Reference-dependent preferences arise from structure learning

Lindsay E. Hunter[1] and Samuel J. Gershman[2]
[1]Department of Psychology and Princeton Neuroscience Institute, Princeton University
[2]Department of Psychology and Center for Brain Science, Harvard University

January 23, 2018

Abstract

Modern theories of decision making emphasize the reference-dependency of decision making under risk. In particular, people tend to be risk-averse for outcomes greater than their reference point, and risk-seeking for outcomes less than their reference point. A key question is where reference points come from. A common assumption is that reference points correspond to expectations about outcomes, but it is unclear whether people rely on a single global expectation, or multiple local expectations. If the latter, how do people determine which expectation to apply in a particular situation? We argue that people discover reference points using a form of Bayesian structure learning, which partitions outcomes into distinct contexts, each with its own reference point corresponding to the expected outcome in that context. Consistent with this theory, we show experimentally that dramatic change in the distribution of outcomes can induce the discovery of a new reference point, with systematic effects on risk preferences. By contrast, when changes are gradual, a single reference point is continuously updated.

Introduction

When people make decisions under risk (e.g., accepting or rejecting a gamble), their proclivity for risk depends critically on whether the outcome of the gamble is perceived as a gain or a loss [1]. If the outcome is perceived as a gain, most people will require an additional incentive (the risk premium) relative to a risk-neutral decision maker in order to accept the gamble, indicating that people are risk-averse for perceived gains. In contrast, most people are risk-seeking for perceived losses. The notion of "perception" is important here, because an objective gain may be perceived as a loss if it is less than expected (e.g., when one receives a surprisingly small raise), and likewise an objective loss may be perceived as a gain if it is greater than expected (e.g., a surprisingly inexpensive ticket). The expectation thus acts as a reference point for subjective valuation.

Modern theories of decision making have sought to formalize the concept of an expectation-based reference point and how it changes based on experience. In an influential line of work, Kőszegi and Rabin [2, 3] proposed that reference points reflect rational expectations based on recent outcomes (see also [4, 5]). In support of this theory, contestants on the TV game show "Deal or No Deal" were more likely to make risky choices when they had recently experienced unfavorable outcomes [6], recapitulating results from laboratory experiments [7, 8]. Similarly, experiments with foraging animals have demonstrated risk-seeking behavior when reward rate is low [9].

A standard assumption is that a single reference point is updated gradually over time as new outcomes are observed. However, this cannot be the whole story, for several reasons. First, the decision making literature is scattered with observations that people can adopt multiple reference points [10, 11, 12, 8]. Kahneman [13] likened the mental co-existence of reference points to ambiguous images (e.g., the Necker cube); the mind does not settle on one or average them together, but instead entertains them all in a state of tension. Second, evidence from other domains suggests that humans organize their knowledge into discrete units (chunks, clusters, contexts, etc.) based on statistical regularities [14, 15, 16, 17, 18, 19], and thus it seems plausible that a similar form of "structure learning" might be invoked to organize the distribution of outcomes into discrete contexts.

In this paper, we pursue this idea theoretically and empirically. Following prior work [2, 3], we posit that reference points reflect rational expectations updated based on recent outcomes. However, we additionally assume that reference points can be discovered *de novo* by structure learning. Adapting a paradigm for studying structure learning in perceptual judgment [14], we present experimental evidence that risk preferences for gambles with outcomes drawn from a fixed distribution are influenced by the distribution of other gambles experienced in the same context (varied across blocks). Crucially, if the fixed and variable distributions are sufficiently different, then the contextual effects are attenuated, indicating that they were assigned to distinct reference points. This attenuation effect is eliminated when the variable distribution is changed gradually across blocks, suggesting that a single reference point is applied to both distributions when their differences are made less salient. These patterns are captured by a Bayesian structure learning model of reference point formation.

Results

Experiment 1

Participants completed a binary choice task (see Materials and Methods for details) in which they made choices between a certain lottery (e.g., 50 points) and a lottery offering a fair chance of doubling or forfeiting the same amount (e.g., 50% chance of 100 points, 50% chance of 0 points). Thus, the expected values of the two options were equivalent.

Unbeknownst to participants, the expected value was randomly sampled from one of two Gaussian distributions (denoted A and B; Figure 1). Distribution A was held fixed across all conditions, whereas the mean of distribution B was varied. According to our structure learning account, which we formalize below, participants should cluster A and B trials together when the means of the distributions [denoted EV(A) and EV(B)] are close, because this is a statistically parsimonious account of the data. In this case (Figure 2A), a single reference point is applied to both A and B trials, corresponding to the expectation of the merged distribution. The reference point for A trials should increase with the mean of B, as long as the two trial types are clustered together (Figure 2B). We can discern evidence for this shifting reference point by measuring the probability of gambling on A trials as a function of EV(B). Risk-averse participants should be more likely to gamble as EV(B) increases (under the assumption that utilities are concave for gains and convex for losses), whereas risk-seeking participants should be less likely to gamble.

