

Learning and Generalization under Ambiguity: An fMRI Study

J. R. Chumbley^{1,2*}, G. Flandin¹, D. R. Bach¹, J. Daunizeau^{1,2}, E. Fehr², R. J. Dolan¹, K. J. Friston¹

1 The Wellcome Trust Centre for Neuroimaging, University College London, London, United Kingdom, **2** Laboratory for Social and Neural Systems Research, Institute of Empirical Research in Economics, University of Zurich, Zurich, Switzerland

Abstract

Adaptive behavior often exploits generalizations from past experience by applying them judiciously in new situations. This requires a means of quantifying the relative importance of prior experience and current information, so they can be balanced optimally. In this study, we ask whether the brain generalizes in an optimal way. Specifically, we used Bayesian learning theory and fMRI to test whether neuronal responses reflect context-sensitive changes in ambiguity or uncertainty about experience-dependent beliefs. We found that the hippocampus expresses clear ambiguity-dependent responses that are associated with an augmented rate of learning. These findings suggest candidate neuronal systems that may be involved in aberrations of generalization, such as over-confidence.

Citation: Chumbley JR, Flandin G, Bach DR, Daunizeau J, Fehr E, et al. (2012) Learning and Generalization under Ambiguity: An fMRI Study. *PLoS Comput Biol* 8(1): e1002346. doi:10.1371/journal.pcbi.1002346

Editor: Tim Behrens, University of Oxford, United Kingdom

Received: July 21, 2011; **Accepted:** November 25, 2011; **Published:** January 19, 2012

Copyright: © 2012 Chumbley et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: JC was funded by the SystemsX, CoMPLEX (UCL) and the Wellcome Trust. This work was supported by Wellcome Trust Program Grants to KJF and RJD. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: j.chumbley@fil.ion.ucl.ac.uk

Introduction

Successful behavior in new situations often requires us to apply ‘rules-of-thumb’. However, acquiring and applying abstract rules from limited experience presents a fundamental computational problem [1]: in which both over- or under-generalization must be avoided [2,3,4,5,6]. Despite their importance, little is known about how neuronal systems learn these rules, and how the delicate balance between past and present information is maintained. Evolutionary arguments suggest that the use of previously learned rules when generalizing to new situations increases adaptive fitness by optimizing behavior [7]. This raises the key question of whether and how generalization is optimized [8]. In this work, we examine whether human subjects combine previously learned rules and current information in an optimal way and identify the brain systems that underlie this combination. Using Bayesian learning theory to specify optimal generalization, we looked for its neural correlates. In particular, we drew on existing evidence that points to the hippocampus as a key structure that is implicated in learning the specifics of a new situation, when previously learned rules may not apply [9,10].

Probabilistic inference in a natural environment is confounded by multiple sources of uncertainty [11,12,13,14], including objective randomness and subjective ignorance [14]. Uncertainty is a key concept here because the confidence about prior beliefs should be weighed against the confidence about new information, when deciding whether to generalize those beliefs to a new situation. Classical reinforcement learning models (e.g. [15,16]) do not represent uncertainty or use generalization to guide learning and behavior: these schemes simply learn the expected value of action-states and only prosper in environments where the current state is sufficient to specify a successful action: see [17] for a

critique and extension. Having said this, several other RL schemes are based on some form of non-probabilistic function approximation and therefore support generalization (see Chapter 8 in [18] for discussion and recent RL approaches in neuroscience that consider generalization in the spatial [19] and temporal [20] case). While recent RL developments in neuroscience incorporate some notion of uncertainty [21], learning and generalization are typically non-probabilistic. In this work we ask if learnt generalizations are accompanied with due uncertainty [22], as prescribed by probability theory.

At the behavioral level, human subjects readily abstract probabilistic rules and use them to generalize [8]. Furthermore, they can distinguish different sources of uncertainty: the unavoidable or irreducible randomness of certain events versus subjective ignorance about the world [12,13,23,24,25]. The latter resembles the concept of subjective *ambiguity* in economics and represents uncertainty about objective risks. For example, the *risk* (or irreducible randomness) associated with a *fair* coin toss is high (50:50); however, there may be subjective ambiguity as to whether the coin is itself fair. This paper examines the function and mechanisms of generalization in the face of ambiguity. While there are good reasons to restrict the term ambiguity to *complete* ignorance [26], we use the term more inclusively to denote the level of uncertainty about the outcome probabilities. This is akin to *estimation* [14] or *second-order* [26] uncertainty (i.e., uncertainty about uncertainty). Ambiguity is subjective and reference-dependent: it ranges from complete ignorance to near certainty and, crucially, can be reduced by generalization in a Bayes-optimal fashion [8]. In other words, if subjects consider their current situation in the light of past experience, they can exploit similarities between the past and present to reduce their ambiguity [27,28]. In our example, ambiguity about a new coin will be

Author Summary

Intelligent behavior requires flexible responses to new situations, which exploit learned principles or abstractions. When no such principles exist, the imperative is to learn quickly from scratch. Behaviorally, we show that subjects learn action-reward relationships in a manner that enables them to generalize rules to new situations. Our fMRI results show that when subjects have no evidence that such a rule exists, medial temporal lobe responses (that reflect uncertainty) predict their augmented learning.

reduced by observing the random behavior of similar coins. This ability to generalize over similar situations is seen readily in behavior and learning [8,14].

In this study, we examined the neuronal correlates of generalization with a special focus on the hippocampus: The hippocampus is involved in generalization [29,30,31,32,33,34] and shows activations that are sensitive to objective uncertainty or risk [23,35]. In this paper, we asked if hippocampal responses also report subjective uncertainty or ambiguity that changes with experience. Specifically, we tested for ambiguity-dependent hippocampal responses, when probabilistic nature of outcomes had to be learned. Furthermore, we hoped to show behaviorally that learning rates were greater in contexts that had more ambiguity. We addressed these questions using a model of our experimental task and, tested whether Bayesian updates or learning could explain behavioral and neurophysiological responses, as measured with fMRI.

Materials and Methods

Subjects and procedure

Nineteen subjects (age 19–31, 11 female) were recruited from the UCL psychology Dept subject pool. All subjects gave informed consent, before reading a brief description of the task which was then performed under fMRI. The study protocol was approved by the local UCL ethics committee.

Image acquisition and analysis

Image acquisition. Images were acquired on a 3 T Allegra head scanner (Siemens Medical Systems) with a head coil for RF transmission and signal reception. We used BOLD signal sensitive T2*-weighted transverse single-shot gradient-echo echo-planar imaging (EPI; flip angle 90°; bandwidth BW, 3551 Hz/pixel; phase-encoding (PE) direction, anterior–posterior; bandwidth in PE direction BWPE, 47.3 Hz/pixel; TE, 30 ms; effective TR, 2600 ms). An automatic 3D-shim procedure was performed at the beginning of each experiment. Each volume contained 40 slices of 2-mm thickness (1-mm gap between slices; field of view, 192×192-mm²; matrix size, 64×64). Sensitivity losses due to susceptibility artifacts were minimized by applying a *z*-shim gradient moment of 0.4 mT/m, a slice tilt of 30°, and a positive PE gradient polarity [36], [37]. Each subject underwent one scanning session, with three breaks. The task was self-timed, and therefore the duration of each session depended on the subject. The first five volumes of each session were discarded to ensure steady-state longitudinal magnetization.

Whole-brain anatomical scans were acquired using a modified driven equilibrium Fourier transform (MDEFT) sequence with optimized parameters [38]. One hundred seventy-six sagittal partitions were acquired with an image matrix of 256×224 (read×phase) and twofold oversampling in read direction (head/foot direction) to prevent aliasing (isotropic spatial resolution 1-

mm;15°; TR/TE/TI, 7.92 ms/2.4 ms/910 ms; BW, 195 Hz/pixel). Spin tagging in the neck was performed to avoid flow artifacts in the vicinity of blood vessels. The flip angle of the tagging pulse was chosen to be 160° to account for B1 losses in the neck. Special RF excitation pulses were used to compensate for B1 inhomogeneity of the transmit coil in superior/inferior and anterior/posterior directions. Images were reconstructed using a standard 3D Fourier Transform, followed by modulus calculation.

Image analysis. Functional imaging data were analyzed with statistical parametric mapping (SPM8; Wellcome Trust Centre for Neuroimaging; www.fil.ion.ucl.ac.uk/spm). EPI images were generated off-line using a generalized reconstruction method based on the measured EPI *k*-space trajectory to minimize ghosting. Motion-corrected images were co-registered to the individual's anatomical MDEFT image and spatially normalized to the Montreal Neurological Institute T1 reference brain template (re-sampled voxel size: 2×2×2-mm).

