Hedging your bets by learning reward correlations in the human brain

Klaus Wunderlich[1*], Mkael Symmonds[1], Peter Bossaerts[2,3] and Raymond J. Dolan[1]

[1] Wellcome Trust Center for Neuroimaging, University College London, London, U.K.

[2] Division of Humanities and Social Sciences, California Institute of Technology, Pasadena, CA

[3] Swiss Finance Institute, Ecole Polytechnique Fédérale Lausanne, Lausanne, Switzerland

[*] To whom correspondence should be addressed at:

Wellcome Trust Center for Neuroimaging
12 Queen Square
London, WC1N 3BG
United Kingdom
Email: k.wunderlich@ucl.ac.uk

Human subjects are proficient at tracking the mean and variance of rewards, and updating these via prediction errors. Here, we addressed whether humans can also learn about higher order relationships between distinct environmental outcomes, a defining ecological feature of contexts where multiple sources of rewards are available. By manipulating the degree to which distinct outcomes are correlated we show that subjects implemented an explicit model-based strategy to learn the associated outcome correlations, and were adept in using that information to dynamically adjust their choices in a task that required a minimization of outcome variance. Importantly, the experimentally generated outcome correlations were explicitly represented neuronally in right mid-insula with a learning prediction error signal expressed in rostral anterior cingulate cortex. Thus, our data show that the human brain represents higher order correlation structures between rewards, a core adaptive ability whose immediate benefit is optimized sampling.

Introduction

Risk is ubiquitous in nature with predation, starvation, adverse environmental change, or lack of reproductive opportunity acting as constant background variables that shape an animals' behavior. Animals evolved a variety of strategies to minimize risk such as diversifying mating behavior (Fox, 2003) or 'bet-hedging'. For example, desert bees mitigate against large temporal variability in rainfall by stabilizing their birth rate (Danforth, 1999; Hopper, 1999). These risk-spreading strategies act to minimize between-year variance in reproductive success in a similar way to cost averaging, where financial investors periodically purchase risky asset to reduce the overall risk of an investment portfolio (Dodson, 1989). Our concern here is with risk as defined by outcome variability, measured from the variance of an outcome distribution. This is a first-order approximation of risk commonly used as a critical decision variable in ecological (Stephens, 1981) and financial (Markowitz, 1952) decision analysis.

While the aforementioned strategies are naive with respect to higher order structure in the environment, organisms can reduce risk even more effectively if they deploy knowledge of how different environmental states occur in relation to each other by representing correlations (Yoshimura and Clark, 1991). Thus, a lion learning that buffalo congregate at water holes on hotter days can reduce the chance of starvation by allocating more predation time to this food source by simply registering that the weather on a particular day is hot. In effect, knowledge of a covariance structure between discrete events allows inferences as to the presence, or in many instances quantity, of one outcome merely by observing a complementary event without actually having to sample on the inferred one.

Risk minimization is also a key concept in financial and insurance markets. Hedging, the process of combining multiple positions in different assets to reduce total risk in a portfolio is a common risk minimization strategy in financial investments (Jorion, 2009). Modern portfolio theory (MPT)(Markowitz, 1952) formalizes the concept of risk-spreading and relies upon correlations between multiple assets to specify how they can be most efficiently combined to maximize returns and minimize risk. Research in decision neuroscience provides extensive evidence for a neural representation of key decision variables (Doya, 2008) with a focus heretofore on value signals, putative inputs to the decision process such as action or goal values, and representations of expected outcome after a choice (Hampton et al., 2006 ; Knutson et al., 2005; Lau and Glimcher, 2007; Padoa-Schioppa and Assad, 2006; Plassmann et al., 2007; Samejima et al., 2005; Wunderlich et al., 2009; Wunderlich et al., 2010). There is now good evidence that fundamental computational mechanisms underlying value-based learning and decision-making are well captured by reinforcement learning algorithms (Sutton and Barto, 1998) where option values are updated on a trial by trial basis via prediction errors (PE) (Knutson and Cooper, 2005; Montague and Berns, 2002; O'Doherty et al., 2004; Schultz et al., 1997). More recently, there is an emergent literature that suggests the brain not only tracks outcome value, but also uncertainty (Huettel et al., 2006; Platt and Huettel, 2008) and higher statistical moments of outcomes such as variance (Christopoulos et al., 2009; Mohr et al., 2010; Preuschoff et al., 2006; Preuschoff et al., 2008; Tobler et al., 2009) and skewness (Symmonds et al., 2010).

An important component of outcomes, namely the statistical relationship *between* multiple outcomes, and what neural mechanisms might support acquisition of this higher order structure has remained unexplored. In principle, there are several plausible mechanisms including the deployment of simple reinforcement learning to form individual associative links (Thorndike,

1911), or a more sophisticated approach that generates decisions based upon estimates of outcome correlation strengths. If the latter strategy is indeed the one implemented by the brain then this entails a separate encoding of correlations and corresponding prediction errors beyond that of action values and outcomes.

Here, we address the question of how humans learn the relationship between multiple rewards when making choices. We fitted a series of computational models to subjects' behavior and found that a model based on correlation learning best explained subjects' responses. Furthermore, we found evidence for a neural representation of correlation learning evident in the expression of fMRI signals in right medial insula that increased linearly with the correlation coefficient between two resources, a normalized measure of the strength of their statistical relationship. A correlation prediction error signal, needed to provide an update on those estimates, was represented in rostral anterior cingulate cortex and superior temporal sulcus. To our knowledge, these behavioral and neural data provide the first evidence that subjects learn the correlative strength between rewards and are able to use this information to make risk-optimal choices.

Results

To investigate how humans learn correlations between outcomes we scanned 16 subjects using fMRI while they performed a 'resource management' game. This task invoked a scenario whereby a power company generates fluctuating amounts of electricity from two renewable energy sources, a solar plant and a wind park. We instructed subjects to create an energy portfolio under a specific goal constraint necessitating keeping the total energy output as constant as possible (**Fig. 1A**). Subjects accomplished this by adjusting weights that determined

how the two resources were linearly combined. A normative best performance is achievable by finding a solution that exploits knowledge of the covariance structure of these resources (**Fig. 1B**), a task design that approximates a simple portfolio problem in finance. Importantly, the outcomes of the two resources co-varied with each other and this correlation between the two outcomes changed probabilistically over time, requiring subjects to continuously update their estimate of the current correlation structure. This task is well suited for assessing subjects' estimate of the correlation strength because a good performance is only accomplished if subjects learn both the distribution of returns for each resource as well as their correlation. We rewarded participants according to how stable they kept the total output of their mixed energy portfolio relative to the variance resulting from an optimal strategy (specified by MPT-calculated optimal weights).

Behavioral model comparison

We speculated that subjects might solve the task by learning the correlative strength between the resources via a correlation prediction error, calculated from the cross product of the individual resources' outcome prediction errors (**Fig. 1C**). This envisages that subjects represent a continuous measure of outcome correlation and update this metric on a trial-by-trial basis. To rule out alternative strategies we examined other computational models that could be used to guide choice in our task, and fitted the free parameters of each model to get model predicted portfolio weights that most closely resembled the actual responses for each subject.

One such alternative model-based strategy is to exploit trial-by-trial evidence to update a representation of the portfolio weights directly instead of first estimating the correlation coefficient. Similar to correlation learning, this model makes assumptions about the structure of

the task and uses individual resource outcomes as a basis for learning. The main difference between the covariance based model and this model is that in the former, subjects update an estimate of the correlation via a prediction error and then translate this correlation strength into task specific weights on every trial, while in the latter the estimate of task dependant weights (i.e. the position on the response slider) are learnt directly. This differentiation is important because the correlation coefficient is a normalized and therefore universal measure of the interdependence between the two outcomes, while appropriate mixing weights are task specific and would need to be relearned if the variances of the individual outcome change or the goal of the task changes from risk minimization to maximization. Both of these strategies are model-based as they require an understanding of how the two individual outcomes interact. There are other potential modes of learning that we also consider. For example, subjects might implement a more simple model-free reinforcement learning based on Q-learning of action values for increasing or decreasing the weights. In contrast to the former approaches requiring subjects to attend to the individual resource outcomes, a subject who updates action values in this model-free way would instead consider the mixed portfolio outcome in every trial and try to minimize its temporal fluctuation using simple outcome based updating. Any change in behavior following a change in correlation between resources would then be due to a relearning of a new optimal mix of actions rather than a more complete knowledge of the structure of the environment. Finally, subjects might use a heuristic of detecting coincidences in the occurrence between outcomes, without a full representation of the strength of correlation.

Out of all tested models, the model based on tracking the correlation coefficient best predicted subjects' behavior (**Fig. 2A, Table 1**). The weights estimated by this model match subjects' behavior very well, as shown by a comparison of model predictions and subjects' actual choices

(**Fig. 2B**) with the regression of actual observed weights on model predicted weights being highly significant in every individual subject ($p<<0.0001$; average R^2 across subjects = 0.77; Table S1). In fact, subjects' responses approximated normatively optimal portfolio weights while subjects attempted to keep the total energy output stable (minimize variance) (**Fig. 2C**). Both model predicted and subjects' actual responses approach normatively optimal weights with some lag, the latter resulting from a need to have multiple observations to reliably detect any change in correlation strength. In effect, subjects' strategy of determining the correlation approximately compared to a normative calculation of the correlation coefficient over the outcomes of the past 10 trials.

Neural representation of the correlation strength

If indeed the brain learns the relationship between two rewards by estimating their covariance then this predicts that we should observe a neural representation of the computations that support this process. Consequently, we tested for fMRI signals that track the covariance or correlation strength, and because the outputs vary, there should also be a signal that updates this information. Based on prior evidence, we predicted activity related to covariance would be seen in insular cortex or striatum, areas implicated in encoding the risk or variance of individual outcomes (Preuschoff et al., 2006; Preuschoff et al., 2008). Consequently, we modeled subjects' trial-by-trial estimates of the correlation coefficient and regressed those model-predicted time series against simultaneously acquired fMRI data. We found BOLD activity in right mid-insula varied with the correlation strength between the outputs of the solar and wind power plants (xyz = 48, 5, -5; Z=4.12; $p<0.001$ FWE corrected; **Fig. 3A**). Right insula was the only region to

survive cluster level whole brain correction and we provide a comprehensive list of all activated areas at a lower threshold ($p<0.001$ uncorrected) in Table 2.