Crucially, the structure learning account also predicts that participants should separate A and B trials into distinct clusters when EV(A) and EV(B) are sufficiently distinct (Figure 2C). Since

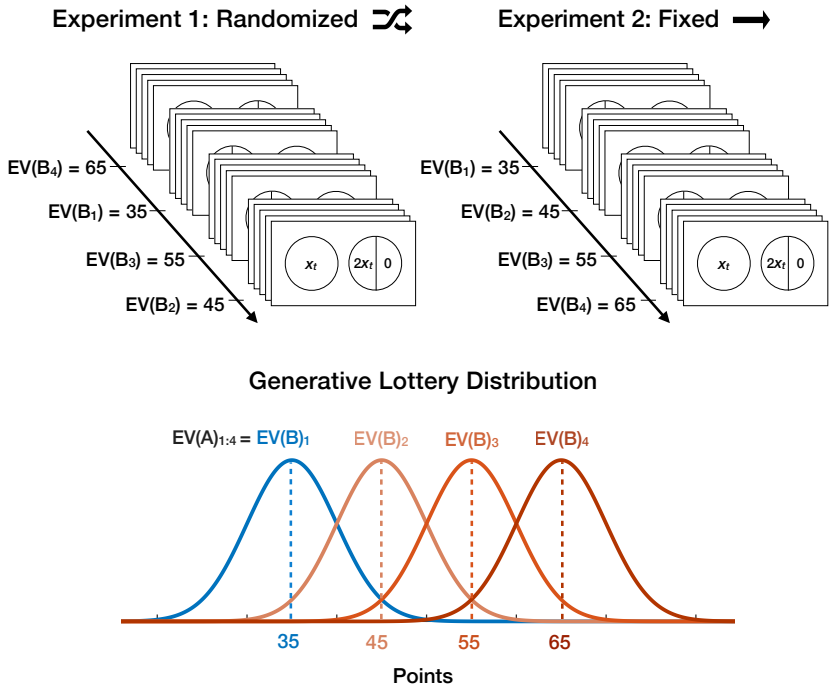

Figure 1: Task Design. On each trial, participants made a choice between a certain reward (displayed as a circle containing the reward amount in points) and a lottery with the same expected value (displayed as a divided circle containing twice the certain reward amount on one half and 0 on the other half). Expected values were drawn from one of two Gaussian distributions (A or B). Distribution A was fixed across all 4 conditions, whereas the mean of distribution B varied across conditions. In Experiment 1, the order of conditions was randomized across blocks; in Experiment 2, the conditions were ordered such that the mean of B increased monotonically across blocks.

EV(A) is always less than or equal to EV(B), this means that the reference point should decrease, decreasing gambling propensity for risk-averse participants and increasing it for risk-seeking participants. Thus, the structure learning account predicts that P(Gamble|A) will be a non-monotonic function of EV(B).

We tested these predictions in a data set of 92 participants (54 designated as risk-averse, 38 designated as risk-seeking; see Materials and Methods for a description of how this designation was determined). The mean of B differed in increments of 10 across experimental conditions: 35, 45, 55, and 65 points, with the lowest value equal to the mean of distribution A. The order of these conditions was randomized across blocks. We reasoned that randomization would make the difference between conditions highly salient.

As shown in Figure 3, both risk-averse and risk-seeking participants changed their risk preference for A trials non-monotonically as a function of EV(B). In accordance with our predictions, risk-averse participants first increased their risk preference and then decreased it, whereas risk-seeking participants did the opposite. Note that, if the expectations-based reference point were based solely on an estimate of overall average reward, we would expect P(Gamble|A) to increase monotonically. Alternatively, if the reference point were based solely on the average reward for A trials, then we

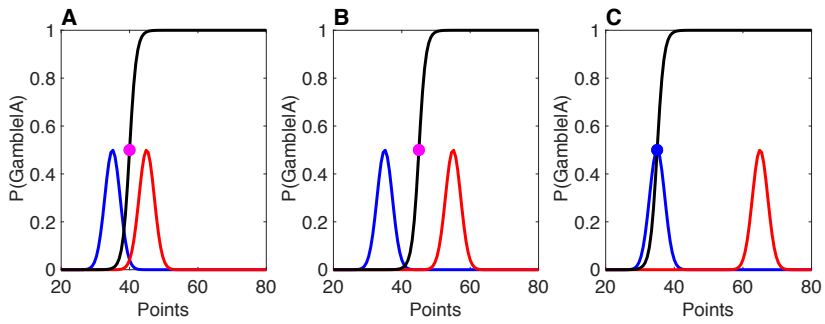

Figure 2: Reference point formation as structure learning. The black curve shows the probability of gambling on A trials, with risk-aversion above the reference point (indicated by a circle), and risk-seeking below the reference point; for risk-seeking participants, this pattern flips. The distribution of expected values for A trials is shown in blue, and the distribution for B trials is shown in red. (A) When the two distributions are close together, a shared reference point (purple circle) is formed based on the merged distribution. (B) For a moderate increase in the mean of the B distribution, the shared reference point increases. (C) For a sufficiently large increase in the mean of B, separate reference points are formed for A and B trials.

would expect no change across conditions. Thus, our experimental results provide strong support for our structure learning hypothesis.

To evaluate these patterns quantitatively, we fit a mixed-effects logistic regression model with the form $\text{P(Gamble|A)} \sim \text{Int.} + \text{EV(B)} + \text{EV(B)}^2$, which allows us to capture quadratic effects of EV(B). We compared this to a model that lacked the quadratic term, and hence can only capture linear effects of EV(B). In both models, the identity of the participant was treated as a random effect, and parameters were estimated separately for risk-averse and risk-seeking groups.

