Title

Asymmetric reinforcement learning: computational and neural bases of positive life orientation

Running title:
Computational and neural bases of positive life orientation

Authors:
Germain Lefebvre[1,2], Maël Lebreton[3,4], Florent Meyniel[5], Sacha Bourgeois-Gironde[2,6] & Stefano Palminteri[1,7]

Affiliations:
[1]Laboratoire de Neurosciences Cognitives (LNC), INSERM U960, Ecole Normale Supérieure, 75005, Paris, France.
[2]Laboratoire d'Économie Mathématique et de Microéconomie Appliquée (LEMMA), Université Panthéon-Assas, 75006, Paris, France
[3]Amsterdam Brain and Cognition (ABC), Nieuwe Achtergracht 129, 1018 WS Amsterdam, the Netherlands
[4]Amsterdam School of Economics (ASE), Faculty of Economics and Business (FEB), Roetersstraat 11, 1018 WB Amsterdam, the Netherlands
[5]Cognitive Neuroimaging Unit, CEA, INSERM, Université Paris-Sud, Université, Paris-Saclay, NeuroSpin center, 91191 Gif/Yvette, France
[6]Institut Jean-Nicod (IJN), CNRS UMR 8129; Ecole Normale Supérieure, 75005, Paris, France.
[7]Institute of Cognitive Neurosciences (ICN), University College London, WC1N 3AR London, UK.

Corresponding author:
Stefano Palminteri (stefano.palminteri@gmail.com)

Abstract
While forming and updating beliefs about future life outcomes, people tend to consider good news and to disregard bad news. This tendency is supposed to support the optimism bias. Whether learning bias is specific to "high-level" abstract belief update or a particular expression of a more general "low-level" reinforcement learning process is unknown. Here we report evidence in favor of the second hypothesis. In a simple instrumental learning task, participants incorporated worse-than-expected outcomes at a lower rate compared to better-than-expected ones. This asymmetry was correlated across subjects with standard measure of dispositional optimism. Multimodal imaging indicated that inter-individual variability in the expression of asymmetric update relies on the dorsomedial prefrontal cortex at both morphological and functional levels. Our results constitute a new step in the understanding of the genesis of optimism bias at the neurocomputational level.

Introduction

"*It is the peculiar and perpetual error of the human understanding to be more moved and excited by affirmatives than negatives; whereas it ought properly to hold itself indifferently disposed towards both alike*" (p. 36)[a].

People typically underestimate the likelihood of negative events and overestimate the likelihood of positive events. This cognitive trait in (healthy) humans is known as the optimism bias and has been repeatedly evidenced in many different guises and populations (Shepperd et al., 2013, 2015; Weinstein, 1980) such as students projecting their salary after graduation (Shepperd et al., 1996), women estimating their risk of getting breast cancer (Waters et al., 2011) or heavy smokers assessing their risk of premature mortality (Schoenbaum, 1997). One mechanism hypothesized to underlie this phenomenon is an asymmetry in belief updating, colloquially referred to as "the good news / bad news effect" (Eil and Rao, 2011; Sharot et al., 2011). Indeed, preferentially revising one's beliefs when provided with favorable compared to unfavorable information constitutes a learning bias which could, in principle, generates and sustains an overestimation of the likelihood of desired events and a concomitant underestimation of the likelihood of undesired events (optimism bias) (Sharot and Garrett, 2016).

This good news/bad news effect has recently been demonstrated in the case where outcomes are hypothetical future prospects associated with a strong a priori desirability or undesirability (estimation of post-graduation salary or the probability of getting cancer) (Shepperd et al., 1996; Waters et al., 2011). In this experimental context, belief formation triggers complex interactions between episodic, affective and executive cognitive functions (Eil and Rao, 2011; Sharot et al., 2011, 2007), and belief updating relies on a learning process involving abstract probabilistic information (Garrett et al., 2014; Moutsiana et al., 2015, 2013; Sharot et al., 2011). However, it remains unclear whether this learning asymmetry also applies to immediate reinforcement events driving instrumental learning directed to affectively neutral options (i.e. with no a priori desirability or undesirability). If an asymmetric update is also found in a task involving neutral items and direct feedbacks, then the good news/bad news effect could be considered as a specific – cognitive – manifestation of a general reinforcement learning asymmetry. If the asymmetry is not found at the basic reinforcement learning level, this would mean that the asymmetry is specific of abstract belief updating, and this would require a theory explaining this discrepancy.

To arbitrate between these two alternative possibilities, we analyzed instrumental behavior of subjects performing a simple two-armed bandit task, involving neutral stimuli and actual and immediate monetary outcomes, with two learning models. The first model (a standard RL algorithm) confounded individual learning rates for positive and negative feedbacks and the second one differentiated them, potentially accounting for learning asymmetries.