We next determined whether the correlation strength is represented either as covariance, a raw measure of how much the two variables fluctuate together, or as the correlation coefficient, a scale invariant metric of the covariance normalized by the standard deviation of each resource. We estimated two additional models using Bayesian estimation, with either the covariance or the correlation coefficient as parametric modulator, and compared the ensuing log-evidence maps in a random effects analysis. Activity in right mid-insula was better described by the correlation coefficient than by covariance (exceedance probability of $p>0.999$). The linear relationship between correlation coefficient and BOLD is visualized in a binned effect size plot (**Fig. 3B**).

We then verified whether this signal was more strongly represented at the time of outcome, when new evidence is available to update estimates, or at choice when subjects actively readjust their allocated weights for the two resources (**Fig. 3C**). In addition to plotting the effect timecourse we tested these neural hypotheses by estimating a design where the correlation coefficient acted as an unorthogonalized parametric modulator of activity at both the time of outcome and time of choice. In this analysis we observed significant effects of correlation strength solely at the outcome time ($Z = 3.60$, $p = 0.01$ FWE corrected) but not at the time of choice ($Z = 2.40$, $p = 0.02$ unc.).

If our behavioral model explains subject's choices and subjects' brain activity represents crucial decision variables in this process then we would expect that brain activity should be particularly well explained in those subjects in whom our model also provides a good choice prediction. This would be expressed in a relationship between the behavioral model fit and the model fit in the

GLM against BOLD data. Consistent with our conjecture, we found a significant positive correlation between R^2 in the behavioral model and R^2 in the MRI analysis ($r = 0.50$, $p < 0.03$; **Fig. 3D**). In effect, this confirms that our model explains a larger proportion of the fluctuation in the neuronal data in those subjects in which the model can also well explain choices.

Neural correlates of correlation prediction errors

A neural representation of correlation strength in our task entails that this estimate is updated over time, a process ascribed to a prediction error signal. Analogous to risk prediction errors for individual rewards (Preuschoff et al., 2008), the cross products of the two outcome prediction errors provide a trial-by-trial estimate of the covariance strength. Using this regressor we found that a correlation prediction error was tracked in fMRI activity in left rostral cingulate cortex (xyz = -15, 44, 7; Z = 4.87; p<0.003 FWE corrected; **Fig. 4;** Table 2).

From correlation to portfolio weights

After observing an outcome, participants may have an imperative to change the slider position if their currently set weights deviate from the estimated new best weights, in other words if they are sub-optimal. We tested for a signal corresponding to the absolute (i.e. unsigned) deviation between current and new weights on the next trial and found corresponding BOLD activity in a region encompassing anterior cingulate (ACC) / dorsomedial prefrontal cortex (DMPFC) (xyz = 6, 26, 34; Z=4.22; p<0.001 FWE corrected) and in right anterior insula (xyz = 42, 23, -5; Z=4.04; p<0.04 FWE corrected) at the time of the outcome (**Fig. 5;** Table 2). In contrast, no areas corresponded directly to the portfolio weight values or a signed updating of weights, signals one

would expect if subjects performed learning over task specific weights instead of the correlation structure between outcomes.

Finally, an optimal solution to our task requires learning of the individual outcome variances in addition to learning the covariance structure. When we tested for neural activity coupled to local temporal fluctuations in the individual outcome variances we replicated previous findings in highlighting a neural representations of outcome risk in striatum (xyz = -18, 5, 10; Z=3.81; p=0.04 small volume corrected; Fig. S3).

Alternative model considerations

As an alternative to learning the correlation coefficient subjects might directly learn the weight representation and perform RL over the weights instead of the correlation coefficient. If that were the case then one would also expect to find a neuronal representation of the weights and weight prediction errors, which were conspicuously absent in our data. Another possibility could be that subjects simplified the problem to detecting outcome coincidences (both outcomes either above or below mean versus one outcome above and the other below mean) instead of fully quantifying the trial-by-trial covariance. In that case we would expect to find a neural signal pertaining to mere outcome coincidences. We found no activations coupled to either the weight or the weight prediction errors, or the trial-by-trial coincidences anywhere in the brain at our omnibus cluster level threshold of $p<0.05$. Together with the inferior behavioral fit of the coincidence model this suggests that subjects quantified the trial-by-trial relationship between outcomes. We also implemented a model-free Q-learning algorithm as further alternative strategy, which was clearly outperformed by the correlation model.

Discussion

We show that human subjects are adept at learning correlations between two dynamic variables, a process also represented neurally. Subjects were highly effective at exploiting this key metric of the statistical relationship between the two individual resources to guide choice in a task requiring minimization of outcome fluctuations. This finding is in contrast to an often proposed model in behavioral finance, which suggests disregarding environmental structure and using fixed weights according to the 1/N rule (Benartzi and Thaler, 2001). Our subjects performed better than this simple heuristic and learnt a more optimal strategy through repeated observations. At a neural level, fMRI signals in right mid-insula were coupled to the current correlation coefficient while activity in rostral anterior cingulate encoded a correlation prediction error, a signal used to update an estimate of the correlation strength based on new evidence in every trial.

Although learning individual outcomes is a central part of decision making, the availability of different rewards are rarely independent of each other in a natural environment. Our results provide the first evidence that subjects also learn the relationship between multiple outcomes by tracking their correlation, and can use this information to decrease overall sampling risk. Commonly observed risk aversion in animals (Kacelnik and Bateson, 1996) and humans (Tversky and Kahneman, 1981) is rational in an evolutionary context, as a small but constant supply of food that always exceeds the critical minimum for survival is far more beneficial to

viability than periods of alternating deficiency and extreme excess. In some other instances risk seeking behavior may occur, such as in gamblers, and may promote exploration and learning. Note however that also in that case a representation of the correlation in the environmental structure is beneficial as this information can be used both for risk minimization or maximization.

To generalize our results to more natural situations we have to ascertain that the findings reflect a specific mechanism of correlation learning instead of incidental task variables. Plausible possibilities include shortcuts such as learning the position on the response slider by a model-free gradient descent mechanism, or using a model-based strategy but without representing individual outcome variances and normalized correlation coefficients and instead directly learning a representation of the portfolio weights. Our behavioral and neural data render all these explanations very unlikely. The best fitting learning rate for outcome variance is similar to the learning rate for correlation and significantly above the one for value for each individual subject. Importantly, we ensured that the signals in our study were not spurious reflections of the individual variances of solar and wind plant outputs by explicitly modeling these signals with additional (unorthogonalized) parametric regressors. A fluctuating trial-by-trial estimate of the outcome variance is also represented neurally in striatum (Fig. S3), an area previously implicated in variance learning (Preuschoff et al., 2006). While these neural signatures of risk and risk prediction errors were somewhat weaker compared to covariance signals, we suggest this observation is due to an amalgamation of signals tracking the two separate resource variances within the same area, and because the variance of the two outcomes fluctuated only slightly over the course of each experimental block. Importantly, we found no significant correlations with

signals pertaining to alternative decision models anywhere in the brain at $p<0.05$ corrected. Specifically, we examined if there was evidence for a direct representation of desired resource weights, or weight prediction errors, signals one would expect instead of the correlation coefficient if subjects used a more task specific strategy. We also did not find significant correlations with a more qualitative measure of coincidences instead of fully quantified correlations. Together with a superior behavioral fit of the correlation learning model, this strongly supports the specificity of our neural results and effectively discounts the possibility that the observed activations here relate to incidental task related learning processes instead of learning the correlation between outcomes.

We found that anterior insula tracked the correlation strength between the outputs in a site slightly posterior to regions previously implicated in tracking variance (Mohr et al., 2010; Preuschoff et al., 2008). Combined these findings suggest that insular cortex may support a general role in processing statistical information about the environment. At the same time, anterior insula has been implicated in representing bodily states and their translation into feelings and possibly awareness (Craig, 2009). Note that the calculus like role proposed here does not contradict the idea that anterior insula represents subjective aspects of experience. Indeed, the Somatic Marker Hypothesis postulates that rational decision theory requires emotional anticipation of outcomes (Bechara et al., 1997), such that seemingly prudent behavior and emotional decision making are intertwined (Paulus et al., 2003). The finding of a slightly posterior encoding of correlation relative to risk also tallies with a structural model for how unconscious state representations might be integrated into a sentient self along a posterior to anterior insula (Craig, 2009). Adequate emotional risk assessment is immediately relevant for fight or flight responses and might therefore require a more direct link to awareness then the

meta parameters of how multiple such variables relate to each other (Bossaerts, 2010). The latter assessment is largely subconscious and may, as implicit function, also be enacted during low-level processing of multidimensional stimuli such as music and rhythm. Interestingly such tasks have previously been associated with insula activation (Koelsch et al., 2006; Platel et al., 1997). Our data show that the brain encodes the correlation coefficient of two outcomes, a normalized value, instead of the covariance itself. In light of previous data (Bunzeck et al., 2010; Padoa-Schioppa, 2009; Seymour and McClure, 2008) this hints that scale invariance is a ubiquitous concept in encoding decision variables in the brain.

The representation of a prediction error in anterior cingulate fits neatly with mounting evidence that this area is involved in learning and behavioral control. Several previous studies report a role for anterior cingulate in an error-driven reinforcement learning system (Kennerley et al., 2006), and in prediction errors for actions (Matsumoto et al., 2007) or social value (Behrens et al., 2008). Together with risk prediction errors in anterior insula (Preuschoff et al., 2008) this teaching signal for correlation strength might belong to a broader system involved in learning the statistical properties of the environment.

We also observed an anticipatory signal reflecting an impetus to shift resource allocations on the next trial in order to keep the total energy output stable. Interestingly, this signal was expressed in a dorsomedial prefrontal cortex (DMPFC) cluster previously linked to updating learning in relation to environmental volatility (Behrens et al., 2007), implying a more general role for this region in adapting behavior to fluctuations in the statistical characteristics of the environment. Most task-modulated activity, including correlation strength, its prediction error, and a signal reflecting the need to alter responses, occurred at the time of outcome rather than at choice. This

suggests that task-relevant computations, including an evaluation of the appropriate action to take after each outcome, occur at the point when individuals can best harvest new evidence. As we focused on the mechanism of learning the correlation strength, rather than on how subjects use this information, this raises the question of how exactly information about a covariance structure is applied in a natural sampling environment. Here, we instantiated this mapping of correlation coefficients into energy resource weights by using the normative function derived from MPT. We assume subjects learnt the form of this non-linear transformation during initial training, but it remains a question for future research how this translation is applied. Based on our present results and previous findings that the brain encodes other statistical parameters such as variance and skewness of outcomes (Preuschoff et al., 2008; Symmonds et al., 2010), we speculate that in more naturalistic environments subjects form structural representations of the world by encoding summary statistical parameters. Such a parameterized representation is both efficient and flexible: the optimal response is dependent upon three parameters, the magnitude, variance and correlation of the available resources, and knowledge of the individual parameters allows fast adaptation in light of changes to any one of them. One way to expand our research to more natural situations could be by changing the cost function to mimic an ecological survival game with perishable outcomes. Such a paradigm would allow one to determine if subjects indeed follow a variance minimizing strategy and incorporate information about reward correlations.