The experimental paradigm

While our goal was to identify domain-general computational processes, the paradigm was framed as a social inference task: Subjects were told that two *groups* of thirty *individuals* had completed a marketing survey. Subjects were then asked to guess, over ten consecutive trials, whether each individual would choose a 'blue' or 'purple' product. Subjects were told they would be paid 'in proportion to the number of correct guesses' and that the two groups were 'geographically and economically unlike one another'. Trial cues (individuals) were faces from the Sterling data-set, whose group membership was indicated by the symbol '*' or 'o' (see Figure 1). Each trial comprised the following sequence: 1) an individual's face was presented along with the symbol indicating their group membership; 2) the response options (blue and purple squares) were then presented, after which 3) the subject responded and 4) received feedback about whether their guess was correct or incorrect. The timeline for a single trial is shown in Figure 1. If subjects did not guess within one second, they were shown the instruction 'ACT FASTER!'. The subject's guess was highlighted until feedback was delivered. Correct guesses were signaled with an auditory beep (500 milliseconds of 500 Hz sine wave) and accumulated in a score bar at the bottom of the screen. Incorrect guesses were indicated by a 500 millisecond burst of white noise (with no increase in their score).

Unbeknown to subjects, individuals from one group had similar preferences, while the other group had more between-individual variability. This meant that subjects had to make guesses about choices in two distinct contexts established by the group an individual belonged to: in the *generalization context* (GC), all individuals chose 'purple' with probability $p=0.8$. In the *ambiguous context* (AC), 'blue' was probabilistically chosen ($p=0.8$) by half of the group members and 'purple' ($p=0.8$) by the other half. To reiterate, subjects were presented with the same face ten times and had to guess whether the individual preferred blue or purple. Each individual was identified as belonging to one group or the other. Every individual preferred one color that was chosen 80% of the time. In the generalization context, all group members preferred the same color, while in the ambiguous context, individual group members preferred blue or purple with equal probability. In both contexts, subjects could learn about any given individual over ten trials.

The generalization context therefore contained a probabilistic rule prescribing the best guess, even in the absence of learning about an individual's preferences. Conversely, in the ambiguous context, subjects had to learn about individual preferences because their group membership provided no clues about what they would

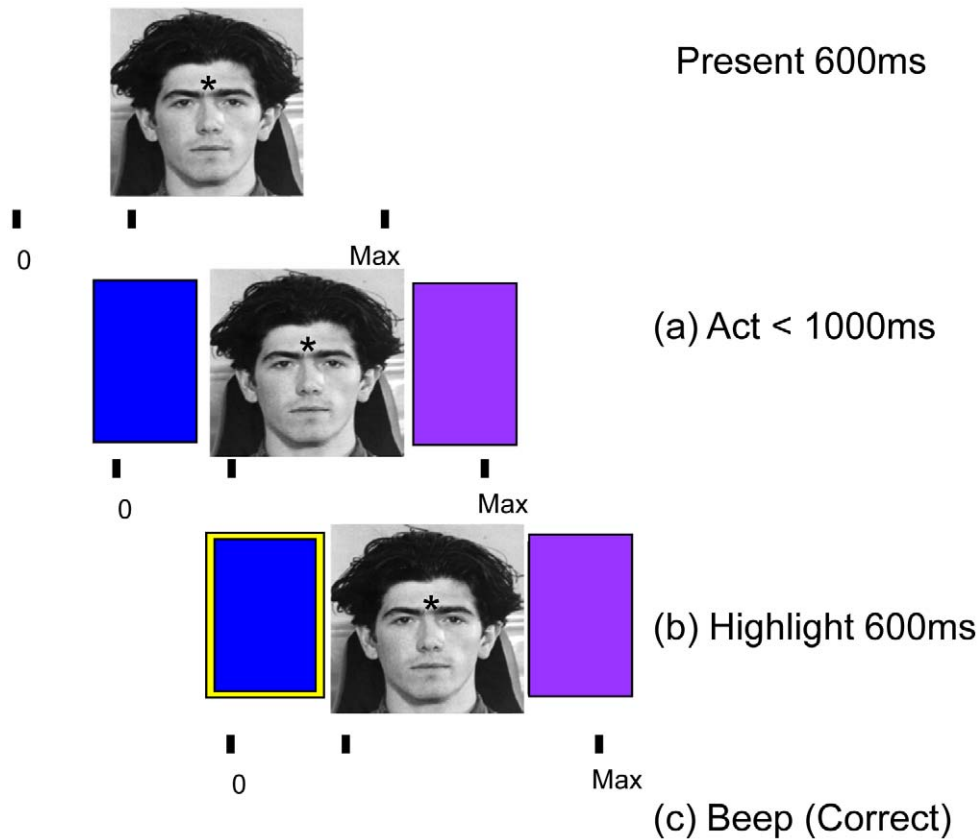


Figure 1. This schematic shows the structure of each trial. A face was presented for 600 ms before two choice options were displayed. The choice options cue the subjects' guess, which was then indicated by a yellow border around the selected option. Audio and visual feedback indicated whether the choice was rewarded (correct) or not (incorrect). doi:10.1371/journal.pcbi.1002346.g001

preferentially choose. Trials were arranged into blocks, in which the same individual was presented for ten consecutive trials. The blocks alternated between AC and GC, with a new individual (face) for each block. This resulted in $30 \times 10 \times 2 = 600$ trials, for thirty individuals, presented ten times for two groups.

The blue and purple options were presented with equal probability on the left and right of the screen on each trial. Individuals (faces) were randomly reassigned to either group, between subjects. All subjects experienced the same feedback contingencies (with randomly reassigned cues). Subjects had three short breaks during the task: for each they were first cued 'PLEASE HAVE A SHORT REST AND RELAX' before being prompted to restart thirty seconds later: 'OK! PLEASE PRESS ANY KEY TO CONTINUE'.

Bayesian modeling versus conventional fMRI analyses

Bayesian learning theory predicts that subjects should learn more quickly about a new individual from the ambiguous group, relative to an individual from the generalization group. This is based upon the assumption that subjects are making Bayes-optimal guesses using a notion of group or context. The increase in learning rate with higher levels of ambiguity is related to increases in learning rate in situations with a high degree of volatility [39] (see below). At the neuronal level, we predicted that increases in learning rate would selectively engage hippocampal processing in the ambiguous context. In other words, hippocampal activation should track changes in ambiguity about an individual's preference as it alternates between AC (high ambiguity) and GC (low

ambiguity) blocks. To quantify ambiguity, we assumed subjects were ideal Bayesian observers who used a model of probabilistic outcomes. We focused on two alternative models to predict subject responses, M1 and M2. Under M1, Bayesian learning combines new information with existing generalizations based on group membership. Conversely, M2 accumulates information about every individual independently, without the benefit of generalization.

To make optimal guesses about the choices of each group member, subjects have to infer their preferences i.e. the probability that this individual will choose a particular option, say 'purple'. We denote this probability with $\theta_i \in [0, 1]$. The information following each trial is equivalent to observing the outcome of a biased coin. We use the random variable o_{n_i} to denote whether the choice of the i^{th} individual was 'purple' ($o_{n_i} = 1$) or 'blue' ($o_{n_i} = 0$): $i \in 1, \dots, I$ in trial $n_i \in 1, \dots, 10$ (subjects encountered $I = 30$ individuals in each of the two groups).

In what follows, we consider alternative models that subjects might have used to infer the $\theta_i \in [0, 1]$. We start with a model that permits generalization and then turn to a version that precludes generalization. We also consider a few alternative models that can be considered as special cases that are of interest from an RL perspective.

Models

M1: Bayes-optimal generalization. The critical feature of M1 is that guesses about each individual are informed by knowledge about group membership. This model supposes that

subjects *jointly* learn about all individuals in a given group. In the generalization context, subjects should be more confident about a new individual from the unambiguous group, relative to the ambiguous group that provides no contextual clues. This differential uncertainty (ambiguity) is our focus. For simplicity, we assumed that subjects generalize within, but not between, groups. In other words, learning in one context was independent of learning in the other. An additional hierarchical level would permit generalization across contexts (e.g., the relative size of each group) and could be modeled with an extension of the Bayesian framework described below [40].