The regression results are summarized in Table 1 (see Table S1 in the Supporting Information for the linear model results). For risk-averse participants, the quadratic model fit yielded significant positive linear and negative quadratic coefficients. For risk-seeking participants, the quadratic model fit yielded significant negative linear and positive quadratic coefficients. The quadratic effects constitute quantitative support for the non-monotonic pattern shown in Figure 3. For both risk-averse and risk-seeking participants, likelihood ratios tests allowed us to reject the linear model relative to the quadratic model (risk-averse: $p < 4.8e-12$, risk-seeking: $p < 2.1e-07$). Moreover, two standard model comparison metrics (Akaike information criterion and Bayesian information criterion) favored the quadratic model (Table 2).

Experiment 2

The results of Experiment 1 indicate that large, abrupt changes in distributional statistics can drive reference point formation. Based on findings from Pavlovian [20], motor [21], and perceptual [22, 23, 15] learning experiments, we hypothesized that subtler changes would obscure the differences between distributions and thus prevent new reference points from being formed. To test this hypothesis, we used the same conditions as in Experiment 1, but ordered them monotonically across blocks, such that each transition between blocks was associated with a relatively gradual 10-point change in the mean of distribution B.

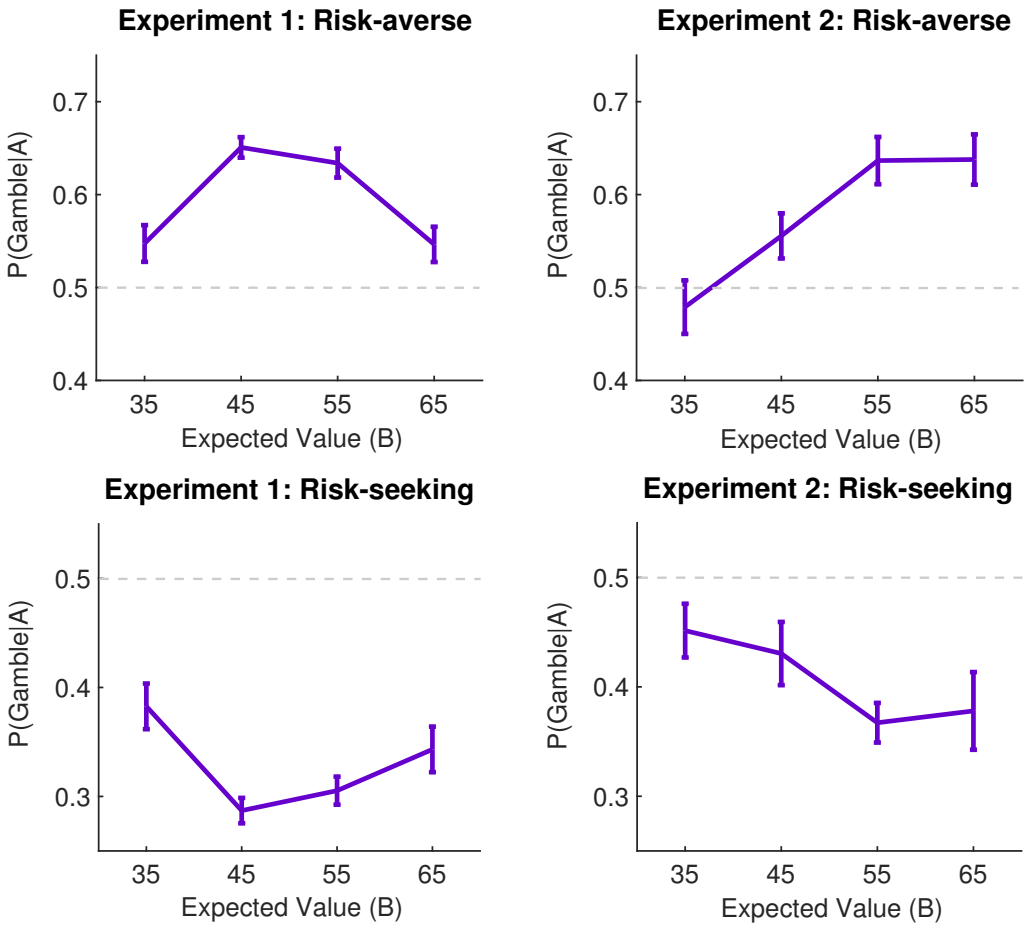

Figure 3: Probability of gambling (choosing the risky option) on lotteries drawn from distribution A, plotted as a function of the mean of distribution B. Results are shown separately for risk-averse (top) and risk-seeking (bottom) participants. Left: Experiment 1 results (mean of B randomized across blocks). Right: Experiment 2 results, where the mean of B increases monotonically across blocks. Error bars represent within-participant standard error of the mean.

Table 1: Parameter estimates for mixed-effects logistic regression model of P(Gamble|A), the probability of choosing the risky option when the expected value is drawn from distribution A. β = regression coefficient; SE = standard error; Int. = intercept; EV(B) = expected value of distribution B.