The recent financial crisis has amply demonstrated that even experts have difficulties regulating correlated risks in the financial domain and investors often deviate from rationality when making financial decisions (Daniel et al., 2002; Kuhnen and Knutson, 2005). In contrast, we show here

that individuals are adept at detecting and responding to correlations and appropriately selecting actions to minimize risk in an intricate learning task. Indeed, this exquisite sensitivity taps into an adaptive and evolutionary conserved ability of implicit neurobiological systems to learn environmental reward structure through trial-by-trial sampling; intrinsic behavior that might even supersede that of financial experts deciding about explicitly described statistics.

Experimental Procedures

Subjects

16 healthy subjects (7 female; 18–35 years old) with no history of neurological or psychiatric illness participated in the study. Two additional pilot subjects from the lab were excluded from the final analysis, as they were already familiar with the hypotheses in the experiment. The study was approved by the Institute of Neurology (University College London) Research Ethics Committee.

Task

To investigate whether and how subjects learn the reward structure in the environment we designed a novel portfolio-mixing task in which knowledge of the correlation between two resource outcomes could improve performance. Subjects' task was to keep the combined output of two power stations as stable as possible (i.e. minimize the variance of an energy portfolio) by mixing the fluctuating outcomes of these two individual resources. They accomplished this by adjusting weights which determined how the resources were linearly combined. A normative best performance is achievable in this task by finding a solution that directly depends on knowledge of the covariance structure of these resources, a task design that approximates a simple portfolio problem in finance (Markowitz, 1952).

We presented the task to subjects as a resource management game that invoked a scenario whereby a power company generates fluctuating amounts of electricity from two renewable

energy sources, a solar plant and a wind park. The resource outputs r_{sun} and r_{wind} were drawn as random numbers in every trial from distributions with a common mean M and variances σ^2_{sun} and σ^2_{wind}. Importantly, the two outcomes co-varied with each other, and the strength of this correlation changed probabilistically over time. This feature encouraged subjects to form an estimate about the mean and variance of the individual outcomes and continually update their assumption about the correlation strength.

Subjects participated in three consecutive experimental blocks, each corresponding to a 21 min long session in an fMRI scanner (Siemens Trio 3T). They were instructed that the correlation would probabilistically change over the course of the study but were not given further details about specific parameters used. We also told subjects that the mean and variance of the two resources would remain constant over one block of the experiment, a simplification to an otherwise quite complex task that enabled subjects to perform well within the settings of this experiment. As our goal was to assess covariance learning (in contrast to learning the values and risk) this did not adversely impact on any mechanism we wanted to observe. However, mean and variance values were different for each block. To give subjects the opportunity to learn these basic statistical parameters (mean and variance) before making portfolio choices, we presented them with a 20-trial observation phase at the beginning of each session. In this phase, which immediately preceded the start of fMRI data acquisition, subjects only observed the individual outcomes of the two resources and did not make any choices. There was no change in the ground truth correlation during this phase. Data from pilot studies and model simulations confirmed that 20 observations of a time series were sufficient to form an estimate of its mean and variance. The observation phase was followed by 84 choice trials, consisting of a 5 sec choice period and a 3 sec outcome period, separated by a blank grey screen of 1-6 sec duration (uniform distribution). The inter-trial-interval was also 1-6 sec (**Fig. 1A**).

The portfolio weights (w_{sun}, w_{wind}) indicate how much of a fraction the portfolio contains from both resources r_{sun} and r_{wind} (Portfolio outcome value $V_p = w_{sun}*r_{sun} + w_{wind}*r_{wind}$). Subjects were allowed to set the portfolio weight w_{sun} within a range between [-1, 2]. Setting negative weights allowed subjects to trade-in a fraction of the trials output from one resource in exchange for multiplying the other output by the same fraction. This concept echoes the possibility of short

selling in financial markets and is important for this task as it permits risk minimizing for positively correlated resources (see section variance minimizing strategies in the supplemental text for further details). The constraint that both weights always add up to 1 automatically determined the weight of the other resource ($w_{wind} = 1\text{-}w_{sun}$). A horizontal line on the choice screen represented the slider during the choice period and icons of a solar and wind plant on both ends indicated which resources were mixed in the portfolio. The parts of the slider involving a negative weight were red while the middle part with both positive weights was shown in white with the center position corresponding to a mix with equal weights. A yellow dot on the slider indicated the current position and portfolio weights were additionally shown numerically next to the resource icon. Subjects were able to make responses during the entire 5s choice period by pressing two buttons on a button box with their right hand. Each button press moved the current slider position a discrete step of 0.1 units in either direction. Moving the slider a step towards the right always increased the weight for sun and decreased the weight for wind. A new choice period started with the portfolio weights from the last trial and subjects were allowed to freely move the slider as many steps in either direction as they wished during the choice period. Importantly, subjects always had to determine the weights for the current trial prior to seeing the actual outcome. Due to inherent stochastic outcomes, and because serial outcomes were independently drawn, the only rational strategy was to set the weights in a way that would yield the least portfolio variance in the long run and this measure depended on the current correlation.

To determine subjects' performance we benchmarked their portfolio fluctuation against the fluctuation of a portfolio with optimal weights. The normative solution was calculated by the risk minimizing formula of portfolio theory (see supplemental text for details). This ensured that subjects were fairly scored given the stochastic outcomes on a trial-by-trial basis (i.e., even if subjects played optimally the portfolio outcome would fluctuate around the target with the amount of fluctuation dependent on the current covariance). Subjects received reimbursement of 10£ flat plus a fraction of the maximum bonus of 45£ in relation to task performance (Table S1). All participants received basic instructive information about hedging strategies (similar to the supplemental text 'variance minimization strategies' and Figure S2) and practiced the task (same number of trials than in the fMRI study but with different parameters for outcome mean and variance) on a separate day prior to scanning. Note, however, that all instructions concerned

exclusively how to set portfolio weights (i.e. how to respond) but not how to learn correlations itself. Therefore this latter process cannot be confounded by the explicit information given here. The reason for using a seemingly intricate portfolio task over having subjects merely report the correlation directly is that explicit assessments of decision variables by self-report are often biased (Kagel and Roth, 1997). Our procedure is in this respect very similar to other commonly used behavioral measures such as auction biddings (Becker et al., 1964; Plassmann et al., 2007) to identify subjects' unbiased value preference. Another advantage of our task is that response behavior does not depend on individually subjective valuation or risk preference. Performance and payout were only related to how close subjects' behavior matched the normative optimal solution (thereby incentivizing an accurate correlation representation) but was independent of the actual amount or variance of the produced energy mix.

Importantly, during the experiment subjects never received direct feedback on their performance at minimizing energy fluctuations (i.e. only saw trial-by-trial outcomes) and the bonus and optimal weights were only revealed after the experiment. We omitted feedback during the task to prevent subjects from using a strategy that is based on optimizing the performance feedback instead of learning the correlation of the individual outcomes. While the portfolio value is shown on every trial, and the deviance of this value from its mean gives some hints to performance, this is only a crude measure of whether the current weights are good because even with optimal weights the amount of portfolio fluctuation depends on the current correlation.

Because the optimal mixing weights (portfolio weights) in our task depend on individual variance from solar and wind power plants and their correlation strength, the best strategy is to learn the variances and correlations by observation of individual outcomes and then translate these estimates into an optimal resource allocation (i.e. weightings). While subjects could learn the statistical properties underlying outcome generation by observation, the outcomes of individual trials were unpredictable. Their task was then to continuously mix the two resources into an energy portfolio and thereby minimize the fluctuation of the portfolio value from trial to trial.

Generation of outcome values

Both resources fluctuated around a common mean, with outcomes drawn from a rectangular distribution with a specific variance. In our task the standard deviation of one resource was always twice that of the other because this maximized the influence of the correlation on the portfolio weights (see Fig. S1 for details). The sequence of correlated random numbers for the two resources were generated by the Cholesky decomposition method (Gentle, 1998). This was realised by first drawing random numbers x_A and x_B for resources A, B from a rectangular distribution. The outcome of the second resource x_B was then modified as $x_B = x_A * r + x_B*sqrt(1-r^2)$, whereby r is the generative correlation coefficient. Finally, x_A and x_B were normalized to their desired standard deviations (in the three blocks: 20/10, 15/30, 10/20) and common means (30, 50, 40). We chose a rectangular distribution to increase the sensitivity of our fMRI experiment in finding neural correlates of covariance and covariance prediction errors as the linear regression against BOLD activity is most sensitive if the values of the parametric modulators are distributed along their entire range. This is not true for normal distributed outcomes, which have proportionally the largest amounts of data close to the mean.

We varied the generative correlation strength in discrete steps of [-.99, -.3, .4, .7, .95, .999]. The observable correlation through sampling by the subject will however very on a continuous scale also between these steps due to stochasticity in the outcomes. A change from the current to a new correlation was determined probabilistically in every trial with a p=0.3 transition probability, under the constraint that a change would only occur after the new correlation became theoretically detectable by an ideal observer that was tracking the correlation coefficient in a sliding window over the past 5 trials. In detail, after the normatively estimated correlation based on the last 5 trials (similar to the sliding window model below) approached the new generative correlation (with a deviation smaller than 0.2), the correlation was allowed to change on all further trials. This prevented overly rapid changes in the generative correlation before subjects could have possibly detected the new correlation coefficient from outcome observations. On average (across subjects and sessions) the correlations changed every 10 trials. To discourage subjects from persevering on a more favorable spot of the response scale that would give a

reasonable result over a wider range of correlations, and instead be forced to track the correlation explicitly, we further implemented an adaptive rule that if subjects' response was both suboptimal (farther from the optimum than 0.2) and they did not change their response within the past 5 trials then the correlation would jump to the farthest extreme (either -.99 or +.999). This increased the penalty on subjects payout at their current weights and ecouraged them to find a better weight allocation. In practice, this constraint came rarely (never for 10 subjects, one or two occurrences in five, and three occurrences in one subject) into use during the fmri experiment.