The form of our model appeals to behavioral evidence that human rule learning resembles non-parametric Bayesian inference [41]. It is also related to a previous [42] Bayesian formulation of rule-learning. (While the latter model focuses on Pavlovian learning, it resembles M1 through inferring the hidden number of subgroups or ‘latent causes’ [42]). In our model, subjects represent the (preferences of) individuals, $\{\theta_1, \dots, \theta_I, \theta_{NEW}\}$ where I is the total number of individuals encountered so far. Subjects represent individual preferences by assigning individuals to subgroups, according to their similarity. Note that while subjects observe group membership, subgroup membership is hidden: There are two hidden subgroups in the ambiguous context, preferring either blue or purple, but only one in the generalization context. By first finding the number and nature of subgroups, optimal Bayesian assignment avoids over-generalization (e.g. incorrectly labeling a new blue-preferring individual as belonging to a known purple-preferring subgroup) and under-generalizing (e.g. failing to recognize that a new purple-preferring individual belongs to a known purple-preferring subgroup). This type of learning has had considerable success in modeling category learning in humans [41] and ‘rationalizes’ non-Bayesian models of generalization in reinforcement learning [43] (see below).

We assume that subjects store the number of times (out of N_i) the i^{th} individual chose ‘purple’. The cumulative counts up to the present trial t , are denoted by $d_i = \{(y_i, N_i - y_i)\}_{i=1}^I$ where $y_i = \sum_{n_i=1}^{N_i} o_{n_i}$. Subjects model their cumulative observations y_i as drawn from a mixture of Binomial distributions of the form $Bin(y|N, \theta)$. Being ignorant of the mixing distribution θ , we assume they use a Dirichlet process G over θ , with concentration parameter α and base distribution G_0 [corresponding to the uninformative conjugate Beta distribution $Beta(1,1)$]. These define the base measure αG_0 . The resulting probabilistic model is

$$\begin{aligned} y_i | \theta_i &\sim Bin(N_i, \theta_i) \\ \theta_i | G &\sim G \\ G | G_0, \alpha &\sim DP(G_0, \alpha) \end{aligned} \tag{1}$$

Here ‘ $R \sim S$ ’ means R has the distribution S ; so the right hand side specifies a distribution. The Dirichlet process, DP, is thus a distribution on distributions and models ambiguity. Because realizations of a DP are discrete with probability one, these models can be viewed as probability measures consisting of a weighted sum of point masses [44,45]; i.e., countably infinite mixtures

$$G = \sum_{k=1}^{\infty} \pi_k \delta(\theta_k^*) \sim DP(G_0, \alpha) \tag{2}$$

Here $\delta(x)$ is a point mass at a single point x , π_k is a stick-breaking process and θ_k^* is distributed as G_0 [46].

The implicit form of generalization is more transparent when we integrate over G to obtain a prior over θ_i in terms of successive conditional distributions (see [46] for a measure-theoretic proof of this integral)

$$\theta_i | \theta_1, \dots, \theta_{i-1} \sim \frac{1}{i-1+\alpha} \left[\sum_{j=1}^{i-1} \delta(\theta_j) + \alpha G_0 \right] \tag{3}$$

This means the prior belief about one individual θ_i depends on knowledge about others sampled from the population, $\sum_{j=1}^{i-1} \delta(\theta_j)$, as well as the initial distribution G_0 . This completes our description of M1 in terms of a likelihood (in Eq. 1) and prior (in Eq. 2/3).

To predict subject’s responses we require M1’s posterior belief about the behavioral contingencies. This quantifies the ambiguity as well as the value of their response options. For posterior inference, one can obtain a sample from the posterior of $\theta = \{\theta_1, \dots, \theta_I\}$ by simulating a Markov chain whose equilibrium distribution is the desired posterior distribution [47]. The simplest approach is to repeatedly sample θ_i from its conditional distribution, given both the data and all other $\theta_{j \neq i}$, denoted by θ_{-i} . This distribution therefore combines the likelihood of θ_i and the prior, conditional on θ_{-i} . This conditional prior for an individual based on previous individuals is given by

$$\theta_i | \theta_{-i} \sim \frac{1}{I-1+\alpha} \left[\sum_{j \neq i} \delta(\theta_j) + \alpha G_0 \right] \tag{4}$$

and derives from the previous equation by noting that i is the last of I observations (i.e. by assuming the θ_i are exchangeable). Introducing the likelihood, this yields the following conditional posterior distribution:

$$\theta_i | \theta_{-i}, y_i \sim b \sum_{j \neq i} Bin(y_i | N_j, \theta_j) \delta(\theta_j) + b \alpha \int Bin(y_i | N_i, \theta) dG_0(\theta) Q_i \tag{5}$$

Where Q_i is the posterior over θ_i , based on the prior G_0 and the single observation y_i with likelihood $Bin(y_i | N_i, \theta_i)$, i.e. $Beta(\theta_i | 1 + y_i, 1 + N_i - y_i)$. Here, b is chosen to ensure that

$$b \left[\sum_{i \neq j} Bin(y_i | N_j, \theta_j) + \alpha \int Bin(y_i | N_i, \theta) dG_0(\theta) \right] = 1.$$

This Bayesian model is related to the non-Bayesian RL model of [43] mentioned in the introduction. In that RL model, each cue is first ‘classified’ before reinforcement learning. A cue is either assigned to a known class of cues based on similarity, or designated exceptional and given its own class. Both perceptual similarity and predictive similarity play a role: do two cues look the same? do they predict the same outcomes? Regarding the latter, *negative prediction errors* from RL reduce perceived similarity between cues in a separate *recognition system*, thereby promoting discrimination over generalization [43]. Our focus is on this predictive similarity. To derive *optimal* generalization, we define predictive similarity as the likelihood of an outcome, given a cue (rather than the inverse magnitude of a negative prediction error). In particular, a cue’s past associations determine if it will be assigned to a known class based on similarity, defined by $Bin(y_i | N_j, \theta_j)$, or assigned to its own class with probability $b \alpha \int Bin(y_i | N_i, \theta) dG_0(\theta)$. The hyperparameter α controls this tradeoff between generalization and discrimination and can itself be learned [48].

Having assumed G_0 is an uninformative Beta distribution, $Beta(\theta|1,1)$, which is conjugate to the likelihood, calculating the integral $\int Bin(y_i|N_i, \theta) dG_0(\theta)$ and sampling from Q_i are straightforward. The simplest algorithm [47] for Gibbs sampling from the full posterior $p(\theta_1, \dots, \theta_I, \theta_{NEW}|d_i)$ including θ_{NEW} is (see [47,48] for further details):

- For $i=1, \dots, I$:
 - Draw a new value from $\theta_i|\theta_{-i}, y_i$ as defined above.
 - With probability $\frac{\alpha}{I-\alpha}$ draw θ_{NEW} from the base distribution G_0 .
 - Otherwise, uniformly draw one existing $\{\theta_i : i=1, \dots, I\}$ and assign its value to θ_{NEW} .

This procedure approximates the trial-by-trial evolution of posterior belief about preferences, $p(\theta_1, \dots, \theta_I, \theta_{NEW}|d_i)$. Once the Markov chain has reached equilibrium, we use a sample of size S from that distribution. Furthermore, any marginal posterior of interest $p(\theta_i|d_i)$ or $p(\theta_{NEW}|d_i)$ is approximated simply as the univariate component of this joint sample [49].

We used two measures of this time-dependent posterior as explanatory variables to predict the behavioral and neurophysiological responses of each subject. Firstly, we operationalized the ambiguity about each new individual using the Shannon entropy $H_i = H(p(\theta_{NEW}|d_i))$. To evaluate this entropy, univariate samples from $p(\theta_{NEW}|d_i)$ were first binned into $L=20$ bins to provide an approximate discrete probability mass function with $\hat{p}_l = \#\{\theta_{NEW} \in l^{th} \text{ bin}\}/S$. In this case, we have $H \approx -\sum_{l=1}^L \hat{p}_l \log \hat{p}_l$. This uncertainty measure characterizes the ambiguity about a new individual, as generalized from experience with other individuals in the group. A differential entropy between AC/GC reflects both the number of subgroups (clusters) and uncertainty about the parameters describing those subgroups. To see this more clearly take an extreme case where all I individuals encountered so far have been attributed to one subgroup, $\theta_1 = \theta_2 = \dots = \theta_I = c$. This subgroup, characterized by a parameter setting of c , is therefore strongly favored as an *a priori* explanation for new individuals (i.e. strong generalization). Applying this condition to Eq. 3 gives

$$\theta_{NEW}|\theta_1, \dots, \theta_I \sim \frac{1}{I+\alpha} [I\delta(c) + \alpha G_0]$$

Intuitively, $\frac{I}{I+\alpha}$ probability mass now rests on just one point mass (for small α this is all the mass). This predictive distribution is therefore less ‘dispersed’ than if there were two or more subgroups (i.e. it has lower entropy). For this reason, when H_i is small, learning is more strongly biased towards belief in one subgroup. In this sense, the entropy can be regarded as a proxy for ambiguity that dictates the ‘learning rate’ or the sensitivity to new information. We used this entropy measure to identify the neurophysiological correlates of ambiguity, using fMRI responses.

