Arbitrary Choice. PLoS ONE. 2013;8(8).
642 25. Johansson P, Hall L, Tärning B, Sikström S, Chater N. Choice Blindness and Preference
643 Change: You Will Like This Paper Better If You (Believe You) Chose to Read It!: Choice
644 Blindness and Preference Change. J Behav Decis Mak. 2014 Jul;27(3):281–9.
645 26. Koster R, Duzel E, Dolan RJ. Action and valence modulate choice and choice-induced
646 preference change. PLoS ONE. 2015;10(3):1–10.
647 27. Luo J, Yu R. The Spreading of Alternatives: Is it the Perceived Choice or Actual Choice
648 that Changes our Preference?: Perceived Choice and Actual Choice in our Preference. J
649 Behav Decis Mak. 2017 Apr;30(2):484–91.
650 28. Luettgau L, Tempelmann C, Kaiser LF, Jocham G. Decisions bias future choices by
651 modifying hippocampal associative memories. Nat Commun. 2020 Dec;11(1):3318.
652 29. Guggenmos M, Sterzer P. A confidence-based reinforcement learning model for
653 perceptual learning. BioRxiv; 2017.
654 30. Daniel R, Pollmann S. Comparing the neural basis of monetary reward and cognitive
655 feedback during information-integration category learning. J Neurosci. 2010
656 Jan;30(1):47–55.
657 31. Akaike H. Akaike, H. (1974). A new look at the statistical model identification. IEEE
658 Transactions on Automatic Control, 19(6), 716–723. doi:10.1109/tac.1974.1100705.
659 IEEE Trans Autom Control. 1974;19(6):716–23.
660 32. Anderson MC, Bjork RA, Bjork EL. Remembering can cause forgetting: Retrieval
661 dynamics in long-term memory. J Exp Psychol Learn Mem Cogn. 1994;20(5):1063–87.
662 33. Hulbert JC, Norman KA. Neural Differentiation Tracks Improved Recall of Competing
663 Memories Following Interleaved Study and Retrieval Practice. Cereb Cortex. 2015
664 Oct;25(10):3994–4008.
665 34. Wimber M, Alink A, Charest I, Kriegeskorte N, Anderson MC. Retrieval induces adaptive
666 forgetting of competing memories via cortical pattern suppression. Nat Neurosci. 2015
667 Apr;18(4):582–9.
668 35. Crist RE, Kapadia MK, Westheimer G, Gilbert CD. Perceptual Learning of Spatial
669 Localization: Specificity for Orientation, Position, and Context. J Neurophysiol. 1997 Dec
670 1;78(6):2889–94.
671 36. Otto TU, Herzog MH, Fahle M, Zhaoping L. Perceptual learning with spatial
672 uncertainties. Vision Res. 2006;46(19):3223–33.
673 37. Parkosadze K, Otto TU, Malania M, Kezeli A, Herzog MH. Perceptual learning of
674 bisection stimuli under roving: Slow and largely specific. J Vis. 2008 Jan 1;8(1):5–5.
675 38. Tartaglia EM, Aberg KC, Herzog MH. Perceptual learning and roving: Stimulus types and
676 overlapping neural populations. Vision Res. 2009;49(11):1420–7.
677 39. Aberg KC, Herzog MH. Does Perceptual Learning Suffer from Retrograde Interference?
678 Greenlee MW, editor. PLoS ONE. 2010 Dec 7;5(12):e14161.
679 40. Yu C, Klein SA, Levi DM. Perceptual learning in contrast discrimination and the

(minimal) role of context. 2004;20:169–82.
41. Kuai S-G, Zhang J-Y, Klein SA, Levi DM, Yu C. The essential role of stimulus temporal patterning in enabling perceptual learning. Nat Neurosci. 2005 Nov;8(11):1497–9.
42. Zhang J-Y, Kuai S-G, Xiao L-Q, Klein SA, Levi DM, Yu C. Stimulus Coding Rules for Perceptual Learning. Fahle M, editor. PLoS Biol. 2008 Aug 12;6(8):e197.
43. Banai K, Ortiz JA, Oppenheimer JD, Wright BA. Learning two things at once: differential constraints on the acquisition and consolidation of perceptual learning. Neuroscience. 2010 Jan;165(2):436–44.
44. Herzog MH, Aberg KC, Frémaux N, Gerstner W, Sprekeler H. Perceptual learning, roving and the unsupervised bias. Vision Res. 2012 May;61:95–9.
45. Rouault M, Dayan P, Fleming SM. Forming global estimates of self-performance from local confidence. Nat Commun. 2019;10(1):1–11.
46. Rouault M, Fleming SM. Formation of global self-beliefs in the human brain. Proc Natl Acad Sci. 2020 Nov 3;117(44):27268–76.
47. Lee ALF, de Gardelle V, Mamassian P. Global visual confidence. Psychon Bull Rev [Internet]. 2021 Mar 25 [cited 2021 Aug 11]; Available from: http://link.springer.com/10.3758/s13423-020-01869-7
48. Clark A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav Brain Sci. 2013 Jun;36(3):181–204.
49. Fleming SM, Massoni S, Gajdos T, Vergnaud J-C. Metacognition about the past and future: quantifying common and distinct influences on prospective and retrospective judgments of self-performance. Neurosci Conscious. 2016;1–12.
50. Diaz JA, Queirazza F, Philiastides MG. Perceptual learning alters post-sensory processing in human decision-making. Nat Hum Behav. 2017;1(january):0035.
51. Lee D, Daunizeau J. Choosing what we like vs liking what we choose: How choice-induced preference change might actually be instrumental to decision-making. PLOS ONE. 2020 May 18;15(5):e0231081.
52. Skipper S, Perktold J. statsmodels: Econometric and statistical modeling with python. In: 9th Python in Science Conference. 2010.
53. Peirce J, Gray JR, Simpson S, MacAskill M, Höchenberger R, Sogo H, et al. PsychoPy2: Experiments in behavior made easy. Behav Res Methods. 2019 Feb;51(1):195–203.
54. Wilson RC, Collins AG. Ten simple rules for the computational modeling of behavioral data. 2019;1–35.
55. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods. 2020 Mar;17(3):261–72.
56. Broyden CG. The Convergence of a Class of Double-rank Minimization Algorithms 1. General Considerations. IMA J Appl Math. 1970 Mar 1;6(1):76–90.
57. Powell MJD. An efficient method for finding the minimum of a function of several variables without calculating derivatives. Comput J. 1964 Jan 1;7(2):155–62.