Correlation learning model

We modeled trial-by-trial values of the correlation strength by using principles of reinforcement learning (Sutton and Barto, 1998). Reinforcement learning generates in every trial a prediction error as the deviation of the experienced outcome R from the predicted outcome. Those prediction errors, multiplied by the learning rate, are then used to update predictions in future trials.

$$\text{resource value: } V_{i,t+1} = V_{i,t} + \alpha^{V}\delta_{i,t} \quad (1)$$

$$\text{value prediction error: } \delta_{i,t} = R - V_{i,t} \quad (2)$$

The squared prediction error is also a measure of the outcome fluctuation and thereby a quantifier of risk. A sequence of continuously large prediction errors indicates that the outcomes greatly fluctuate, whereby a sequence of small prediction errors indicate that prediction is precise with little deviation. We used this to model the risk h for both resources:

$$\text{resource variance: } h_{i,t+1} = h_{i,t} + \alpha^{R}\varepsilon_{i,t} \quad (3)$$

$$\text{variance prediction error: } \varepsilon_{i,t} = \delta_{i,t}^{2} - h_{i,t} \quad (4)$$

We then extended this model from independent outcomes to the interaction of outcomes, whereby the product of the individual prediction errors measures the co-variation of two outcomes:

$$\text{resource covariance: } cov_{t+1} = cov_{t} + \alpha^{C}\zeta_{t} \quad (5)$$

$$\text{covariance prediction error: } \zeta_{t} = h_{s1,t}\, h_{s2,t} - cov_{t} \quad (6)$$

The correlation coefficient ρ was then defined as the covariance normalized by the individual standard deviations of the two involved outcomes:

$$\text{resource correlation: } \rho_t = cov_t / (sqrt(h_{s1,t}) * sqrt(h_{s2,t})) \quad (7)$$

In every trial the correlation coefficient was finally translated into a position on the response slider using the normative function $(h_{2,t} - cov_t) / (h_{1,t} + h_{2,t} - 2*cov_t)$, which is derived in the supplemental text. This relationship (Fig. S1) did not change over the entire course of the experiment (because we always used the same ratio of 1:2 between outcome σ).

We kept the mean of the resource outcomes constant during each session and therefore the optimal strategy was indeed to not update those variables once a reliable estimate had been formed during the observation phase of each block. In fact, the best fitting learning rate for resource values was consistently very small across subjects (average 0.08), confirming that, as intended by the design, subjects indeed treated the mean as a stable value after the initial observation period and adjusted their learning rate downwards to reflect this steady nature (Behrens et al., 2007).

We investigated whether subjects used different learning rates for variance and covariance learning or whether these processes could be described by a single parameter. We did this by comparing a model with separate parameters for variance and covariance learning with a model that used a common parameter for both learning processes. We found that the reduced model could describe learning as well as the full model if model complexity is considered (Table 1). Note that both overall mean value and variance were constant during the experiment but the best fitting learning rate for variance was higher than for value. This suggests that, in contrast to mean outcome value, subjects continuously updated their estimate of individual risk in response to local temporal fluctuations in the individual variances.

We therefore used the reduced model with a common risk/covariance learning parameter to generate fMRI regressors. Parameter estimates were fit for every individual subject using least squares minimization between model predicted weights and actual weights set by the subject (see below).

Alternative models

We created several alternative models that do not require learning of covariance information. Those models are described in the supplemental experimental procedures.

Model comparison

We compared how well each model predicted subjects' behavior by fitting the free parameters of each model such that the mean squared sum of the deviation between model predicted (w^m) and subjects' weights (w^s) was minimized. As measure of model fit we then calculated the Bayesian information criterion (BIC) (Schwarz, 1978) as

$$\mathrm{BIC} = 2L + k\ln(n) \tag{8}$$

$$L = \frac{1}{2}n\left(\log(2\pi) + \log\left(\sum_i^n (w_i^s - w_i^m)^2 / n\right) + 1\right) \tag{9}$$

where L is the negative log likelihood function, n = 252 trials and k the number of free model parameters (Table 1). We also calculated a generalized r^2-statistics for each model, which is a standardized measure of model fit analogous to accounted variance (Nagelkerke, 1991). It is computed as $r^2 = 1 - \frac{L}{L_{random}}$.

Stimuli

Stimuli were presented on a gray background using Cogent 2000 (www.vislab.ucl.ac.uk/cogent.php) running in MATLAB. Stimuli were presented using an LCD projector running at a refresh rate of 60 Hz, viewed by subjects via an adjustable mirror.

FMRI data acquisition

Data were acquired with a 3T scanner (Trio, Siemens, Erlangen, Germany) using a 12-channel phased array head coil. Functional images were taken with a gradient echo T2*-weighted echo-planar sequence (TR = 3.128 s, flip angle = 90°, TE = 30 ms, 64 × 64 matrix). Whole brain coverage was achieved by taking 46 slices in ascending order (2 mm thickness, 1 mm gap, in-plane resolution 3 × 3 mm), tilted in an oblique orientation at -30deg to minimize signal dropout in ventrolateral and medial frontal cortex (Weiskopf et al., 2006). Subjects' head was restrained with foam pads to limit head movement during acquisition. Functional imaging data were acquired in three separate 415-volume runs, each lasting about 21 minutes. The first five volumes of each run were discarded to allow for T1 equilibration. A B0-fieldmap (double-echo FLASH, TE1 = 10 ms, TE2 = 12.46 ms, 3x3x2 mm resolution) and a high-resolution T1-weighted anatomical scan of the whole brain (MDEFT sequence, 1x1x1 mm resolution) were also acquired for each subject.

FMRI data analysis

Image analysis was performed using SPM8 (rev. 3911; www.fil.ion.ucl.ac.uk/spm). EPI images were realigned and unwarped using field maps (Andersson et al., 2001). Each subject's T1 image was segmented into gray matter, white matter, and cerebrospinal fluid, and the segmentation parameters were used to warp the T1 image to the SPM Montreal Neurological Institute (MNI) template. These normalization parameters were then applied to the functional data. Finally, the normalized images were spatially smoothed using an isotropic 8-mm full-width half-maximum Gaussian kernel.

Functional MRI (fMRI) time series were regressed onto a composite general linear model (GLM) containing regressors representing the time of the choice, the time of the outcome screen, and any button presses during the choice period. The outcome regressor was modulated by a number of model derived decision variables. Modulators for outcome were: prediction errors for the individual resource outcomes and the portfolio outcome (δ_1, δ_2, δ_p), the absolute deviation of the portfolio outcome from the target ($|\delta_p|$), resource risk (h_1, h_2), risk prediction errors (ε_1, ε_2), the correlation strength of the resources (ρ), and the correlation prediction error (ζ). A further modulator captured the anticipated magnitude of actual weight updating in the next trial ($|w_t$-

w_{t+1}|). In contrast to the default procedure in SPM, we entered all regressors and modulators independently (without serial orthogonalization) into the design matrix. Thereby only the additional variance that cannot be explained by any other regressor is assigned to the effect, preventing spurious confounds between regressors (Andrade et al., 1999; Draper and Smith, 1998). Specifically, this ensured that the observed effects of correlation strength and correlation prediction error are solely accountable by effects not explained by signals relating to the variance of individual outcomes.

The regressors were convolved with the canonical HRF, and low frequency drifts were excluded with a high-pass filter (128-s cutoff). Short-term temporal autocorrelations were modeled using an AR(1) process. Motion correction regressors estimated from the realignment procedure were entered as covariates of no interest. Statistical significance was assessed using linear compounds of the regressors in the GLM, generating statistical parametric maps (SPM) of t values across the brain for each subject and contrast of interest. These contrast images were then entered into a second-level random-effects analysis using a one-sample t test against zero.

Anatomical localization was carried out by overlaying the t-maps on a normalized structural image averaged across subjects, and with reference to an anatomical atlas (Duvernoy, 1999). All coordinates are reported in MNI space (Mazziotta et al., 2001). Unless otherwise noted, all statistics are FWE corrected at the cluster level for multiple comparisons at $p<0.05$ with a height threshold of $p<0.001$ (using the cluster level statistics implementation within SPM). Small volume correction in the outcome variance contrast for striatum was performed within a 12mm sphere around the seed voxel coordinates (xyz = -10, 3, 3), which were taken from (Preuschoff et al., 2006).

Region of interest analysis

We extracted data for all region of interest analyses using a cross-validation leave-one-out procedure: we reestimated our main second-level analysis 16 times, always leaving out one subject. Starting at the peak voxel for the correlation signal in right insula and for the correlation prediction error in rACC we selected the nearest maximum in these cross-validation second-level

analyses. Using that new peak voxel we then extracted the data from the left-out subject and averaged across voxels within a 8mm sphere around that peak.

Binned effect size plots

To create the effect size plots of the parametric decision variables we first divided the values in our parametric modulator into quartiles and estimating the average BOLD response in relation to each bin. We did this by sorting all trials into four bins according to the magnitude of the model predicted signal and defined the 25th, 50th, 75th, and 100th percentile of the range. Then we created and estimated for each subject a new GLM with four new onset regressors containing the trials of each bin. The parameter estimates of these onset regressors represent the average height of the BOLD response for all trials in that bin. The data plots in Figs. 2 B and 3 B are the average parameter estimates (across all subjects in the cross-validation analyses) converted to percent signal change. This analysis was performed using algorithms in the rfxplot toolbox for SPM (Glascher, 2009).

Covariance / Correlation comparison

For the test whether bold activity in right insula is better explained by a linear relationship with covariance or correlation we estimated two additional GLMs on BOLD data, each with only one regressor (either model predicted covariance or the correlation coefficient) using Bayesian estimation (Friston et al., 2002). This produced a log-evidence map for each model and we calculated average log evidences across all voxels within our region of interest for every subject and performed a random effects model comparison (Stephan et al., 2009). This analysis suggests that the correlation coefficient can explain BOLD activity in midinsula better than covariance (Dirichlet alpha = 16.9 for correlation vs. 1.1 for covariance; posterior probability (correlation) $p = 0.94$, exceedance probability (probability that the correlation model is more likely) ≈ 1.0).

Effect size time course plots

To visualize the nature of the BOLD response to the correlation coefficient as timecourse plot over the entire trial we upsampled the entire extracted bold signal to 100ms (the effective temporal resolution of the averaged timecourse is higher than the TR because our stimulus presentation was jittered relative to slice acquisition), split the signal into trials and resampled such that the onset of the choice screen is at time 0 and the onset of the outcome screen at 8.5 seconds in every trial. We then estimated a GLM across trials for every timepoint in each subject independently. Lastly, we calculated group average effect sizes at each time point, and their standard errors. The graph in Figure 2C shows the time series of effect sizes throughout the trial for the regressor of interest. This method for plotting the effect size timecourse of a parametrically modulated regressor is also described in detail elsewhere (Behrens et al., 2008).

Timing of correlation representations

To investigate whether subjects carried out task related computations at the time of the outcome or at the time of choice, we estimated a separate GLM that was similar to the main GLM described above except for an additional parametric modulator at the time of choice for the correlation coefficient, i.e. the correlation coefficient modulated both the regressor at the time of the choice screen and the outcome screen.