We have emphasized that greater prior ambiguity (i.e. higher number of inferred subgroups/higher predictive entropy) is accompanied by a diminished *a priori* bias. This affords observations more influence over posterior belief. Another influential hypothesis is that uncertainty influences choice by promoting exploration itself [50,51] i.e. what to learn about *vs.* how fast to learn in the current situation. To simplify things, we chose a task with no exploration-exploitation trade-off. Specifically, because every trial in our task provides feedback on the value of the chosen action *and* counterfactual information about the other unchosen action, there is no information to be gained from exploring the less valuable action.

Subjects do not know the true expected reward, $\theta_{i,TRUE}$, for choosing ‘purple’ when faced with the i^{th} individual (they do not know that individual’s preference). Let $\eta_{n_i+1}^{M1}$ then denote their

subjective, expected reward for guessing ‘purple’ on the n_i+1^{th} trial faced with the i^{th} individual, following a total of t trials under M1. This expectation is defined by weighting possible values of θ_i according to their current plausibility, giving $\eta_{n_i+1}^{M1} = \int \theta_i p(\theta_i|d_i) d\theta_i$. This is just the posterior expectation of θ_i and can be approximated by

$$\eta_{n_i+1}^{M1} \approx \frac{1}{S} \sum_{s=1}^S \theta_{i,s} \tag{6}$$

Here, each $\theta_{i,s}$ is an MCMC sample from the posterior $p(\theta_i|d_i)$ conditional on all observations to date. This replaces an analytic expectation with an empirical expectation (converging according to the law of large numbers). An exactly analogous approximation yields the predicted value for a new cue:

$$\eta_{NEW,1}^{M1} = \int \theta_{NEW} p(\theta_{NEW}|d_i) d\theta_{NEW} \approx \frac{1}{S} \sum_{s=1}^S \theta_{NEW,s} \tag{7}$$

We now turn to some alternative models.

M2: Rescorla-Wagner without generalization. To assess the predictions of M1 in relation to a null model, we also considered the predictions under M2, where subjects learn about each individual without generalization. Under this assumption, the expected reward (correct choice) can be modeled with classical Rescorla-Wagner learning [15].

$$\begin{aligned} \eta_{n_i+1}^{M2} &= \eta_{n_i}^{M2} + \alpha \delta_{n_i} \\ \delta &= o_{n_i} - \eta_{n_i}^{M2} \\ \text{and} \\ \eta_{NEW}^{M2} &= 0.5 \end{aligned} \tag{8}$$

Where o_{n_i} is still the binary outcome on trial n with the i^{th} cue. One implementation of M2 - akin to habit learning - would be to separately initialize the value of guessing ‘purple’ or ‘blue’ ($\eta_{PURPLE,t}$, $\eta_{BLUE,t}$) to zero and update each only when the corresponding action was taken [52]. Guesses could then be modeled according to $p(PURPLE) \propto \exp(\lambda \eta_{PURPLE,t})$, where λ controls the stochastic precision of the guess. However, because subjects are told that exactly one option is correct, each outcome is informative about the counterfactual (unchosen) option. We therefore initialized η_{n_i} (the value of the purple guess on trial n in the presence of each cue i) to $1/2$, and defined the value of the blue choice as $1 - \eta_{n_i}$. Subsequent outcomes $o_{n_i} \in \{0,1\}$ push $\eta_{n_i}^{M2}$ up or down as specified by M2. This agent therefore uses counterfactual data (from the unchosen option), but does not generalize between individuals. We fit the free ‘learning rate’ parameter α by

minimizing the error function, $E(\alpha) = \sum_{i=1}^I \sum_{n_i=1}^{N_i} (\eta_{n_i}(\alpha) - x_{n_i})^2$,

where $x_{n_i} \in \{0,1\}$ indicates which option the subject guessed on the corresponding trial (coded as $PURPLE=1, BLUE=0$). $E(\alpha)$ was evaluated numerically for different values of α with increments of 0.001 within parameter space.

M3–M5: Additional models. The resulting sequence of value for our two models η^{M1}, η^{M2} are plotted as a function of trial number in the dashed and solid curves of Figure 2, for the sequence of face cues and outcomes presented to our subjects (i.e. the sequence of blue/purple choices made by each individual). These correspond to the value of guessing purple under a model

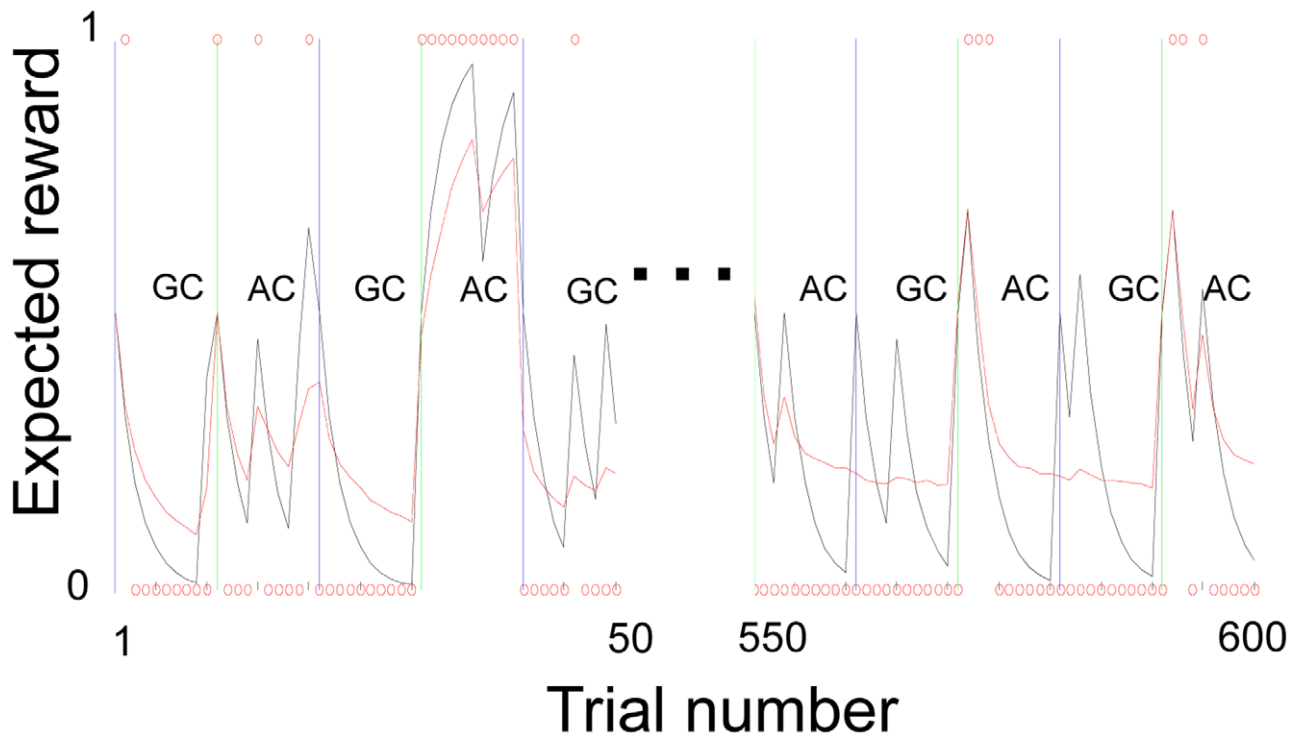


Figure 2. Subject's predicted value (expected reward) for guessing 'purple', according to model-based (M1, red-dashed) and model-free (M2, black) schemes. M1 tracks current information when necessary (AC), and otherwise exploits generalization to limit the impact of spurious outcomes on action (GC). M2 is ignorant about each new individual and myopically chases reward. Red circles indicate the actual guesses of a typical subject.

doi:10.1371/journal.pcbi.1002346.g002

with (M1) and without (M2) the facility for generalization. For ease of visualization, Figure 2 and 3 interpolate discrete-time model predictions to form smooth curves. The vertical lines demarcating the context (ambiguous or generalization) are centered on the first trial of each block.