		β	SE	p-value
Experiment 1				
	Risk-Averse			
	Int.	-0.67	(0.24)	5.6e-03 ***
	EV(B)	1.1	(0.20)	1.4e-08***
	$EV(B)^2$	-0.22	(3.6e-02)	1.1e-09***
	Risk-Seeking			
	Int.	0.44	(0.22)	5.6e-02 *
	EV(B)	-1.33	(0.29)	1.4e-06***
	$EV(B)^2$	0.25	(5.5e-02)	1.1e-06***
Experiment 2				
	Risk-Averse			
	Int.	-0.69	(0.46)	0.13
	EV(B)	0.67	(0.34)	4.8e-02*
	$EV(B)^2$	-0.08	(6.7e-02)	0.23
	Risk-Seeking			
	Int.	0.19	(0.32)	0.54
	EV(B)	-0.50	(0.44)	0.26
	$EV(B)^2$	0.066	(9.4e-02)	0.48

Data from 39 participants (18 risk-averse, 21 risk-seeking) were analyzed using the same procedure described for Experiment 1. Consistent with our hypothesis that gradual change would obscure the distributional differences between conditions and lead to a single shared reference point, risk-averse participants increased their risk preference on A trials monotonically as a function of EV(B), while risk-seeking participants decreased their risk preference monotonically (Figure 3).

Mixed-effects logistic regression confirmed this observation statistically. In contrast to the results of Experiment 1, we found no evidence for quadratic effects in either participant group (Table 1). Quantitative model comparison metrics favored the linear over the quadratic model (Table S2), and parameter estimates (summarized Table S1) showed significant linear effects of EV(B). Taken together, these results support our claim that a shared reference point tracked the gradually increasing EV(B) across blocks.

Computational modeling

To account for experimental results, we generalize the theory of expectation-based reference point updating to incorporate structure learning. Our point of departure is the idea that the reference

Table 2: Model comparison metrics for regression analyses. AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion.

		AIC		BIC	
		Quadratic	Linear	Quadratic	Linear
Experiment 1	*Risk-Averse*	6703.7	6754.7	6763.0	6787.6
	Risk-Seeking	4141.1	4169.7	4197.2	4200.9
Experiment 2	*Risk-Averse*	2147.6	2144.2	2197.2	2171.7
	Risk-Seeking	2438.7	2461.2	2489.7	2489.5

point corresponds to an agent's expectation [2, 3, 4], rationally updating over time using Bayes' rule. Because the agent expects that rewards may arise from multiple latent causes [20, 15], the expectation (and thus the reference point) can sometimes "jump" rather than adapt slowly. Most importantly for present purposes, structure learning can give rise to multiple reference points within the same context.

We develop the computational model in three stages. First, we describe the hypothetical data-generating process that characterizes the agent's internal model of the world. Second, we formalize the inference problem facing the agent: to form beliefs about structure (latent causes) and the distribution of rewards associated with each latent cause. Third, we describe how beliefs are translated into a choice policy. We then fit the model to our data and show that it can capture the key phenomena observed in our experiments.

Let $x_t \in \mathbb{R}$ denote the reward payoff for the certain option (or equivalently the expected value of the risky option) on trial t. This reward is drawn from a Gaussian distribution with mean μ_t^k and variance σ^2, where $k = z_t$ indicates the latent cause responsible for trial t. We assume a Gaussian prior over the initial condition μ_0^k, with mean 0 and standard deviation σ_0^2. To allow for gradual changes in the payoff distribution over time, we assume that the mean follows a Gaussian random walk: $\mu_t^k \sim \mathcal{N}(w\mu_{t-1}^k, q)$, where $w \in [0, 1]$ is a decay parameter that controls the rate of mean-reversion.

To allow for larger, abrupt changes in the payoff distribution, we assume that the latent cause can change over time, either by resampling an old latent cause or sampling a new latent cause. Following previous work on structure learning [20, 15], we model the prior over latent causes with a Chinese restaurant process (CRP) [24, 25], which generates assignments of trials to latent causes according to the following sequential stochastic process:

$$P(z_t = k|z_{1:t-1}) = \begin{cases} \frac{M^k}{t-1+\alpha} & \text{if } M^k > 0, \\ \frac{\alpha}{t-1+\alpha} & \text{if } M_k = 0. \end{cases}$$

where M^k is the number of trials assigned to latent cause k up to trial t, and $\alpha \geq 0$ is a parameter controlling the number of latent causes. When $\alpha = 0$, all trials are assigned to the same latent cause, and in the limit $\alpha \to \infty$, all trials are assigned to different latent causes. More generally, the expected number of latent causes after t trials is $\alpha \ln t$.

The computational problem facing the agent at time t is to infer the joint posterior over latent causes and their associated expected reward, as stipulated by Bayes' rule:

$$P(z_t, \mu_t|x_{1:t}) \propto P(x_t|z_t, \mu_t, x_{1:t-1})P(z_t)P(\mu_t),$$

where the index $1:t$ indicates the set of all trials from 1 to t. The likelihood $P(x_t|z_t, \mu_t, x_{1:t-1})$ and priors $P(z_t)P(\mu_t)$ are given by the generative process described above. Details about tractably approximating the posterior can be found in the Supporting Information.

We assume that preferences follow a reference-dependent quadratic utility function:

$$u(x; r) = x - r + \rho(x - r)^2,$$

where r denotes the reference point (see below) and ρ controls the curvature of the utility function. In our experimental task, the expected utility of option $c \in \{\text{certain}, \text{risky}\}$ is then given by:

$$\begin{aligned} V_t(\text{certain}) &= u(x_t; r_t), \\ V_t(\text{risky}) &= u(x_t; r_t) + \rho(x_t - r_t)^2. \end{aligned}$$

The curvature parameter ρ controls risk preferences: $\rho < 0$ implies concavity for payoffs above the reference point (risk aversion) and convexity for payoffs below the reference point (risk seeking), as in Prospect Theory [1]. This pattern reverses for $\rho > 0$, and $\rho = 0$ implies risk neutrality. The quadratic utility function can also be understood as a special case of a mean-variance choice model [26, 27].