Representation of portfolio weights

We investigated the questions if subjects might learn task specific portfolio weights instead of the more universal correlation between outcomes by estimating a separate GLM. This was similar to the main GLM except that the parametric modulator ρ was replaced by the portfolio weight w and the correlation prediction error ζ was replaced by the signed weight prediction error ($w_{t+1} - w_t$). The nonlinear relationship between ρ and w allows us to differentiate between representations of correlation and weights on the neural level.

Representation of outcome coincidences

To test for a neural representation of more qualitative coincidences instead of the correlation coefficient with estimated another GLM, similar to the main GLM except that the parametric modulators ρ and ζ were replaced by a binary parametric modulator with a coincidence value of $sign(td_1)*sign(td_2)$.

Relationship between explained variance in behavioral model and BOLD data

To test for a relationship between behavior and neural model fit we compared R^2 (explained variance) in the behavioral model with the R^2 in the fmri GLM. An R^2 value for the behavioral model was calculated for every subject by regressing trial-by-trial model predicted choice on subject's actual choices. We calculated the R^2 value for the fmri regression as the proportion of variance in BOLD that was explained by our interest regressors in relation to the total variance ($R^2 = RSS_{reg} / RSS_{tot}$), where RSS_{reg} equals the explained variance (variance of the predicted timecourse $y^{pred} = Xb$, X = design matrix and b the regression coefficient) and RSS_{tot} is the variance of the bold signal after adjusting for block and nuisance effects.

We also tested the influence of potential confounding variables on this relationship, namely the fitted learning rate and the average absolute amount of weight updating per trial, by calculating partial correlations. This analysis confirmed a significant correlation between behavioral and neural fit (r_{xy}=0.54, p=0.04) after accounting for potential confounds. Furthermore, there was no relationship between these potential confounds and neural fit (r_{ay}=0.12, p=0.66; $r_{|w|y}$=-0.14, p=0.63).

Psychophysiological interaction (PPI) analysis

We performed posthoc an exploratory PPI analysis (Friston et al., 1997) to investigate changes in functional connectivity with right mid insula at the time of outcome (when almost all task related activity was observed). The PPI term was Y x P, with Y being the BOLD timecourses in the insula ROI and P indicating the time during the outcome screen. We then entered the seed region BOLD Y, the psychological variable P, and the PPI interaction term into a new GLM. Findings from this analysis are reported in Figure S4.

Acknowledgements

This study was supported by a Wellcome Trust Program Grant and Max Planck Award.

References

Andersson, J.L., Hutton, C., Ashburner, J., Turner, R., and Friston, K. (2001). Modeling geometric deformations in EPI time series. Neuroimage *13*, 903-919.

Andrade, A., Paradis, A.L., Rouquette, S., and Poline, J.B. (1999). Ambiguous results in functional neuroimaging data analysis due to covariate correlation. Neuroimage *10*, 483-486.

Bechara, A., Damasio, H., Tranel, D., and Damasio, A.R. (1997). Deciding advantageously before knowing the advantageous strategy. Science *275*, 1293-1295.

Becker, G., DeGroot, M., and Marschak, J. (1964). Measuring utility by a single-response sequential method. Behav Sci *9*, 226-232.

Behrens, T.E., Woolrich, M.W., Walton, M.E., and Rushworth, M.F. (2007). Learning the value of information in an uncertain world. Nat Neurosci *10*, 1214-1221.

Behrens, T.E.J., Hunt, L.T., Woolrich, M.W., and Rushworth, M.F.S. (2008). Associative learning of social value. Nature *456*, 245-249.

Benartzi, S., and Thaler, R. (2001). Naive diversification strategies in defined contribution savings plans. American Economic Review *91*, 79-98.

Bossaerts, P. (2010). Risk and risk prediction error signals in anterior insula. Brain Struct Funct *214*, 645-653.

Bunzeck, N., Dayan, P., Dolan, R.J., and Duzel, E. (2010). A common mechanism for adaptive scaling of reward and novelty. Hum Brain Mapp.

Christopoulos, G.I., Tobler, P.N., Bossaerts, P., Dolan, R.J., and Schultz, W. (2009). Neural correlates of value, risk, and risk aversion contributing to decision making under risk. J Neurosci *29*, 12574-12583.

Craig, A.D. (2009). How do you feel--now? The anterior insula and human awareness. Nat Rev Neurosci *10*, 59-70.

Danforth, B.N. (1999). Emergence dynamics and bet hedging in a desert bee, Perdita portalis. Proc Biol Sci. *266*.

Daniel, K., Hirshleifer, D., and Teoh, S.H. (2002). Investor psychology in capital markets: evidence and policy implications. Journal of Monetary Economics *49*, 139–209.

Dodson, L. (1989). Three Dimensions of Diversification. Journal of Accountancy *167*, 131-139.

Doya, K. (2008). Modulators of decision making. Nat Neurosci *11*, 410-416.

Draper, N.R., and Smith, H. (1998). Extra Sums of Squares and Tests for Several Parameters being Zero. In Applied Regression Analysis (Wiley).

Duvernoy, H.M. (1999). The Human Brain. Surface, Blood Supply and Three-Dimensional Section Anatomy, 2nd edn (NewYork: Springer).

Fox, C.W. (2003). Bet-hedging and the evolution of multiple mating. Evolutionary ecology research *5*, 273.

Friston, K.J., Buechel, C., Fink, G.R., Morris, J., Rolls, E., and Dolan, R.J. (1997). Psychophysiological and modulatory interactions in neuroimaging. Neuroimage *6*, 218-229.

Friston, K.J., Penny, W., Phillips, C., Kiebel, S., Hinton, G., and Ashburner, J. (2002). Classical and Bayesian inference in neuroimaging: theory. Neuroimage *16*, 465-483.

Gentle, J.E. (1998). Cholesky Factorization. In Numerical Linear Algebra for Applications in Statistics (Springer), pp. 93-95.

Glascher, J. (2009). Visualization of group inference data in functional neuroimaging. Neuroinformatics *7*, 73-82.

Hampton, A.N., Bossaerts, P., and O'Doherty, J.P. (2006). The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. J Neurosci *26*, 8360-8367.
Hopper, K.R. (1999). Risk-spreading and bet-hedging in insect population biology. Annu Rev Entomol *44*, 535-560.
Huettel, S.A., Stowe, C.J., Gordon, E.M., Warner, B.T., and Platt, M.L. (2006). Neural signatures of economic preferences for risk and ambiguity. Neuron *49*, 765-775.
Jorion, P. (2009). Financial Risk Manager Handbook, 5th edn (Wiley).
Kacelnik, A., and Bateson, M. (1996). Risky Theories--The Effects of Variance on Foraging Decisions. Integrative and Comparative Biology *36*, 402-434.
Kagel, J.H., and Roth, A.E. (1997). Handbook of Experimental Economics (Princeton University Press).
Kennerley, S.W., Walton, M.E., Behrens, T.E., Buckley, M.J., and Rushworth, M.F. (2006). Optimal decision making and the anterior cingulate cortex. Nat Neurosci *9*, 940-947.
Knutson, B., and Cooper, J.C. (2005). Functional magnetic resonance imaging of reward prediction. Curr Opin Neurol *18*, 411-417.
Knutson, B., Taylor, J., Kaufman, M., Peterson, R., and Glover, G. (2005). Distributed Neural Representation of Expected Value. J. Neurosci. *25*, 4806-4812.
Koelsch, S., Fritz, T., DY, V.C., Muller, K., and Friederici, A.D. (2006). Investigating emotion with music: an fMRI study. Human brain mapping *27*, 239-250.
Kuhnen, C.M., and Knutson, B. (2005). The neural basis of financial risk taking. Neuron *47*, 763-770.
Lau, B., and Glimcher, P.W. (2007). Action and outcome encoding in the primate caudate nucleus. J Neurosci *27*, 14502-14514.
Markowitz, H. (1952). Portfolio Selection. The Journal of Finance *7*, 77-91.
Matsumoto, M., Matsumoto, K., Abe, H., and Tanaka, K. (2007). Medial prefrontal cell activity signaling prediction errors of action values. Nat Neurosci *10*, 647-656.
Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J., Zilles, K., Woods, R., Paus, T., Simpson, G., Pike, B., *et al.* (2001). A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). Philos Trans R Soc Lond B Biol Sci *356*, 1293-1322.
Mohr, P.N.C., Biele, G., and Heekeren, H.R. (2010). Neural Processing of Risk. Journal of Neuroscience *30*, 6613-6619.
Montague, P.R., and Berns, G.S. (2002). Neural Economics and the Biological Substrates of Valuation. Neuron *36*, 265-284.
Nagelkerke, N. (1991). A Note on a General Definition of the Coefficient of Determination. Biometrika *78*, 691-692.
O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., and Dolan, R.J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. Science *304*, 452-454.
Padoa-Schioppa, C. (2009). Range-adapting representation of economic value in the orbitofrontal cortex. J Neurosci *29*, 14004-14014.
Padoa-Schioppa, C., and Assad, J.A. (2006). Neurons in the orbitofrontal cortex encode economic value. Nature *441*, 223-226.