To ensure we had not overlooked other explanations for the subjects' responses, we performed secondary analyses, to establish the explanatory power of M1 in the context of alternative models, M3–M5. M3 was a generalization of M2 [15], which represents and learns the value of contextual (group-membership) cues. This agent therefore represents 62 cues (60 faces, 2 contextual cues). On each trial, M3 calls and updates both the context (group) and individual (face) cues presented on that trial. Defining o_t as the outcome on trial $t = 1, \dots, 600$ and $\eta_{i,t}^{M3}$ as the instrumental value of choosing purple, faced with the i^{th} cue, updates were implemented with

$$\begin{aligned} \eta_{i,t+1}^{M3} &= \eta_{i,t}^{M3} + \alpha \delta_{i,t} \\ \delta &= o_{n_i} - \sum_i I(i) \eta_{i,t}^{M3} \end{aligned} \quad (9)$$

where the indicator $I(i)$ is 1 whenever the i^{th} cue is present and 0 otherwise. Only values for cues actually present in trial t are updated. Like M2, this agent uses counterfactual information (from the unchosen option). Specifically, the value of choosing 'purple' on trial n was $\sum_i I(i) \eta_{i,t}^{M3}$ and the value of choosing 'blue' was $1 - \sum_i I(i) \eta_{i,t}^{M3}$. Each $\eta_{i,t}^{M3}$ was initialized to 0.25 so that prior to learning 'blue' or 'purple' were equally valuable; i.e., $0.5 = \sum_i I(i) \eta_{i,t}^{M3}$. The free parameter α was fit to each subject's guesses using the same procedure as for M2. M4 modeled a

Bayesian learner that *over-generalizes*. It has an identical mathematical form to M1 but unlike M1 does not distinguish between contexts. It treats all individuals from the two contexts/populations indiscriminately i.e. as part of one 'meta-population'. M5 modeled a Bayesian learner that *under-generalizes*; In other words, it can represent uncertainty but cannot generalize. M5, like M1 and M4, models observations associated with any one individual as $y_i | \theta_i \sim \text{Bin}(N_i, \theta_i)$, but differs in the prior. In particular, each individual is treated independently with no generalization within or between groups (the prior over individuals factorizes). Specifically, we use a Beta prior that resets the prior for the i^{th} cue to uniform $\text{Beta}(\theta_i | 1, 1)$, irrespective of its experience with other cues. This agent shares a key feature of M2 - resetting the predictions for each new cue to 0.5 and learning without generalization. For subsequent trials, it calculates the expected value of choosing purple as

$$(y_i + 1) / (N_i + 2) \quad (10)$$

where y_i is still the count of correct purple choices with cue i (see above and [49]). In practice this agent's predictions are similar to M2.

Relating model predictions to data

Behavior. We used logistic regression to predict trial-by-trial choices from the value (expected reward) based on M1, while including the value derived from models M2–M5 as additional nuisance covariates. We calculated within-subject point estimates of the partial regression coefficients of M1 predictions, before testing for significant (nonzero) effects at the between-subject level using standard classical statistics.

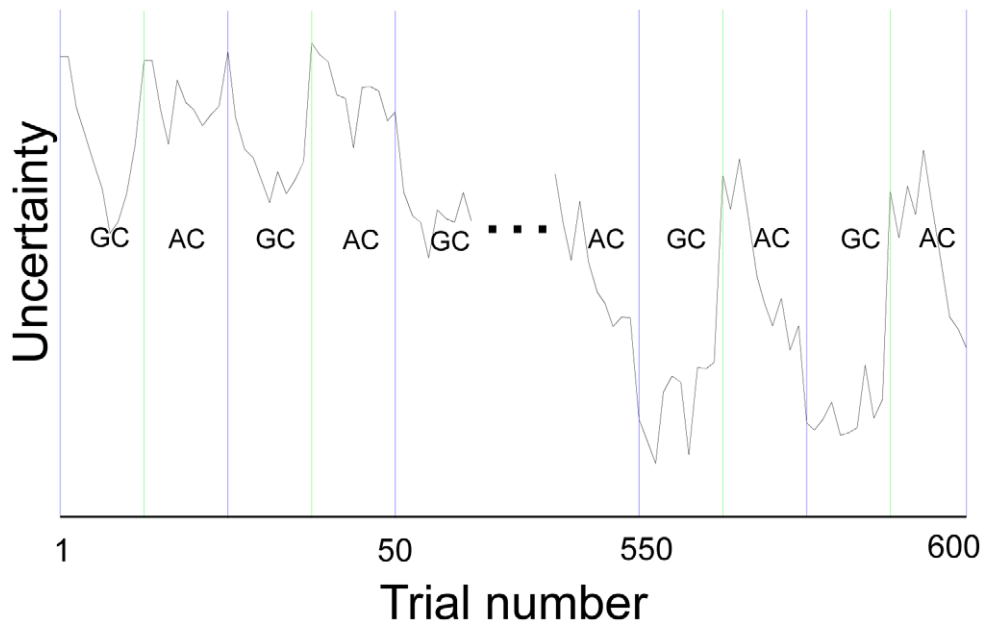


Figure 3. Degree of learning formalized as prior uncertainty about the reinforcement contingencies, in M1. Without evidence of a contextual norm (in the AC) subjects are uncertain about what to do with an unfamiliar person, and must learn quickly. This time-series, convolved with a hemodynamic response function predicted hippocampal fMRI responses (see main text). doi:10.1371/journal.pcbi.1002346.g003

fMRI. Our analyses of the fMRI data used a conventional approach in which the parametric effects of variables from our formal Bayesian model were used to predict the amplitude of fMRI responses, after convolution with a suitable hemodynamic response function [53]. Because the majority of experimental variation in the model predictions is between conditions (AC vs. GC), we arranged these conditions in a block design to ensure high efficiency. We could therefore choose either a conventional analysis that simply tested for condition effects or a model-based analysis that used parametric variations within and between conditions. To exploit our formal model, we used the more comprehensive model-based analysis: The fMRI data were modeled using a general linear convolution model, whose explanatory variables comprised stimulus functions convolved with a canonical hemodynamic response function. These stimulus functions comprised delta functions modulated by the following:

- 1) The prediction ‘risk’ under M1 (time-locked to the choice presentation):
- 2) The reward predictions under M1, conditioned on the subject’s choice (time-locked to the choice):
- 3) The model-based (Shannon) surprise at the outcome under M1 (time-locked to the outcome):
- 4) The signed model-based prediction error under M1, conditional on subject’s choice (time-locked to the outcome):
- 5) the trial outcome: correct/incorrect, coded at 1,0 respectively (time-locked to the outcome).

In terms of our hypothesis, these regressors can be regarded as modeling nuisance effects. Our final regressor was the key effect of interest; namely, the trial specific ambiguity as measured by the Shannon entropy above. It is this measure that reflects an encoding of contextual uncertainty that weakens generalization. The entropy entered as parametrically modulated delta functions at the time of choice, but before feedback. Six columns describing scan-specific rigid body translations and rotations were included as confounds. The data was temporally filtered to remove low-frequency drifts below 1/128 Hz.

Results

Figure 2 shows the value (expected reward) of each choice according to the two main learning models we considered, together with a typical subject’s guesses. Model 1 (M1) generalizes, while Model 2 (M2) cannot. For each subject, we used logistic regression to explain their choices in terms of these predictions and a constant term. Using a between-subject summary-statistic approach, we applied a two-tailed Student’s *t*-test to the subject-specific logistic regression coefficients associated with the predictions of M1 (red-dashed curve, Figure 2). We rejected the null hypothesis that this effect was equal to zero ($p=0.00041$, $n=19$). Interestingly, the size of the M1 regression coefficient predicted the total number of rewards obtained by each subject (correlation $\rho=0.651$, $p=0.0025$, $n=19$). This illustrates that generalization is evident behaviorally and pays off.