Over two experiments, we found subjects' behavior was better explained by the asymmetric model, with an overall difference in learning rates consistent with preferential learning from positive, compared to negative, prediction errors. Furthermore, this tendency of using the asymmetric model covaried with optimistic trait as measured with standard psychometric scale (LOT-R). The good news/bad news effect can then be considered as a particular consequence of a more general asymmetry in reinforcement learning, accounting for the optimism bias.

Our and previous studies suggest that the good news/bad news effect is highly variable across subjects. Behavioral differences in optimistic beliefs and optimistic update have been shown to be reflected by differences in brain activation in the prefrontal cortex (Sharot et al., 2011). However the question remains, whether or not such interindividual differences are related to specific anatomical, supposedly developmental, variability as revealed by voxel-based morphometry (VBM). Our imaging results indicate that the inter-individual variability in the tendency in optimistic learning correlates with grey matter density in the dorsal prefrontal cortices. Finally, our fMRI results confirmed at the functional level the assumptions of our computational model as well as the VBM results.

Results

Behavioral task and dependent variables

Healthy subjects performed a probabilistic instrumental learning task with monetary feedbacks, previously used in brain imaging, pharmacological and clinical studies[13–15] (**Figure 1A**). In this task, options (abstract cues) were presented in fixed pairs (i.e.

[a]Bacon, F. (1939). *Novum organum*. In Burtt, E. A. (Ed.), The English philosophers from Bacon to Mill (pp. 24-123). New York: Random House. (Original work published in 1620)
[b]Original French citation: "Qu'est-ce qu'optimisme? disait Cacambo. – Hélas! dit Candide, c'est la rage de soutenir que

conditions). In all conditions each cue was associated with a stationary probability of reward. In asymmetric conditions, the two reward probabilities differed between cues (25/75%). From asymmetric conditions we extracted the rate of "correct" response (selection of the best option) as a measure of performance (**Figure 1B, left**). In symmetric conditions, both cues had the same reward probabilities (25/25% or 75/75%), such that there was no intrinsic "correct response". In symmetric conditions we extracted for each subject and each symmetric pair, a "preferred response" rate, defined as the choice rate of the option most frequently selected by a given subject (i.e.by definition in more than 50% of trials). The preferred response rate in the 25/25% conditions should be taken as a measure of the tendency to overestimate the value of one instrumental cue compared to the other, in absence of actual outcome-based evidence. (**Figure 1B, right**). In a first experiment (N=50) that subjects performed while being fMRI scanned, the task involved reward (+0.5€) and reward omission (0.0€), as the best and worst outcome respectively. In a second purely behavioral experiment (N=35), the task involved reward (+0.5€) and punishment (-0.5€), as the best and worst outcome respectively. In addition to performing the instrumental learning task, subjects of this second experiment filled the Life Orientation Test – Revised (LOT-R), a standard assessment of optimistic trait.

Computational models

We fitted the behavioral data with two reinforcement-learning models (Sutton and Barto, 1998). The "reference" model was represented by a standard Rescorla-Wagner (Rescorla and Wagner, 1972), thereafter referred to as RW model. The RW model learns option values by minimizing reward perdition errors. It uses a single learning rate (alpha: a) to learn from positive and negative prediction errors. The "target" model was represented by a modified version of the RW model, thereafter referred to as RW± model. In the RW± model, learning from positive and negative prediction errors is governed by different learning rates (alpha: a^+ and alpha: a^- respectively). For $a^+ > a^-$ the RW± model instantiates optimistic reinforcement learning (i.e. the good news/bad news effect); for $a^+ = a^-$, the RW± instantiates realistic (or unbiased) reinforcement learning, just as in the RW model (the RW model is thus nested in the RW± model); finally for $a^+ < a^-$ the RW± instantiates pessimistic reinforcement learning. In both models the choices are taken by feeding the option values into a softmax decision rule, whose exploration/exploitation trade-off is governed by a parameter (b).