Paulus, M.P., Rogalsky, C., Simmons, A., Feinstein, J.S., and Stein, M.B. (2003). Increased activation in the right insula during risk-taking decision making is related to harm avoidance and neuroticism. NeuroImage *19*, 1439-1448.
Plassmann, H., O'Doherty, J., and Rangel, A. (2007). Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. J Neurosci *27*, 9984-9988.
Platel, H., Price, C., Baron, J.C., Wise, R., Lambert, J., Frackowiak, R.S., Lechevalier, B., and Eustache, F. (1997). The structural components of music perception. A functional anatomical study. Brain : a journal of neurology *120 (Pt 2)*, 229-243.
Platt, M.L., and Huettel, S.A. (2008). Risky business: the neuroeconomics of decision making under uncertainty. Nat Neurosci *11*, 398-403.
Preuschoff, K., Bossaerts, P., and Quartz, S.R. (2006). Neural differentiation of expected reward and risk in human subcortical structures. Neuron *51*, 381-390.
Preuschoff, K., Quartz, S.R., and Bossaerts, P. (2008). Human insula activation reflects risk prediction errors as well as risk. J Neurosci *28*, 2745-2752.
Samejima, K., Ueda, Y., Doya, K., and Kimura, M. (2005). Representation of action-specific reward values in the striatum. Science *310*, 1337-1340.
Schultz, W., Dayan, P., and Montague, P.R. (1997). A neural substrate of prediction and reward. Science *275*, 1593-1599.
Schwarz, G.E. (1978). Estimating the dimension of a model. Annals of Statistics *6*, 461-464.
Seymour, B., and McClure, S.M. (2008). Anchors, scales and the relative coding of value in the brain. Curr Opin Neurobiol *18*, 173-178.
Stephan, K.E., Penny, W.D., Daunizeau, J., Moran, R.J., and Friston, K.J. (2009). Bayesian model selection for group studies. Neuroimage *46*, 1004-1017.
Stephens, D.W. (1981). The logic of risk-sensitive foraging preferences. Animal Behaviour *29*, 628-629.
Sutton, R.S., and Barto, A.G. (1998). Reinforcement Learning: An Introduction (MIT Press, Cambridge, MA).
Symmonds, M., Bossaerts, P., and Dolan, R.J. (2010). A behavioral and neural evaluation of prespective decision-making under risk. Journal of Neuroscience *30*.
Thorndike, E.L. (1911). Animal Intelligence: An Experimental Study of the Associate Processes in Animals (New York: Macmillan).
Tobler, P.N., Christopoulos, G.I., O'Doherty, J.P., Dolan, R.J., and Schultz, W. (2009). Risk-dependent reward value signal in human prefrontal cortex. Proc Natl Acad Sci U S A *106*, 7185-7190.
Tversky, A., and Kahneman, D. (1981). The framing of decisions and the psychology of choice. Science *211*, 453-458.
Weiskopf, N., Hutton, C., Josephs, O., and 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.
Wunderlich, K., Rangel, A., and O'Doherty, J.P. (2009). Neural computations underlying action-based decision making in the human brain. Proceedings of the National Academy of Sciences *106*, 17199-17204.
Wunderlich, K., Rangel, A., and O'Doherty, J.P. (2010). Economic choices can be made using only stimulus values. Proc Natl Acad Sci U S A *107*, 15005-15010.
Yoshimura, J., and Clark, C. (1991). Individual adaptations in stochastic environments. Evolutionary Ecology *5*, 173-192-192.

Figure Captions

Figure 1. *Experimental Design.* (**A**) Subjects were presented with a slider to set portfolio weights that determine the fraction of each resource (wind or solar power) in the energy mix (screen 1). The weights could be set within the range [-1,2], with a fixed relationship that both weights always add up to 1, i.e. $w_{wind} = 1\text{-}w_{sun}$. The trial outcome (screen 2) displayed the individual resource values for sun and wind, and the portfolio value of the combined mix (calculated by the weights from screen 1). (**B**) Optimal portfolio weight w_{sun} ($w_{wind} = 1 - w_{sun}$) increases as a function of the correlation coefficient between sun and wind outcomes. The background color indicates portfolio standard deviation (blue = small sd, red = large sd). Optimal portfolio weights (for variance minimization) are displayed as white line, the gray lines indicate the 10% interval around the optimal choice (a deviation of that amount from the optimal weights would result in a 10% higher sd). (**C**) The correlation estimate ρ (red line) is updated from trial to trial (x-axis) via a correlation prediction error ζ (green stems) and then in a second step used to allocate weights in every trial. ζ is calculated as the cross product between the two resource outcome prediction errors (gray bars). The correlation coefficient that was used to generate the data in this illustration is -.60 during the first 10 trials and afterwards changes to +0.80 (dashed line). Learning of ρ from ζ is depicted here for a learning rate of 0.2.

Figure 2. *Model fit and behavior* (**A**) The correlation learning model explained subjects' behavior best. Plotted are the Bayesian Information Criterions, which are corrected for the different levels of complexity in the models (smaller values are better). The r^2 value represents the proportion of behavioral variance explained by each model. (**B**) Regression of actual weights on model predicted weights. Data is pooled over all subjects; for single subject results see Table 1. Note that the deviations at the extremes are a result from bounding the possible weight range

at -1 and 2; any behavioral errors at the boundary could therefore happen only in one direction. (**C**) Both the response of a representative subject (blue) and the model predicted weights (red) approach the normative best response under full knowledge of the generative correlation (black line) with some lag, which results from the time necessary to observe changes in correlation. Subjects responded after a 20-trial long observation-only phase (not shown).

Figure 3. *Neural representation of correlation strength.* (**A**) Neural activity in mid-insula cortex correlated with the trial-by-trial model predicted correlation strength between the two resource values at the presentation of the outcome screen. (**B**) Effect size plots (average percent signal change across subjects). Data plotted separately for trials in which the model predicted correlation strength was low and high in four bins (25 / 50 / 75 / 100 percentile of correlation range, errors bars = SEM). Activity in insula increased linearly with the correlation coefficient (which is, in contrast to the covariance, normalized by the standard deviations of the resources). Data were extracted using a cross-validation (leave-one-out) procedure to ensure independence of data used for localization and effect measure. (**C**) Timecourse plot of effect size for the correlation coefficient regressor. The correlation coefficient is represented at the time of the outcome screen, when new evidence becomes available, but not during the choice period. (**D**) Comparison of explained variance in the behavioral model with the explained variance in the fMRI analysis. Fluctuations in BOLD activity in mid-insula can be particularly well explained within those subjects whose behavior is also well explained by the model ($r = 0.50$, $p=0.03$). Each dot represents one subject and the line is the regression slope.

Figure 4. *Neural representation of correlation prediction errors.* (**A**) Activity in rostral cingulate cortex correlated with the correlation prediction error. (**B**) Effect size plots (similar to Fig. 2 B) for the cluster confirm a linear effect.

Figure 5. *Absolute weight updates*. Activity in ACC/DMPFC and anterior insula correlated, at the time of the outcome screen, with the absolute amount that subjects update the resource allocation weights during the following choice.

Table 1: Model comparison and model fit

Distribution of subjects' individual maximum likelihoods and parameter estimates. Medians of best-fitting parameters for the compared learning algorithms. Parameters were fit to individual subjects across the three scanning blocks.

NLL = model evidence (negative log likelihood, smaller is better); BIC = model evidence corrected for complexity (Bayesian information criterion).

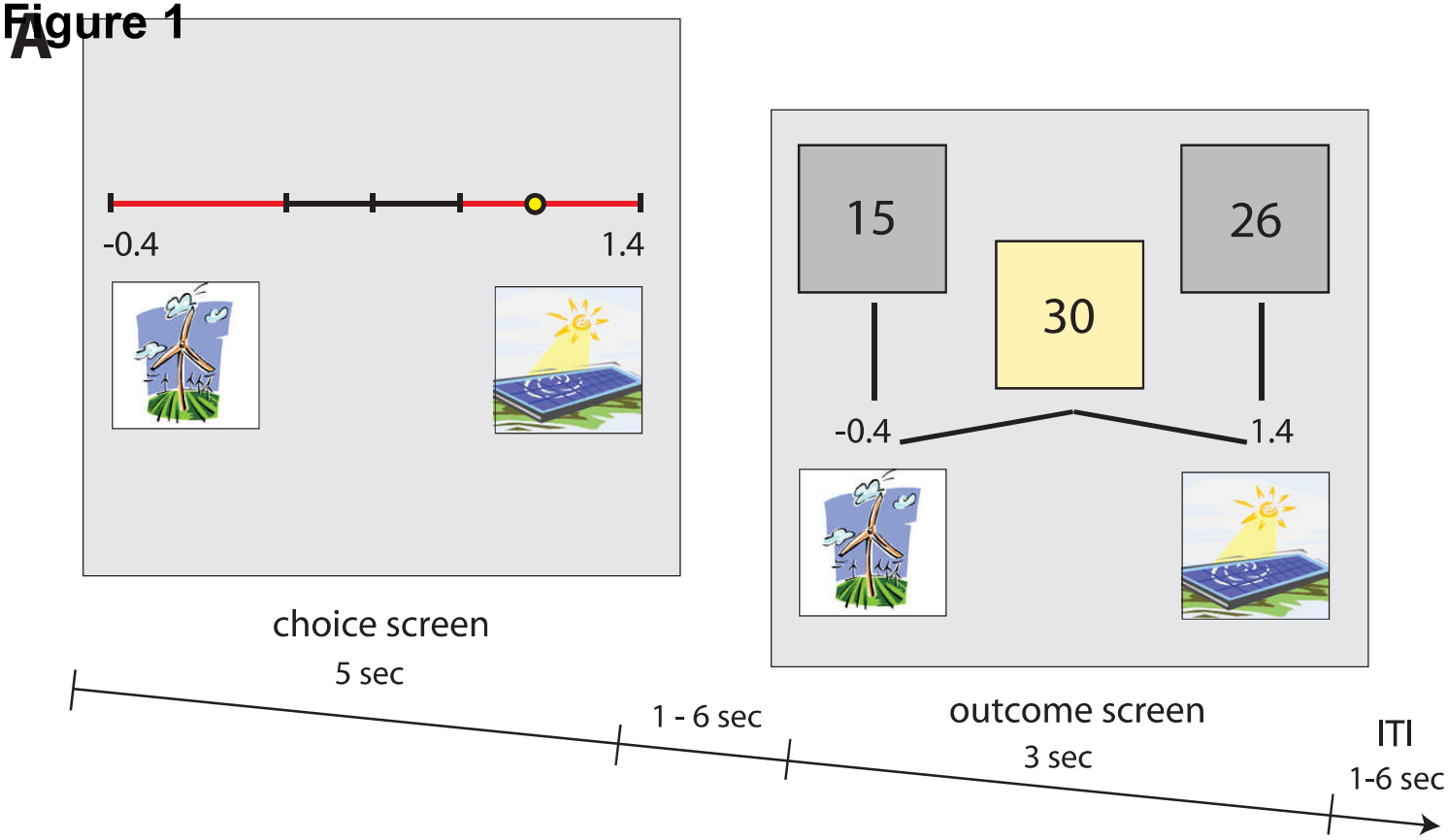
The (pseudo) r^2 value measures how well the model can capture subjects' behavior (see methods). The r^2-forecast measure uses a similar normalization to quantify how well the model could estimate the ground truth correlation. To estimate this value we refit the parameters of each model to estimate ground truth correlations, pooled over all sessions and subjects.