Under an expectations-based reference point model, r corresponds to the expected payoff $\mathbb{E}[x]$. In our case, this expectation on trial t is given by:

$$\mathbb{E}[x_t] = w \sum_k P(z_t = k|x_{1:t-1})\hat{\mu}^k_{t-1},$$

where $\hat{\mu}^k_t = \mathbb{E}[x_t(c)|z_t = k]$ denotes the posterior mean payoff for latent cause k (see Supporting Information).

To allow for some stochasticity and bias in choice, we model the choice policy with a logistic sigmoid function $f(v) = 1/(1 + e^{-v})$:

$$\begin{aligned} P(c_t = \text{risky}) &= f(V_t(\text{risky}) - V_t(\text{certain}) + \psi) \\ &= f(\rho(x_t - r_t)^2 + \psi), \end{aligned}$$

where c_t is the choice on trial t, and ψ models an overall bias for risk seeking ($\psi > 0$) or risk aversion ($\psi < 0$).

The data from Experiments 1 and 2 were fit with two versions of the structure learning model (see Materials and Methods for model-fitting procedures): the full model ($\alpha > 0$) that can learn multiple reference points, and a restricted model ($\alpha = 0$) that learns a single reference point. Figure 4 shows the gambling probabilities for both models. The full model is able to capture the key findings: (1) a non-monotonic risk preference on A trials as a function of EV(B) in Experiment 1; (2) a monotonic risk preference in Experiment 2; and (3) opposite patterns of modulation for risk-averse and risk-seeking participants. The critical feature of the model is its ability to segregate A and B into separate latent causes when their expected values are sufficiently different. The importance of this feature is highlighted by the fact that the restricted model is unable to capture the non-monotonic risk preference in Experiment 1.

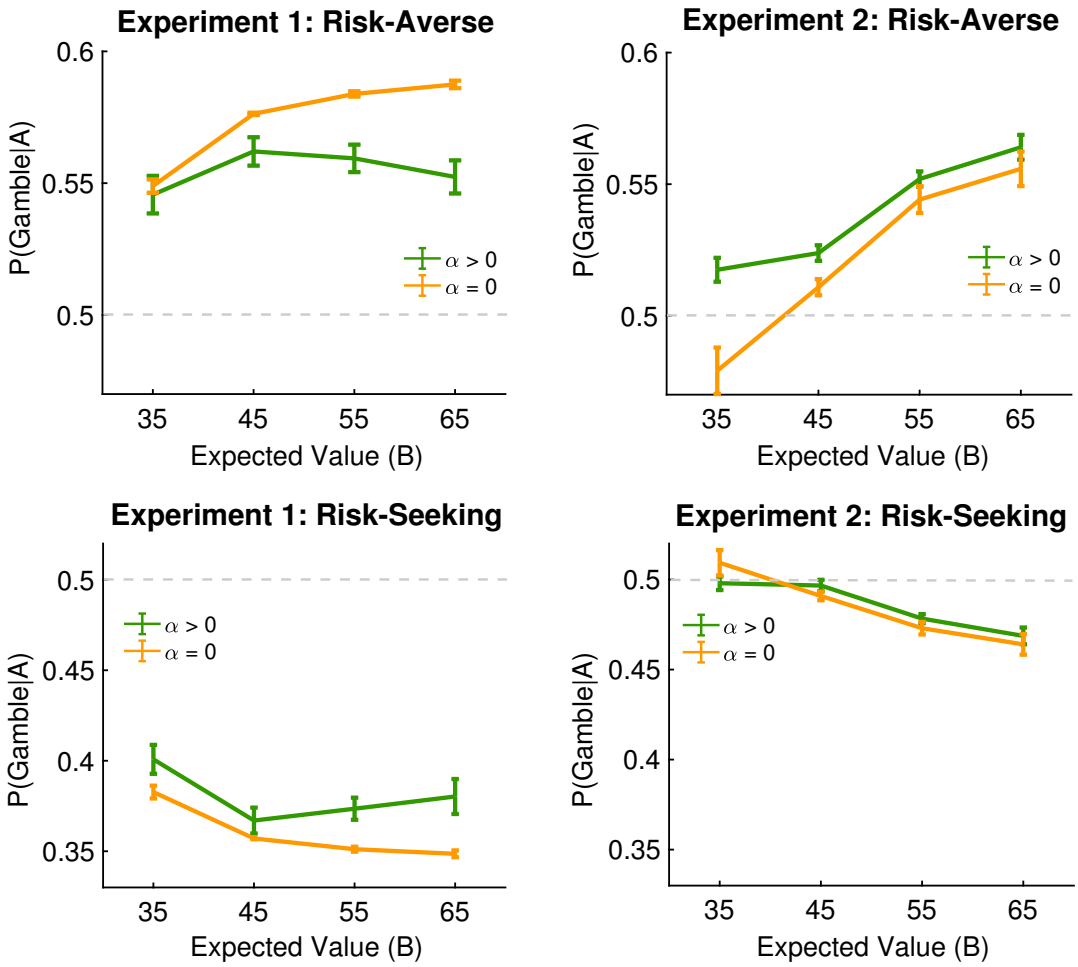

Figure 4: Model fits for the probability of gambling (choosing the risky option) on lotteries drawn from distribution A, plotted as a function of the mean of distribution B. Results are shown separately for risk-averse (top) and risk-seeking (bottom) participants. The $\alpha > 0$ curve shows the fit of the full structure learning model that adaptively infers new reference points, and the $\alpha = 0$ curve shows the fit of the restricted model in which all trials are forced to use the same reference point. Left: Experiment 1 results (mean of B randomized across blocks). Right: Experiment 2 results, where the mean of B increases monotonically across blocks. Error bars represent within-participant standard error of the mean.