Model comparison and model parameters analysis

We implemented Bayesian model comparison to establish which model better accounted for the behavioral data. For each model we estimated the optimal free parameters by maximization the likelihood of the participants' choices, given the models and sets of parameters. For each model and each subject, we calculated the Bayesian Information Criterion (BIC) by penalizing the maximum likelihood with the number of free parameters in the model. Group-level BIC analysis indicated that the RW± model better explains the behavioral data compared to the RW model (BIC_{RW}=99.4±4.4, $BIC_{RW\pm}$=93.6±4.7; t(49)= 2.9, p=0.006, paired t-test), even accounting for its additional degrees of freedom (**Table 1**). RW± being the best fitting model we compared the learning rates fitted for positive (good news: a^+) and negative (bad news: a^-) prediction errors. We found a^+ significantly higher compared to a^- (t(49)= 3.8, p<0.001 paired t-test). To summarize, model comparison indicated that, in our simple instrumental learning task, the best fitting model is the model with different learning rates for learning from positive and negative predictions errors (RW±). Crucially, learning rates comparison indicated that instrumental values are preferentially updated following positive prediction errors, which is consistent with an optimistic bias operating when learning from immediate feedback (optimistic reinforcement learning).

Computational characterization of inter-individual variability

To explore inter-individual variability, we computed for each subject the between-model BIC difference ($DBIC=BIC_{RW} - BIC_{RW\pm}$). The ΔBIC quantifies at the individual level the goodness of fit improvement moving from the RW to the RW± models, hence approximates a measure of optimism. Subjects with a positive DBIC (N=25, in the first experiment) are subjects whose behavior is better explained by the RW± model (thereafter refereed as RW± subjects); subjects with a negative DBIC (N=25, in the first experiment) are subjects whose behavior is better explained by the RW model (thereafter refereed as RW subjects) (**Figure 2A**). Importantly the maximum likelihood of reference model (RW) was not different between the two groups of subjects, indicating similar baseline quality of fit (t(48)= -1.0, p=0.314, two-sample t-test)..

To characterize the computational profiles of these two groups of subjects and the validity of the ΔBIC as a measure of optimism, we analyzed and compared their free parameters (α^+ and α^- and 1/β) (figure 2C). Learning rates fitted with the RW± model entered a two-way ANOVA with group (RW and RW±) and learning rates type (a^+ and a^-) as respectively between- and within-subjects factors. The ANOVA showed a main effect of learning rate type (F(1,48)=16.5, P<0.001) with a^+ higher than a^-. We also found a main effect of group (F(1,48)=10.48, P=0.002) and a significant group x learning type interaction (F(1,48)= 7.8, P=0.007). Post-hoc tests revealed that average learning rates for positive prediction errors were not different among the two groups, a^+_{RW}=0.45 ± 0.08 and $a^+_{RW\pm}$= 0.27 ± 0.06 (t(48) = 1.7, p=0.086, two-sample t-test). On the contrary, average learning rates for negative prediction errors were significantly different between groups a^-_{RW}= 0.41 ± 0.08 and $a^-_{RW\pm}$= 0.04 ± 0.02 (t(48)= 4.6, p<0.001, two-sample t-test). In addition, an asymmetry in learning rates was detected within the RW± group, where a^+ was higher than a^- (t(24)=5.1, p<0.001, paired t-test) but not within RW group (t(24)=0.9, p=0.399, paired t-test). Thus, RW± subjects specifically drove the learning rates asymmetry found in whole population. On the other side the RW subject display "realistic" (as opposed to "optimistic") instrumental learning (**Figure 2B and 2C**).

Interestingly, the choice randomness (captured by the 1/β, "temperature" parameter) was also found to be significantly different between the two groups of subject, $1/\beta_{RW}$=0.20 ± 0.05 and $1/\beta_{RW\pm}$=0.06 ± 0.01. (t(48)= 2.9, p=0.006, two-sample t-test). This suggests that optimistic reinforcement learning, observed in RW± subjects, is also associated with exploitative, as opposite to explorative, behavior (**Figure 2C**). Finally, the ΔBIC (our classification variable) was found significantly correlated significantly with the normalized learning rate asymmetry ($[a^+ - a^-]/[a^+ + a^-]$; R=0.5455, p<0.001) and with the temperature (R=-0.3343, p=0.0177). To summarize, RW± subjects tend to weight more positive feedbacks and, as a consequence, to exploit more consistently the previously rewarded options (optimism). Both computational features of this optimistic computational phenotype are quantitatively captured by the ΔBIC index.

Behavioral signature distinguishing optimistic from realistic subjects

In order to analyze the behavioral consequences of optimistic, as opposed to realistic, learning and to confirm our model-based results with model-free behavioral observations, we compared the task's dependent variables between our two groups of subjects (**Figure 2D, Table 2**). Correct response rate did not differ between groups (t(48)=-0.7323, p=0.467, two-sample t-tests). However, the preferred response rate in the 25/25% condition was significantly higher for RW± group in comparison to RW group (t(48)= -3.4, p=0.001, two-sample t-test).