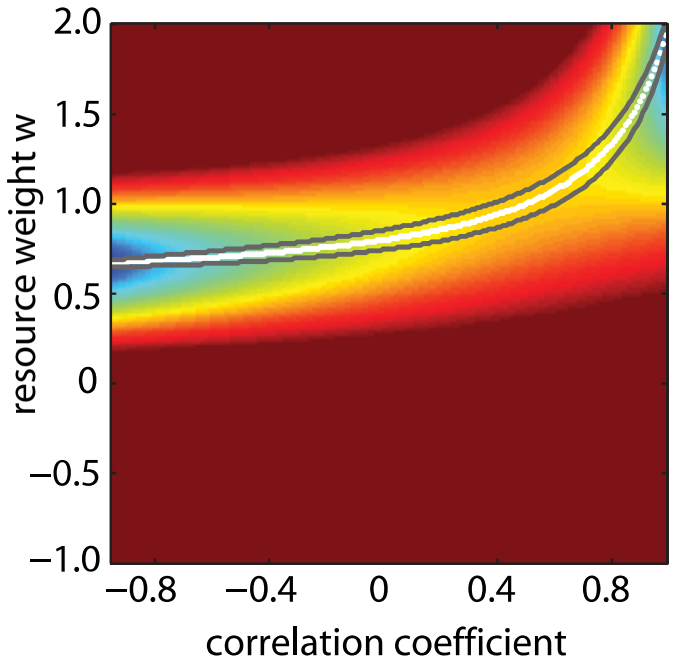
Model\ parameters	Correlation var/cov	Correlation val/var/cov	Q-Learning	coincidence	Sliding window	1/N	Random choice
α–Value	0.08	0.08					
α–Risk	0.25	0.25		0.16			
α–Covariance	0.25	0.26		0.34			
Learning rate			0.23				
w step width			0.10				
Window length					9.93		
n parameters	2	3	2	2	1	0	0
NLL	87.92	86.20	197.20	142.93	110.96	283.85	384.67
r^2	.77	.78	.49	.61	.71	.26	0
BIC	186.90	188.99	405.45	296.92	227.44	567.69	769.35
r^2 forecast	.77	.77	.40	.56	.70	.26	0

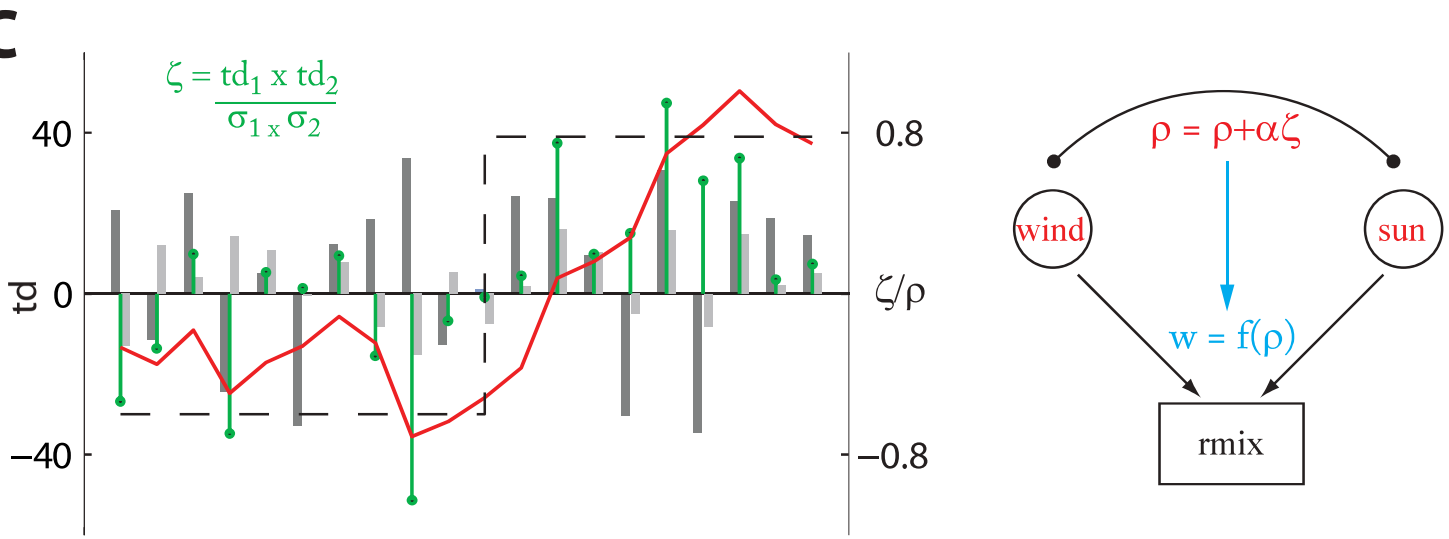
Table 2: Significant activations in statistical parametric analysis.
All peaks are thresholded p<0.001 uncorrected; listed are all clusters with an extent ≥ 3 voxels. * significant at p<0.05 FWE cluster level correction in entire brain. Coordinates in MNI space.

x	*y*	*z*	*Z*	*voxels*	*p (FWE)*	*Region*	*Hemi*
Correlation coefficient (ρ)							
48	**5**	**-5**	**4.12**	**59**	**0.001***	**mid insula**	**R**
60	-1	-5	3.87	"	"	mid insula (extending into superior temporal sulcus)	"
48	-7	-2	3.77	"	"	"	"
-60	-1	-17	3.85	18	0.61	superior temporal sulcus	L
-51	-10	-17	3.20	"	"	"	"
-18	-16	1	3.56	19	0.44	thalamus	L
9	-55	37	4.50	8	0.96	precuneus	R
12	-61	-5	3.33	5	0.90	occipital cortex	L
-54	-40	4	3.31	4	0.96	superior temporal sulcus	L
Correlation prediction error (ζ)							
-15	**44**	**7**	**4.87**	**36**	**0.003***	**rostral ACC**	**L**
-54	-25	-5	4.01	43	0.14	superior temporal sulcus	L
-57	8	-23	3.95	4	0.99	anterior superior temporal sulcus	L
-42	-61	37	3.63	17	0.80	inferior parietal lobe	L
-60	-1	-14	3.61	10	0.93	superior temporal sulcus	L
-63	-7	-8	3.48	"	"	"	"
12	-13	52	3.57	3	0.91	medial cingulate gyrus	R
36	-10	7	3.23	3	0.99	posterior insula	R
Absolute weight update							
6	**26**	**34**	**4.22**	**135**	**0.001***	**ACC / DMPFC**	**R**
-9	29	25	3.50	"	"	"	L
42	**23**	**-5**	**4.04**	**55**	**0.04***	**anterior insula**	**R**
15	**-64**	**34**	**3.95**	**40**	**0.04***	**precuneus**	**R**
51	26	22	3.81	7	0.73	DLPFC	R
15	-31	26	3.73	15	0.38	cerebellum	R
0	-19	-2	3.71	29	0.20	VTA vicinity	
-33	17	-5	3.57	21	0.69	anterior insula	L
-12	2	58	3.37	7	0.88	SMA	L
0	-52	-35	3.18	6	0.97	cerebellum	
Risk (average contrast over individual risk from both outcomes, h1/h2)							
45	-4	-14	3.69	7	0.76	posterior insula	R
45	-76	34	3.67	3	0.98	posterior parietal cortex	R
18	-28	4	3.60	3	0.86	thalamus	R
-21	2	7	3.55	7	0.85	striatum	L
-42	-55	-35	3.38	3	0.99	cerebellum	L
Risk prediction errors (average contrast over individual risk PE from both outcomes, $\varepsilon_1/\varepsilon_2$)							
-24	23	-8	3.13	3	0.98	anterior insula	L

Figure 1

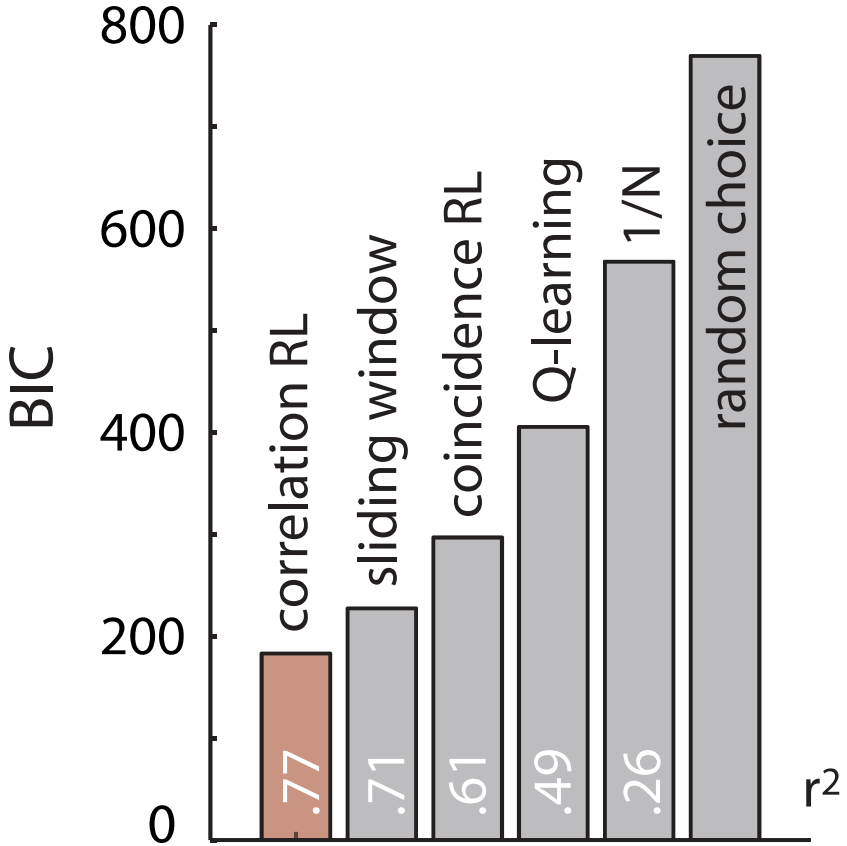

A
-0.4
1.4
15
26
30
-0.4
1.4
choice screen
5 sec
1 - 6 sec
outcome screen
3 sec
ITI
1-6 sec