Discussion

Reference points play a fundamental role in theories of decision making, yet where they come from and how they change with experience has been an enduring puzzle. The research reported here sheds light on these questions, demonstrating that reference points arise from inferences about latent structure. The key idea is that new reference points are created when the prize distribution undergoes a dramatic change. In contrast, reference points are updated incrementally when the prize distribution undergoes gradual change.

The structure learning account of reference point formation makes a non-trivial prediction, which we confirmed experimentally: whereas small changes in the mean of the prize distribution can shift the reference point up or down, thereby altering risk preferences, large changes will actually have a *diminished* effect due to the creation of a new reference point. In other words, the magnitude of change affects risk preference non-monotonically. This pattern holds both for risk-seeking and risk-averse participants. Importantly, when the mean changes gradually across blocks of the experiment, risk preference changes monotonically, in accordance with our prediction that gradual change will lead to reference point updating rather than the creation of a new reference point.

Our model of reference point formation is a significant generalization of the expectations-based reference point model [2, 3]. The core idea remains the same—reference points reflect expectations—but allows different expectations to form in different contexts. The notion that reference points are context-dependent has been widely acknowledged in psychology, but without a systematic formal treatment like the one proposed here [10, 11, 12, 8, 13].

A strong claim of our theory is that the brain tracks multiple reference points across time, invoking different reference points in a context-dependent manner. Brain imaging could be used to obtain independent evidence for this claim by identifying a neural correlate of the reference point, which could then be used to predict variability in risk preferences. Functional MRI studies have exploited this idea for a fixed structure [28, 29, 30], but have not yet investigated the role of structure learning.

While we have focused on expectations about prizes, the same logic can be applied to other economic variables, such as probabilities and delays. One interesting question is whether reference points apply to the joint space of economic variables, or whether these variables are dissociable, with reference points forming and updating independently. Either scenario could be formalized in our modeling framework, and could be addressed experimentally by orthogonally manipulating the magnitude of changes in different variables simultaneously.

In summary, our findings implicate an important and hitherto unappreciated role for structure learning in decision making. These findings dovetail with results in a diverse set of domains, including memory [15, 31], social cognition [16], categorization [32], and perception [33], all of which involve some form of structure learning. Indeed, our experimental design closely mirrored previous experiments on perceptual judgment [14]. A common modeling framework, based on nonparametric Bayesian inference, can explain many aspects of behavior across these different domains, suggesting that there may be a set of basic computational principles that govern structure learning in the brain.

Materials and Methods

Participants

We recruited 200 individuals to complete the experiment online using the Amazon Mechanical Turk (MTurk) service. Participants had to correctly answer questions to ensure comprehension of the instructions before proceeding. In addition to $2 base pay, each participant was awarded a bonus payment based on the realization of a randomly selected trial. Points were converted into dollars such that the minimum bonus was $1 and the maximum was $2. Participants gave informed consent, and all procedures were approved by Harvard Universitys Institutional Review Board.

Exclusion criteria

In line with recommendations for studies conducted using Amazons Mechanical Turk (AMT) service, careful instructions and a priori exclusion criteria were applied to ensure data quality [34]. Of the 186 participants who progressed to the end of the experiment, Fourteen were excluded for failing to provide responses for greater than 15% of trials overall or 20% of trials within a given block, and an additional 27 participants were excluded from analysis due to biased selection of risky vs. safe, or left vs. right option on over 90% of trials. Finally, 17 participants were excluded because they showed no effect of expected value on choice behavior.

Procedure

Participants completed 200 trials in which they chose between a certain option (guaranteed x points) and a risky option (2x points with probability 0.5). The side of the screen on which the risky option appeared was counterbalanced across trials. Participants were given an unlimited amount of time to make each choice (subject to the 15 minute deadline for completion of the task). Participants did not receive feedback about the outcome of their choice. Once a response was recorded, the task transitioned to a 1.5 second inter-trial interval during which a fixation cross appeared. To discourage the use of simple heuristics (e.g., always choosing the risky lottery gamble) and promote sustained attention, catch trials were embedded randomly within each block. On these trials (4 total across the entire experiment), the options were mismatched so that either the certain or risky option had twice the expected value of the other.

Design

Experiment 1 consisted of 4 conditions (50 trials per condition) presented in random order across participants. Each condition differed only in the mean of B, which took on values of 35, 45, 55, or 65 points. Distribution A had a fixed mean of 35 points across all conditions. The variance of both distributions was equal to 5 points. Experiment 2 was almost identical to experiment 1, except that the conditions were presented in order of increasing mean.