In order to validate the ability of RW± model to capture this difference, we performed simulations using both models and submitted them to the same statistical analysis as actual choices (**Figure 2D**). The RW± model simulated preferred response rate was significantly higher for RW± group compared to the RW group (t(48)= -5.4496, p<0.001, two-sample t-test), which replicated human behavior. However, the simulated preferred response rates from the RW model were similar in the two groups (t(48)= -0.6, p=0.566, two-sample t-test), which departed from our observations in real subjects. We found that a higher preferred response rate in the 25/25% condition was a specific signature of optimistic learning: in poorly rewarding environment (and where there is no intrinsic correct response), optimistic subjects tend to overestimate the value of one of the two options (**Supplementary Figure 1**). Finally, we found that the preferred response rate in the 25/25% condition was significantly correlated with our classification computational variable ΔBIC (R=0.6366, p<0.001). The preferred response rate thus provides a model-free signature of optimistic reinforcement learning that is congruent with our computational analysis: the preferred response

rate was higher in RW± group in comparison to RW group and only simulations realized with RW± model were able to replicate this pattern of responses.

Anatomical signature distinguishing optimistic from realistic subjects

To investigate the neuroanatomical basis of the inter-individual computational variability observed in our task, we used voxel-based morphometry (VBM). We devised a multiple regression model with our computational classification variable ΔBIC as a continuous variable of interest. We used the ΔBIC since it captures two the computational features distinguishing the RW± from the RW subjects: the learning rate asymmetry and the higher tendency to explore. This analysis showed a significant positive correlation between theΔBIC and grey matter density in only two brain regions; the dorsolateral and dorsomedial prefrontal cortices (DLPFC and the DMPFC; **Figure 3A** and **B**). This result implies that the better the behavioral fit by the RW± (optimistic learning), the higher is the grey matter density in DLPFC and DMPFC. Moreover, in order to validate this model-based result with a model-free analysis, we also analyzed the correlation between the grey matter density in these regions and preferred response rate in the 25/25% condition. We found significant correlations in both the DMPFC and the DLPFC (respectively R=0.45, p<0.001, and R=0.42, p=0.003; **Figure 3C** and **D**). Our VBM then indicates that the computational variability observed in our task is linked to structural difference in the dorsal prefrontal cortex (both in the medial and the lateral part). More precisely, the more a subject's computational strategy diverges from an unbiased - "realistic" - model, the higher is the grey matter density in these areas. These results lend further support, based on biological data, to the computational phenotyping based on the learning task.

Functional signature distinguishing optimistic from realistic subjects

To investigate the functional consequence of the anatomical difference between the RW± and RW subjects (i.e. different grey matter density in in DLPFC/DMPFC), we analyzed the brain activity in these regions, using functional Magnetic Resonance Imagining (fMRI). We devised a general linear model in which we modeled as separated events the choice and the outcome onset, each modulated by different parametric modulators. The choice onset was modulated by the reaction time (RT), which counts as a proxy of decision value and complexity and reflects the decision process(Kolling et al., 2012; Shenhav et al., 2014). The outcome onset was modulated by the outcome obtained (0.5€ or 0.0€) in a given trial. Overall, we found both the DMPFC and the DLPFC as significantly encoding reaction times (t(49)=5.9751, P<0.001 and t(49)=2.0873, P=0.042 respectively, one-sample t-tests). However regression coefficients were not different between the two groups in either region (t(48)=-0.0253, P=0.9799 and t(48)=0.1658, P=0.869 respectively, two-sample t-tests). On the contrary, when all subjects were analyzed together, the activity in the DMPFC and DLPFC did not correlate with the outcome value (t(49)=0.5967, P=0.5535 and t(49)=0.648, P=0.52 respectively, one-sample t-tests). We tested whether the two computational phenotypes differed in the way their DMPFC and DLPFC react to the outcome value. We found a significant difference in the DMPFC (t(48)=-2.4919, p=0.016, two-sample t-test); post-hoc comparison showed that the DMPFC activity covaried positively with the outcome value in the RW± group (t(24)=2.1326, P=0.043, one-sample t-test), but not in the RW group (t(24)=-1.3669, P=0.1843, one-sample t-test). We found a similar trend in DLPFC, but the effect did not reach significance (**Figure 4E**). It therefore seems that the difference in grey matter density identified with VBM between the RW and the RW± groups has a functional counterpart. More precisely, the two phenotypes differ in that only in the RW± the DMPFC responds to the outcome value. On the other side, decision process related activations (RT) were found to be similar in the two groups.