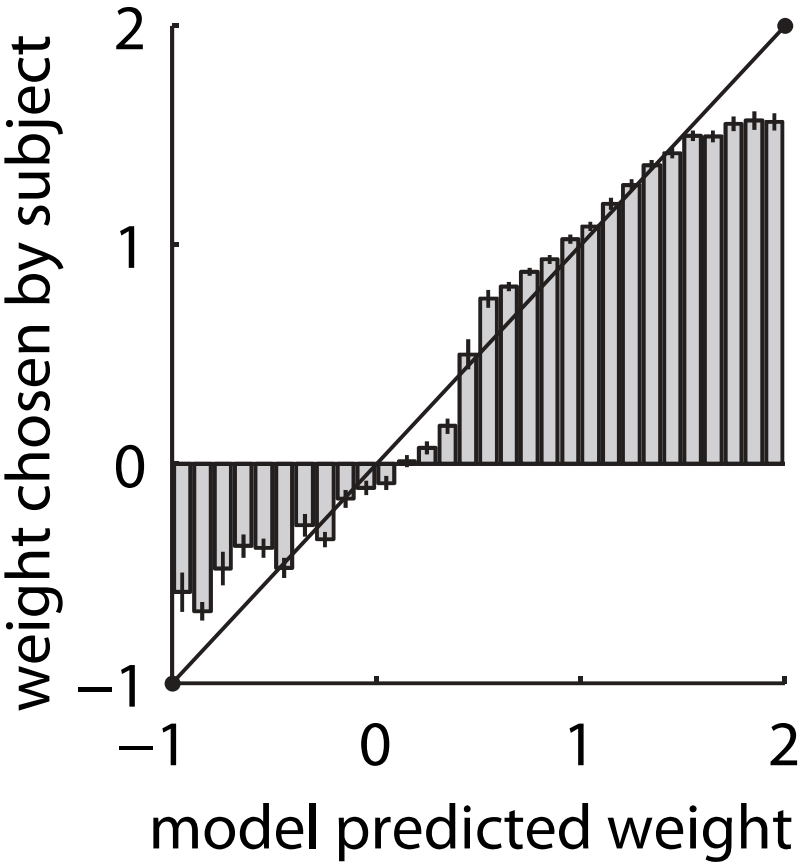

B
2.0
1.5
1.0
0.5
0
−0.5
−1.0
resource weight w
−0.8
−0.4
0
0.4
0.8
correlation coefficient


C
ζ = td1 x td2 / σ1 x σ2
40
0
−40
td
0.8
ζ/ρ
−0.8

ρ = ρ+αζ
wind
sun
w = f(ρ)
rmix

Figure 2


800
600
400
200
0
BIC
correlation RL
sliding window
coincidence RL
Q-learning
1/N
random choice
.77
.71
.61
.49
.26
r2


B
2
1
0
−1
−1
0
1
2
weight chosen by subject
model predicted weight

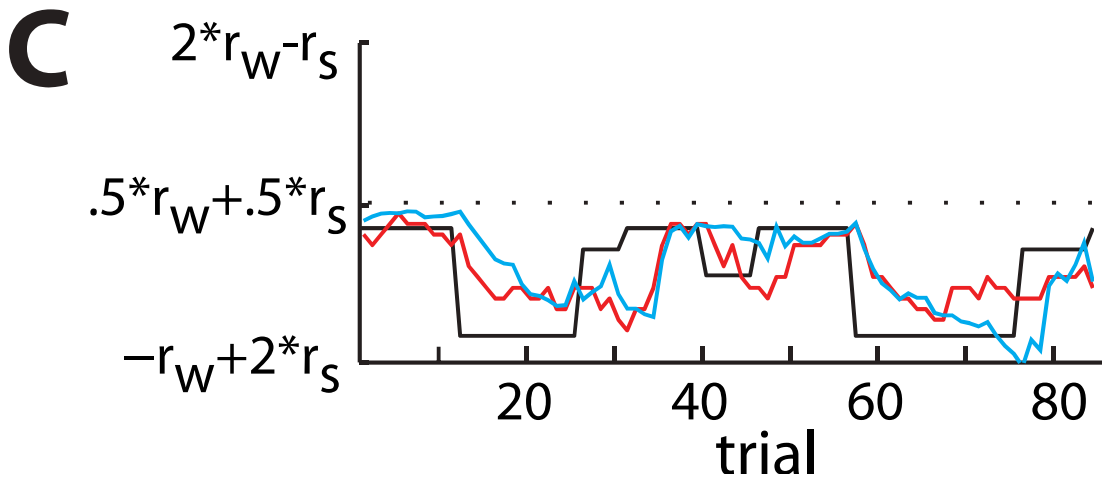

C
2*rw-rs
.5*rw+.5*rs
−rw+2*rs
20
40
60
80
trial

Figure 3

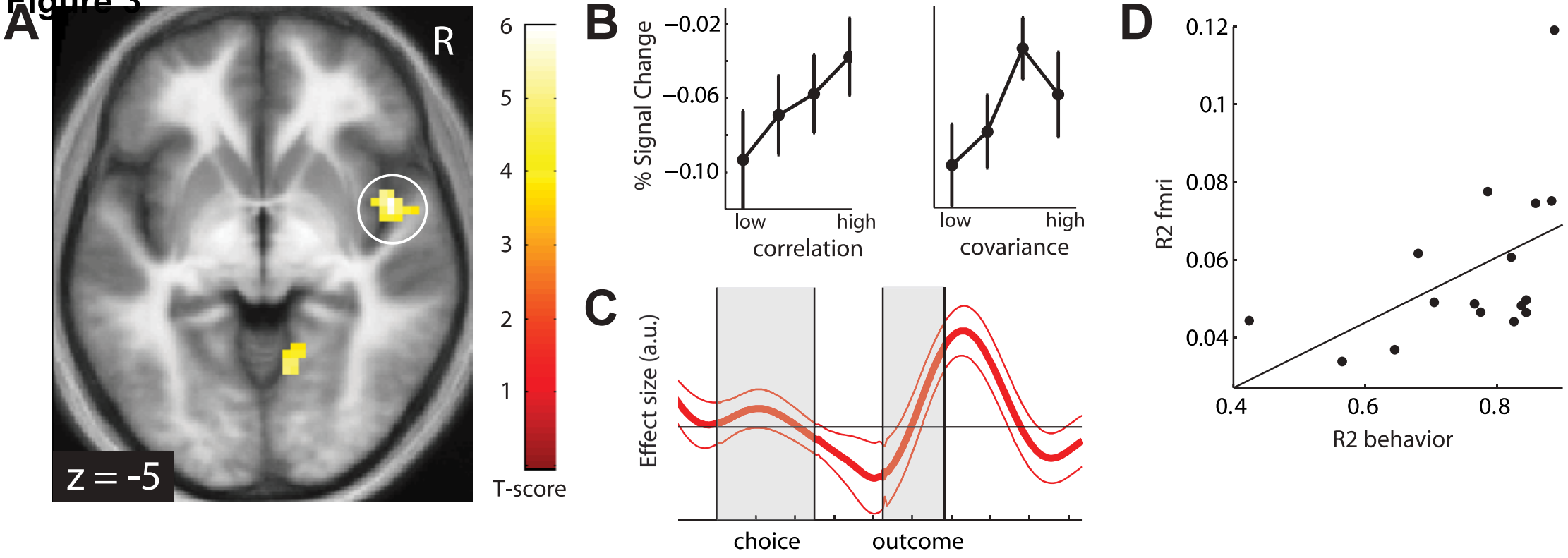
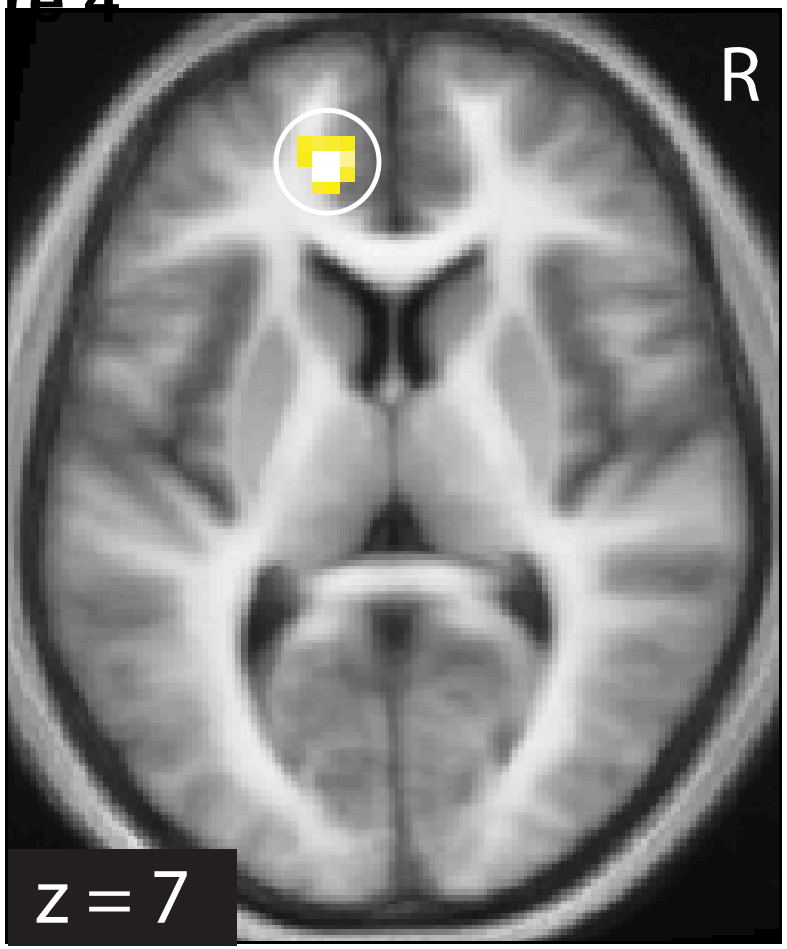

A
R
z = -5
6
5
4
3
2
1
T-score
B
% Signal Change
−0.02
−0.06
−0.10
low
high
correlation
low
high
covariance
C
Effect size (a.u.)
choice
outcome
D
R2 fmri
0.12
0.1
0.08
0.06
0.04
0.4
0.6
0.8
R2 behavior

Figure 4



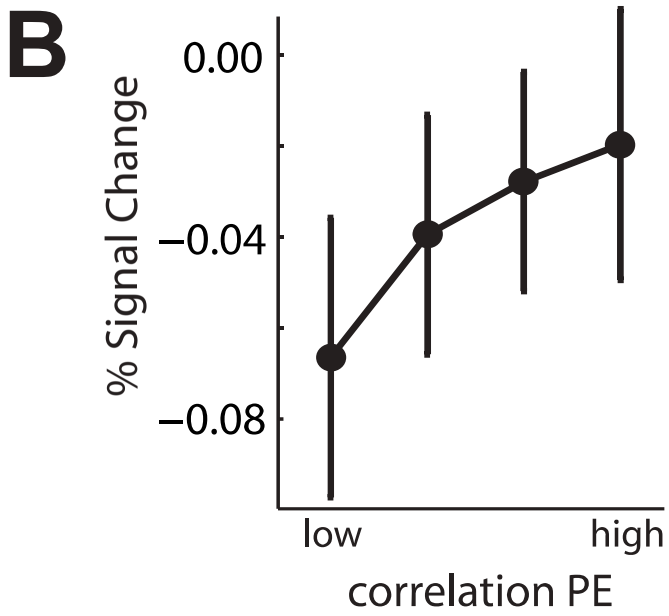


Figure 5