A secondary behavioral analysis assessed the specificity of M1 predictions by examining the explanatory power of M1 in the context of the alternative models, M2 to M5. For each subject, we used logistic regression to explain subject’s choices as a mixture of predictions from five models (M1 to M5), plus a constant term. Having estimated the logistic regression model for each subject, we again considered the subject-specific estimates for the coefficient reporting on M1 predictions. A two-tailed Student’s *t*-test on the M1 coefficients was highly significant, $p=0.00069$, $N=19$. No other model coefficients reached significance.

To summarize, we used standard regression techniques to ask if, having accounted for competing models, a component of choice behavior reflects Bayes-optimal generalization (M1). Specifically, we included several model predictions in one linear model and estimated the partial regression coefficient for the predictor of interest (action-values derived from M1). One can therefore [54] conclude that, over and above competing models, behavior can be predicted by M1. Because Models M2 and M3 have a free parameter this conclusion is conservative: having been pre-fit to

subject's behavior, these models have an explanatory advantage that is unavailable to M1 (or M4 and M5). In contrast to M2 (Eq. 8), M1 attempts to explain behavior via abstract computational principles, not detailed mechanisms. Its predictions have no free parameters. Rather, its predictions are based only on the subject's observations under ideal Bayesian assumptions. We have demonstrated that this model predicts behavior, above and beyond that explained by the other models considered. In what follows, we now ask whether the brain encodes ambiguity [see e.g. [39] for a similar approach].

While M1 differs from other models in many ways, the important aspect for the fMRI analysis is that M1 provides an ambiguity measure. We therefore tested the null hypothesis that the fMRI signal is sensitive to ambiguity, as quantified by the Shannon entropy of prior belief (see above). In our fMRI data, fourteen subjects satisfied the inclusion criteria for a second-level between subject analysis (no interruptions to the scanner session or rapid head movement, as estimated by co-registration). We conducted regional and whole-brain analyses. All fMRI results presented here are based on the same general linear model, including the confounding factors (i.e., with nine regressors). In view of our specific hypothesis, region of interest (ROI) analyses asked whether activity within bilateral hippocampi tracked ambiguity about the current contingencies. Figure 4a shows the anatomy of the ROI.

Figure 3 depicts ambiguity about a new individual (alternating block-wise between GC and AC blocks). As discussed, this dictates the relative influence of the current observation on belief updates (higher when there is high ambiguity). The parameter estimates associated with the entropy regressor above were averaged over bilateral hippocampal voxels for each subject, using the AAL atlas [55]. We applied a two-tailed Student's *t*-test to these subject-specific summaries, testing the null hypothesis that hippocampal responses do not covary with ambiguity. We were able to reject this null hypothesis with a correct $p=0.0245$, $n=14$. Repeating the analysis on unilateral right and left hippocampus separately provided similar results ($p=0.0303$, $p=0.0328$, respectively). (These latter two results examine the separate contribution of each hemisphere to our bi-lateral effect. These tests are not statistically independent of the bi-lateral test and were not subject to additional correction.) There was no significant difference between left and right hippocampi. Our results therefore suggest that neuronal activity encodes the same sorts of variables that arise in our Bayes-optimal computations and, consequently, may be performing some form of approximate Bayesian inference.

As with the behavioral data, we next examined the between-subject correlation between the hippocampal ambiguity coefficients and the total number of rewards attained in the experiment (correlation $\rho=0.554$, $p=0.0399$, $n=14$). Testing for separate correlations in left and right hippocampal effects gave respectively: $\rho=0.624$, $p=0.0172$ and $\rho=0.4265$, $p=0.1283$ ($n=14$).

In an exploratory whole brain analysis, we then smoothed the data with a Gaussian Kernel $FWHM=[4mm\ 4mm\ 4mm]$ and re-estimated the general linear model above using a conventional SPM analysis with whole brain correction for multiple comparisons [56]. Two right-hemisphere clusters survived correction for cluster-extent (using a height threshold of 3). The first region ($p=0.04$ FWE corrected) subsumed a right hippocampal region, mostly hippocampus and amygdala, but also putamen, as defined with the AAL atlas [55]. The second region ($p=0.031$ FWE corrected) encompassed the fusiform gyrus and precuneus, with a spill-over into a calcarine region. These regions are shown in maximum intensity projection format in Figure 4b (this display format shows voxels with maximum intensity that fall on parallel

lines traced from the viewpoint to the plane of projection as in a standard X-Ray). Orthogonal views of the anterior activation at its local maximum are shown in Figure 4c. For illustration purposes, Figure 4d shows the mean times series in this anterior region, averaged over all subjects. All of the above fMRI analyses were based on the same model, which included the nuisance regressors listed in **Relating model predictions to data: fMRI**. None of these nuisance effects could explain the variation in hippocampal responses that was explained by our Bayes-optimal generalization model (M1).

Discussion

Behaviorally, we have shown that subjects learn action-reward relationships in a manner that enables them to generalize rules to new situations. Crucially, this enables subjects to adapt their learning rate to provide an optimal balance between pre-existing generalizations and new information. We established this by showing that the accuracy of subjects' guesses evolved over trials in a way that was predicted by Bayes-optimal generalization, using a statistical model equipped with prior beliefs that allowed for contextual ambiguity. Furthermore, we established that a significant component of hippocampal responses could be explained by fluctuations in ambiguity under this model. These regionally specific responses were also significant in a whole brain SPM analysis.

We provide empirical support for a model that explains how experience moderates decision making. In this model, the bias towards rule-based choices is determined by low ambiguity. We show that both learning and hippocampal responses are attenuated when the underlying rule is learned and applied in an unambiguous context. Conventional 'model-free' reinforcement learning cannot easily explain such effects because these schemes do not include contextual ambiguity. As noted in the introduction, one recent variant of reinforcement learning [43] is relevant here: In this two-system learning theory, generalization between observable cues rests both on their perceptual similarity and their predictive similarity (do cues look the same? do they predict the same outcomes?). The authors of [43] contrast normal learning with under/over-generalization or 'under/over willingness to generate a new state' p 97. We have used a single model that formalizes this optimality by drawing on principles of optimal probabilistic generalization (see [42] for a related model). As in [43], our model generalizes by classifying observable cues before acting. Unlike [43], it invokes an explicit representation of subjective ambiguity to mediate and optimize this generalization. There remains an interesting challenge to relate our formulation and results to classical RL schemes. Interestingly the authors of [43] speculate that the neuronal systems mediating generalization depend on the hippocampus (and PFC); because these systems are flexible, the rules by which observable cues are classified can easily be changed to permit new discriminations. These speculations are entirely consistent with our findings.

As in previous treatments [14], we distinguish uncertainty about objective, observable events (e.g., the risk of getting 'tails' in a fair coin flip) from subjective ambiguity about unobservable states or parameters (is the coin really fair?). While the hippocampus has been implicated in the former [23,35,57], the latter is central to computational accounts of contextual learning and inference; e.g. [1,22]. Using a Bayes-optimal model, our work provides the first evidence that the hippocampus tracks contextual ambiguity about hidden or latent variables.

Previous work [11,14,39,58] has addressed how ambiguity mediates the influence of uncued *temporal* variability (volatility) on