Identification of risk preferences

Participants' were sorted according to risk preference by fitting a logistic regression model to choice data as a function of expected value x_t on trial t, $P(c_t = \text{risky}) = f(\beta x_t)$, where $f(\cdot)$ is the logistic

sigmoid function. Participants whose regression coefficient β was negative were identified as risk-averse, and participants whose regression coefficient was positive were identified as risk-seeking. To avoid confounding this identification procedure with our effects of interest, only choices from the EV(B) = EV(A) = 35 condition were used for this analysis.

Model fitting

Each participant's choice data were fit separately using maximum likelihood estimation of parameters. For the full structure learning model, the free parameters were ψ, ρ, and α. For the restricted model, the free parameters were ψ and ρ. The remaining parameters were fixed as follows for both models: $\sigma_0^2 = 25, \sigma^2 = 0.05, q = 0.005, w = 0.95$. Numerical optimization was used to find maximum likelihood estimates of the free parameters, with 5 random initializations to avoid local optima.

Model comparison

We used two standard metrics for model comparison: the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC).

$$\text{BIC} = -2L + K \ln N$$
$$\text{AIC} = -2L + 2K,$$

where L is the maximum likelihood value, K is the number of parameters, and N is the number of data points. Both metrics balance model fit (L) against model complexity (K), but the BIC penalizes complexity more strongly.

Acknowledgements

S.J.G. was supported by the National Institutes of Health (CRCNS R01-1207833).

References

[1] Kahneman D, Tversky A (1979) Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the econometric society* 47:263–291.

[2] Kőszegi B, Rabin M (2006) A model of reference-dependent preferences. *The Quarterly Journal of Economics* 121:1133–1165.

[3] Kőszegi B, Rabin M (2007) Reference-dependent risk attitudes. *The American Economic Review* 97:1047–1073.

[4] Rigoli F, et al. (2016) A bayesian model of context-sensitive value attribution. *eLife* 5:e16127.

[5] Rigoli F, Mathys C, Friston KJ, Dolan RJ (2017) A unifying bayesian account of contextual effects in value-based choice. *PLOS Computational Biology* 13:e1005769.

[6] van den Assem MJ, Baltussen G, Thaler RH (2008) Deal or no deal? decision making under risk in a large-payoff game show. *The American Economic Review* 98:38–71.

[7] Thaler RH, Johnson EJ (1990) Gambling with the house money and trying to break even: The effects of prior outcomes on risky choice. *Management Science* 36:643–660.

[8] Sullivan K, Kida T (1995) The effect of multiple reference points and prior gains and losses on managers' risky decision making. *Organizational Behavior and Human Decision Processes* 64:76–83.

[9] Kacelnik A, Bateson M (1996) Risky theories–the effects of variance on foraging decisions. *American Zoologist* 36:402–434.

[10] Wang XT, Johnson JG (2012) A tri-reference point theory of decision making under risk. *Journal of Experimental Psychology: General* 141:743.

[11] Koop GJ, Johnson JG (2012) The use of multiple reference points in risky decision making. *Journal of Behavioral Decision Making* 25:49–62.

[12] Neale MA, Bazerman MH (1991) *Cognition and Rationality in Negotiation*. (Free Pr).

[13] Kahneman D (1992) Reference points, anchors, norms, and mixed feelings. *Organizational Behavior and Human Decision Processes* 51:296–312.

[14] Gershman SJ, Niv Y (2013) Perceptual estimation obeys occam's razor. *Frontiers in Psychology* 4.

[15] Gershman SJ, Radulescu A, Norman KA, Niv Y (2014) Statistical computations underlying the dynamics of memory updating. *PLoS Computational Biology* 10:e1003939.

[16] Gershman SJ, Pouncy HT, Gweon H (2017) Learning the structure of social influence. *Cognitive Science* 41:545–575.

[17] Gobet F, et al. (2001) Chunking mechanisms in human learning. *Trends in Cognitive Sciences* 5:236–243.

[18] Mathy F, Feldman J (2012) Whats magic about magic numbers? Chunking and data compression in short-term memory. *Cognition* 122:346–362.

[19] Orbán G, Fiser J, Aslin RN, Lengyel M (2008) Bayesian learning of visual chunks by human observers. *Proceedings of the National Academy of Sciences* 105:2745–2750.

[20] Gershman SJ, Jones CE, Norman KA, Monfils MH, Niv Y (2013) Gradual extinction prevents the return of fear: implications for the discovery of state. *Frontiers in Behavioral Neuroscience* 7.

[21] Kagerer FA, Contreras-Vidal JL, Stelmach GE (1997) Adaptation to gradual as compared with sudden visuo-motor distortions. *Experimental Brain Research* 115(3):557–561.

[22] Wallis G, Bülthoff HH (2001) Effects of temporal association on recognition memory. *Proceedings of the National Academy of Sciences* 98:4800–4804.

[23] Preminger S, Blumenfeld B, Sagi D, Tsodyks M (2009) Mapping dynamic memories of gradually changing objects. *Proceedings of the National Academy of Sciences* 106:5371–5376.

[24] Aldous DJ (1985) Exchangeability and related topics in *École d'Été de Probabilités de Saint-Flour XIII1983*. (Springer), pp. 1–198.

[25] Pitman J, , et al. (2002) Combinatorial stochastic processes.

[26] Bell DE (1995) Risk, return, and utility. *Management Science* 41:23–30.

[27] dAcremont M, Bossaerts P (2008) Neurobiological studies of risk assessment: a comparison of expected utility and mean-variance approaches. *Cognitive, Affective, & Behavioral Neuroscience* 8:363–374.