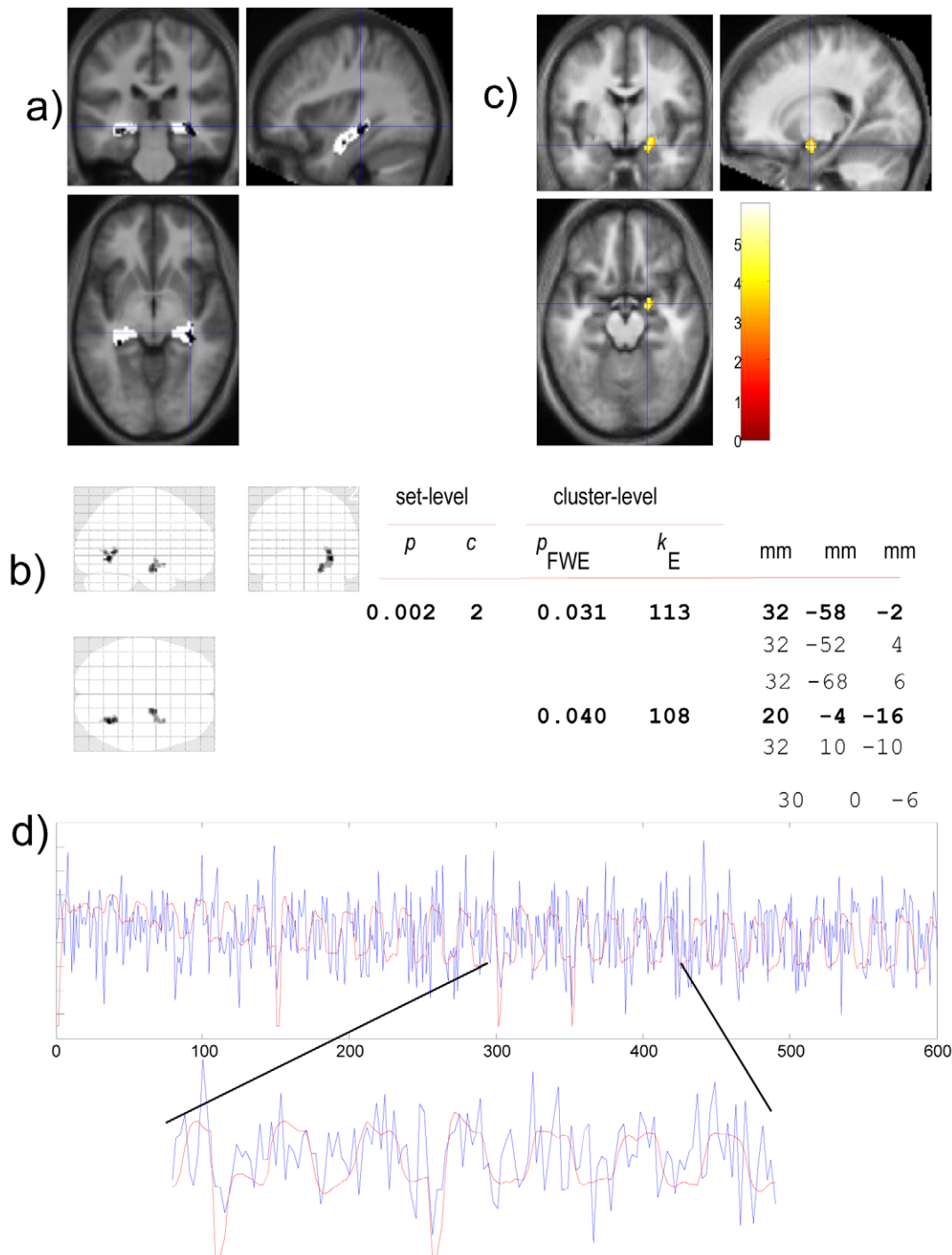


Figure 4. Structural MRI, mask and functional activation. (a) Structural region-of-interest (white) on the subjects' average anatomical image. For visualization, black blobs exceed the 0.05 level uncorrected threshold corresponding to a Student's t (13 df). See the main text for statistical inference at a corrected $p < 0.05$ level. (b,c) Whole brain analysis of the effect of ambiguity. (b) Shows a glass-brain view (maximum intensity projection) of significant activations. (c) Shows the anterior activation, which included right anterior hippocampus and amygdala as defined, superimposed on the subjects' average anatomical image. (d) The observed fMRI trial-by-trial time-series (blue) averaged over all subjects for the hippocampal activation identified in our whole brain analysis (see main text). The model-based ambiguity is shown in red. Note that the model only captures the slow changes in observed responses over blocks as the contingencies are learned. doi:10.1371/journal.pcbi.1002346.g004

learning. We asked if variability in response requirements to different cues influences creates ambiguity and influences learning. In the current study, we manipulated the uncertainty about the behavioral contingencies over contexts, rather than time, and

showed that associative learning adapts accordingly. Further work could examine whether neuromodulatory manipulations influence this effect; e.g., by selectively facilitating synaptic gain as predicted by [11,59]. The role of dopamine deserves special attention, given

prior work with Pavlovian or simpler instrumental tasks [60]. Additionally, given that the amygdala is able to modulate memory storage in non-amygdala brain areas [61], multi-region *in vivo* recordings could disclose interactions with the hippocampus in these tasks. Interestingly, the amygdala activation in our whole-brain analyses is consistent with previous work implicating the amygdala in the representation of ambiguity [24,62]. However, previous studies were unable to address whether ambiguity regulates learning, as predicted theoretically. In line with Bayesian learning theory, our results suggest that learning (updating beliefs) can be guided by optimal probabilistic constraints, generalized from previous experience.

The learning rate in (model-free) reinforcement learning prescribes the sensitivity of belief updates to current information. When this information is under or over-weighted, inefficient learning ensues. While classical RL is non-probabilistic (i.e. has a degraded uncertainty representation [22]), it may in principle address this challenge by incorporating something akin to an ‘ambiguity-dependent’ or ‘surprise-dependent’ learning rate. For example, attempts have been made to optimise learning rates [63,64] in both stationary and non-stationary settings [65]. Bayesian learners use the rules of probability to achieve this balance by weighing new information against pre-existing generalizations. The relative weight of the latter depends upon ambiguity (the relative confidence in prior beliefs about the current context). When pre-existing beliefs are held with a high

degree of confidence, they generally accommodate new observations, by down-weighting their impact. Such abilities to balance different sources of information and constraints are at the heart of adaptive behavior [66]. For example, appropriate social behavior requires communal norms, while retaining sensitivity to individual inclinations and preferences. The (social) learning task in this paper is a first step in this direction. Conversely, aberrant generalization has widespread consequences [2,3,5,67]. The framework used in this study may provide an experimental framework to quantify dysfunctional generalization in specific patients; e.g., over-generalized schemata which persist despite contradictory evidence, as seen in depressive and delusional states and its associated pathophysiology at the neuronal level.

Acknowledgments

We would like to thank Peter Dayan, Jon Rosier, Read Montague, Antonio Rangel, John O’Doherty, Grit Hein and our anonymous reviewers for invaluable discussions and feedback.

Author Contributions

Conceived and designed the experiments: JRC KJF. Performed the experiments: JRC DRB. Analyzed the data: JRC. Contributed reagents/materials/analysis tools: JRC JD EF GF. Wrote the paper: JRC RJD KJF. Mathematical/computational modelling: JRC.

References

- Bishop C (2006) Pattern Recognition and Machine Learning. Springer.
- Beck A (1993) Cognitive Therapy and the Emotional Disorders. NY: Penguin.
- Dayan P, Huys QJM (2008) Serotonin, inhibition, and negative mood. *PLoS Comp Biol* 4: e4.
- Huys Q, Vogelstein J, Dayan P (2009) Psychiatry: insights into depression through normative decision-making models. In: Koller D, Schuurmans D, Bengio Y, Bottou L, eds. *Advances in Neural Information Processing Systems 21*. Cambridge (Massachusetts): MIT Press. pp 729–736.
- Carrell PL, Eisterhold JC (1983) Schema theory and ESL reading pedagogy. *TESOL Quarterly* 17: 553–573.
- Hestenes D (1987) Toward a modeling theory of physics instruction. *Am J Phys* 55: 440–454.
- Tenebaum JB, Griffiths TL (2001) Generalization, similarity, and Bayesian inference. *Behav Brain Sci* 24: 629–+.
- Griffiths TL, Kemp C, Tenenbaum JB (2008) Bayesian models of cognition. In: Ron Sun, ed. *The Cambridge handbook of computational cognitive modeling*. Cambridge University Press.
- O’Reilly RC, Norman KA (2002) Hippocampal and neocortical contributions to memory: advances in the complementary learning systems framework. *Trends Cogn Sci* 6: 505–510.
- Wittmann BC, Bunzeck N, Dolan RJ, Duzel E (2007) Anticipation of novelty recruits reward system and hippocampus while promoting recollection. *Neuro Image* 38: 194–202.
- Yu AJ, Dayan P (2005) Uncertainty, neuromodulation, and attention. *Neuron* 46: 681–692.
- Chater N, Tenenbaum J, Yuille A (2006) Probabilistic models of cognition: Conceptual foundations. *Trends Cogn Sci* 10: 287–291.
- Schultz W, Preusschoff K, Camerer C, Hsu M, Fiorillo C, et al. (2008) Explicit neural signals reflecting reward uncertainty. *Philos Trans R Soc Lond, Ser B: Biol Sci* 363: 3801.
- Payzan-LeNestour E, Bossaerts P (2011) Risk, Unexpected Uncertainty, and Estimation Uncertainty: Bayesian Learning in Unstable Settings. *PLoS Comp Biol* 7: e1001048.
- Rescorla R, Wagner A (1972) Variations in the Effectiveness of Reinforcement and Nonreinforcement. In: *Classical Conditioning II: Current Research and Theory*. New York: Appleton-Century-Crofts.
- Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. *Science* 275: 1593–1599.
- Zilli EA, Hasselmo ME (2008) Modeling the role of working memory and episodic memory in behavioral tasks. *Hippocampus* 18: 193–209.
- Sutton R, Barto A (1998) Reinforcement learning: An introduction. Cambridge: The MIT Press.
- Gustafson NJ, Daw ND (2011) Grid Cells, Place Cells, and Geodesic Generalization for Spatial Reinforcement Learning. *PLoS Comp Biol* 7: e1002235.
- Ludvig EA, Sutton RS, Kehoe EJ (2008) Stimulus representation and the timing of reward-prediction errors in models of the dopamine system. *Neural Comput* 20: 3034–3054.
- Preusschoff K, Bossaerts P (2007) Adding prediction risk to the theory of reward learning. *Ann N Y Acad Sci* 1104: 135–146.
- Cox RT (1946) Probability, Frequency and reasonable expectation. *Am J Phys* 14: 1–13.
- Strange BA, Duggins A, Penny W, Dolan RJ, Friston KJ (2005) Information theory, novelty and hippocampal responses: unpredicted or unpredictable? *Neural Netw* 18: 225–230.
- Hsu M, Lin H, McNamara P (2008) Neuroeconomics of decision-making in the aging brain: the example of long-term care. *Adv Health Econ Health Serv Res* 20: 203–225.
- Hsu M, Bhatt M, Adolphs R, Tranel D, Camerer CF (2005) Neural systems responding to degrees of uncertainty in human decision-making. *Science* 310: 1680.
- Bach DR, Hulme O, Penny WD, Dolan RJ (2011) The Known Unknowns: Neural Representation of Second-Order Uncertainty, and Ambiguity. *J Neurosci* 31: 4811.
- Kemp C, Perfors A, Tenenbaum JB (2007) Learning overhypotheses with hierarchical Bayesian models. *Dev Sci* 10: 307–321.
- Congdon P (2003) *Applied bayesian modelling*. John Wiley & Sons Inc.
- Greene A, Gross W, Elsingher C, Rao S (2006) An fMRI analysis of the human hippocampus: inference, context, and task awareness. *J Cogn Neurosci* 18: 1156–1173.
- Shohamy D, Wagner A (2008) Integrating memories in the human brain: hippocampal-midbrain encoding of overlapping events. *Neuron* 60: 378–389.
- Zeithamova D, Preston A (2010) Flexible Memories: Differential Roles for Medial Temporal Lobe and Prefrontal Cortex in Cross-Episode Binding. *J Neurosci* 30: 14676.
- Heckers S, Zalesak M, Weiss A, Ditman T, Titone D (2004) Hippocampal activation during transitive inference in humans. *Hippocampus* 14: 153–162.
- Preston A, Shrager Y, Dudukovic N, Gabrieli J (2004) Hippocampal contribution to the novel use of relational information in declarative memory. *Hippocampus* 14: 148–152.
- Eichenbaum H, Cohen N (2004) *From conditioning to conscious recollection: Memory systems of the brain*. Oxford University Press, USA.
- Harrison LM, Duggins A, Friston KJ (2006) Encoding uncertainty in the hippocampus. *Neural Netw* 19: 535–546.
- Weiskopf N, Hutton C, Josephs O, Deichmann R (2006) Optimal EPI parameters for reduction of susceptibility-induced BOLD sensitivity losses: A whole-brain analysis at 3 T and 1.5 T. *NeuroImage* 33: 493–504.
- Weiskopf N, Hutton C, Josephs O, Turner R, Deichmann R (2007) Optimized EPI for fMRI studies of the orbitofrontal cortex: compensation of susceptibility-induced gradients in the readout direction. *MAGMA* 20: 39–49.

38. Deichmann R, Schwarzbauer C, Turner R (2004) Optimisation of the 3D MDEF T sequence for anatomical brain imaging: Technical implications at 1.5 and 3 T. *NeuroImage* 21: 757–767.
39. Behrens TEJ, Woolrich MW, Walton ME, Rushworth MFS (2007) Learning the value of information in an uncertain world. *Nat Neurosci* 10: 1214–1221.
40. Teh YW, Jordan MI, Beal MJ, Blei DM (2006) Hierarchical dirichlet processes. *J Am Stat Assoc* 101: 1566–1581.
41. Griffiths TL, Canini KR, Sanborn AN, Navarro DJ (2007) Unifying rational models of categorization via the hierarchical Dirichlet process. Proceedings of the Twenty-Ninth Annual Conference of the Cognitive Science Society.
42. Gershman SJ, Blei DM, Niv Y (2010) Context, learning, and extinction. *Psychol Rev* 117: 197.
43. Redish AD, Jensen S, Johnson A, Kurth-Nelson Z (2007) Reconciling reinforcement learning models with behavioral extinction and renewal: Implications for addiction, relapse, and problem gambling. *Psychol Rev* 114: 784–805.
44. Sethuraman J (1994) A constructive definition of Dirichlet priors. *Statistica Sinica* 4: 639–650.
45. Teh Y (2010) Dirichlet Processes. In: Sammut CW, Geoffrey I, eds. *Encyclopedia of Machine Learning* Springer.
46. Blackwell D, MacQueen JB (1973) Ferguson distributions via Polya urn schemes. *Ann Stat* 1: 353–355.
47. Neal RM (2000) Markov chain sampling methods for Dirichlet process mixture models. *J Comput Graph Stat* 9: 249–265.
48. Escobar MD, West M (1995) Bayesian density estimation and inference using mixtures. *J Am Stat Assoc* 90: 577–588.
49. Gelman A (2004) *Bayesian data analysis*. Florida. CRC press.
50. Daw ND, O’Doherty JP, Dayan P, Seymour B, Dolan RJ (2006) Cortical substrates for exploratory decisions in humans. *Nature* 441: 876–879.
51. Kakade S, Dayan P (2002) Dopamine: generalization and bonuses. *Neural Netw* 15: 549–559.
52. Dayan P, Niv Y, Seymour B, Daw ND (2006) The misbehavior of value and the discipline of the will. *Neural Netw* 19: 1153–1160.
53. O’Doherty JP, Hampton A, Kim H (2007) Model-based fMRI and its application to reward learning and decision making. *Ann N Y Acad Sci* 1104: 35.
54. Dobson AJ (2002) An introduction to generalized linear models. CRC Pr I Llc.
55. Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, et al. (2002) Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage* 15: 273–289.
56. Friston KJ, Holmes A, Poline JB, Price CJ, Frith CD (1996) Detecting activations in PET and fMRI: Levels of inference and power. *NeuroImage* 4: 223–235.
57. Vanni-Mercier G, Mauguire F, Isnard J, Dreher JC (2009) The Hippocampus Codes the Uncertainty of Cue-Outcome Associations: An Intracranial Electrophysiological Study in Humans. *J Neurosci* 29: 5287–5294.
58. den Ouden HEM, Daunizeau J, Roiser J, Friston KJ, Stephan KE (2010) Striatal prediction error modulates cortical coupling. *J Neurosci* 30: 3210.
59. Hasselmo ME (1995) Neuromodulation and cortical function - modeling the physiological basis of behaviour. *Behav Brain Res* 67: 1–27.
60. Dayan P, Niv Y (2008) Reinforcement learning: The Good, The Bad and The Ugly. *Curr Opin Neurobiol* 18: 185–196.
61. Cahill L, Weinberger NM, Roozendaal B, McGaugh JL (1999) Is the amygdala a locus of “conditioned fear”? Some questions and caveats. *Neuron* 23: 227–228.
62. Prévost C, McCabe JA, Jessup RK, Bossaerts P, O’Doherty JP (2011) Differentiable contributions of human amygdalar subregions in the computations underlying reward and avoidance learning. *Eur J Neurosci* 34: 134–145.
63. Mackintosh NJ (1983) *Conditioning and associative learning*. Oxford: Clarendon Press.
64. Pearce JM, Hall G (1980) A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychol Rev* 87: 532.
65. Sutton RS (1992) Gain adaptation beats least squares. *Citeseer*. pp 161–166.
66. Foraker S, Regier T, Khetarpal N, Perfors A, Tenenbaum J (2009) Indirect Evidence and the Poverty of the Stimulus: The Case of Anaphoric One. *Cogn Sci* 33: 287–300.
67. Huys QJM, Dayan P (2009) A Bayesian formulation of behavioral control. *Cognition* 113: 314–328.