[28] De Martino B, Kumaran D, Holt B, Dolan RJ (2009) The neurobiology of reference-dependent value computation. *Journal of Neuroscience* 29:3833–3842.

[29] Rigoli F, Friston KJ, Dolan RJ (2016) Neural processes mediating contextual influences on human choice behaviour. *Nature communications* 7:12416.

[30] Rigoli F, Rutledge RB, Dayan P, Dolan RJ (2016) The influence of contextual reward statistics on risk preference. *NeuroImage* 128:74–84.

[31] Gershman SJ, Monfils MH, Norman KA, Niv Y (2017) The computational nature of memory modification. *eLife* 6.

[32] Anderson JR (1991) The adaptive nature of human categorization. *Psychological Review* 98(3):409.

[33] Austerweil JL, Griffiths TL (2011) A rational model of the effects of distributional information on feature learning. *Cognitive Psychology* 63:173–209.

[34] Crump MJ, McDonnell JV, Gureckis TM (2013) Evaluating amazon's mechanical turk as a tool for experimental behavioral research. *PloS one* 8(3):e57410.

[35] Sanborn AN, Griffiths TL, Navarro DJ (2010) Rational approximations to rational models: alternative algorithms for category learning. *Psychological Review* 117:1144–1167.

[36] Wang L, Dunson DB (2011) Fast Bayesian inference in Dirichlet process mixture models. *Journal of Computational and Graphical Statistics* 20:196–216.

Supporting Information (SI)

SI Text

As described in the main text, the agent observes the expected payoff $x_t \in \mathbb{R}$ on trial t and uses this observation to update her belief about the latent cause z_t responsible for the observation, as well as her belief about the mean of the payoff distribution μ_{z_t} associated with the latent cause. Here we provide the details of how the update is computed.

The posterior over z_t is given by:

$$\begin{aligned} P(z_t|x_{1:t}) &\propto P(x_t|x_{1:t-1}, z_t)P(z_t|x_{1:t-1}) \\ &= \sum_{z_{1:t-1}} P(x_t|x_{1:t-1}, z_{1:t})P(z_t|z_{1:t-1})P(z_{1:t-1}|x_{1:t-1}). \end{aligned} \tag{1}$$

Because the sum over $z_{1:t-1}$ is computationally intractable, we use a "local" maximum *a posteriori* (MAP) approximation [32, 35, 15], which replaces the marginalization with a maximization:

$$P(z_t|x_{1:t-1}) \approx P(z_t|\hat{z}_{1:t-1}) \tag{2}$$

$$P(x_t|x_{1:t-1}, z_t) \approx P(x_t|z_t, \hat{z}_{1:t-1}) \tag{3}$$

where $P(z_t|\hat{z}_{1:t-1})$ is the Chinese restaurant process (CRP) and $\hat{z}_{1:t-1}$ is defined recursively according to:

$$\hat{z}_t = \operatorname*{argmax}_k P(z_t = k|x_{1:t}, \hat{z}_{1:t-1}). \tag{4}$$

The local MAP approximation does not in general yield the history of latent causes with the highest posterior probability, because it does not update past assignments after observing new information. However, it is often sufficiently accurate, as attested by its use in machine learning applications [36].

Using the local MAP approximation, the likelihood is given by:

$$P(x_t|z_t = k, \hat{z}_{1:t-1}) = \mathcal{N}(x_t; \hat{\mu}^k_{t-1}, w^2\lambda^k_{t-1} + q + \sigma^2), \tag{5}$$

where $w \in [0, 1]$ is a decay parameter, q is the diffusion noise variance, σ^2 is the observation noise variance, $\hat{\mu}^k_{t-1}$ is the posterior mean for cause k after observing trials 1 to $t-1$, and λ^k_t is the posterior variance. The mean and variance are updating according to the Kalman filtering equations:

$$\hat{\mu}^k_t = w\hat{\mu}^k_{t-1} + \eta^k_t P(z_t = k|x_{1:t-1})(x_t - w\hat{\mu}^k_{t-1}), \tag{6}$$

$$\lambda^k_t = w^2\lambda^k_{t-1} + q + \eta^k_t P(z_t = k|x_{1:t-1})\lambda^k_{t-1}, \tag{7}$$

where η^k_t is the Kalman gain (learning rate), given by:

$$\eta^k_t = \frac{w^2\lambda^k_{t-1} + q}{w^2\lambda^k_{t-1} + q + \sigma^2}. \tag{8}$$

Table S1: P(Gamble|A) as a linear function of EV(B). β = regression coefficient; SE = standard error; Int. = intercept; EV(B) = expected value of distribution B.

	X	β	SE	Pr($>|t|$)
Experiment 1				
	Risk-Averse			
	Int.	0.42	(0.15)	6.6e-03 ***
	EV(B)	2.6e-03	(5.3e-02)	0.96
	Risk-Seeking			
	Int.	- 0.67	(0.19)	4.0e-04***
	EV(B)	-0.10	(6.4e-02)	0.11
Experiment 2				
	Risk-Averse			
	Int.	-0.30	(0.30)	0.31
	EV(B)	0.27	(8.2e-02)	9.9e-04 ***
	Risk-Seeking			
	Int.	-0.11	(0.27)	0.70
	EV(B)	-0.17	(8.7e-02)	0.05