Optimistic life attitude correlates with optimistic reinforcement learning

In order to validate our "low level" (computational and instrumental) measures of optimistic reinforcement learning in respect to classical "high level" (psychological and attitudinal) measure of optimism, we ran an additional experiment. In this second behavioral experiment (N=35), subjects completed the Life Orientation Test revised (LOT-R) assessing optimism trait, in addition to our task. Importantly the computational and behavioral results obtained in this second experiment fully replicated the behavioral and computational findings reported above (see **Supplementary Materials** and **Supplementary FigureS2**). We derived from the LOT-R scale the optimism score that we found correlated with the ΔBIC (our model-based index of optimistic learning: R=0.3814, p=0.024) (**Figure 4A** and **C**). We also found a positive and significant correlation between the LOT-R and the preferred response rate in the 25/25% condition (our model-free signature of optimistic behavior: R=0.3350, p=0.049). Thus, the more the subjects are optimist according the "high level" standard psychological measure of optimism, the more they show "low level" computational and behavioral signatures of the RW± phenotype.

Discussion

We found that, in a simple instrumental learning task involving neutral visual stimuli associated to actual monetary rewards, participants preferentially updated option values following better-than-expected, compared to worse-than-expected, outcomes. This learning asymmetry was replicated in two experiments and proved to be robust across different conditions (see **Supplementary Materials**).

At the individual level, the learning asymmetry was differentially expressed across subjects. We were able to capture this variability with a computational measure (the ΔBIC), which quantifies the extent to which a subject's fit improves when moving from the unbiased "realistic" learning model to a biased one. Importantly, this computational metric was strongly related to a (model-free) behavioral signature of optimistic reinforcement learning: the preferred response rate in the poorly rewarding condition.

We further tested the validity of our computational and behavioral assessment of inter-individual variability by confronting it to external (i.e. task-independent) neurobiological and psychometric measures. First, quantitative neuroanatomical imaging indicated that grey matter density in both the DMFPC and the DLPFC positively correlates with both computational and behavioral measures of optimistic reinforcement learning. This relation was found to have specific functional consequences in the DMPFC, in the form of differential neural response to outcome value in realistic compared to optimistic subjects. Finally, computational and behavioral measures of optimistic learning also correlated with a standard measure of dispositional optimism (LOT-R), providing a link between computational and personality traits.

Our results support the hypothesis that the good news/bad news stands as a core psychological process generating and maintaining unrealistic optimism (Eil and Rao, 2011). In addition, our study has the originality of showing that this effect is not specific to probabilistic belief updating, and that the good news/bad news effect can parsimoniously be considered as an amplification of a primary instrumental learning asymmetry. In other terms, following nomenclature recently proposed by Sharot and Garrett, we found that asymmetric update applies to both "estimation errors" and "prediction errors" (Sharot and Garrett, 2016).

The asymmetric model (RW±) included two different learning rates following positive and negative prediction errors and we found the "positive" learning rate higher compared to the "negative" one (Doll et al., 2011; Niv et al., 2015). Distinct learning rates should be taken as supplementary evidence that the systems used to learn from positive and negative prediction errors are dissociable across different brain circuits and areas (Frank et al., 2004; Palminteri et al., 2012). Note that since our first experiment did not implicate losses, but only reward omissions, our results cannot be interpreted as a consequence of loss aversion (Kahneman and Tversky, 1979). In other terms the learning asymmetry is not outcome sign-based, but prediction error sign-based.

However, higher learning rates for positive compared to negative prediction errors was not the only computational metric distinguishing optimistic from realistic subjects. In fact, we also found that optimistic subjects had a greater tendency to exploit previously rewarded option, as opposed to realistic subjects who were more prone to explore both options. Importantly the higher stochasticity of realistic subjects was associated neither with lower performance in the asymmetrical conditions, nor with a lower baseline quality of fit, as measured by the maximum likelihood. This overexploitation tendency was particularly striking in the symmetrical 25/25% condition, in which

both options are poorly rewarding compared to the average task reward rate

We found an interesting association between optimistic update and lower propensity to explore. Indeed, the tendency to ignore negative feedback about chosen options was linked to considering the preferred option better than it is, and hence to stick to this preference. A natural link between optimism and such "conservatism" is not new; it can be dated back to Voltaire's work "*Candide ou l'Optimisme*", where the belief of "living in the best of the possible worlds" was consistently associated with a strong rejection and condemnation of progress and explorative behavior. In the word's of the 18th century philosopher:

"Optimism," said Cacambo, "What is that?" "Alas!" replied Candide, "It is the obstinacy of maintaining that everything is best when it is worst"[b].

Accordingly, optimism bias has been recently recognized as an important psychological factor helping maintain inaction regarding pressing social problems, such as climate changes (Gifford, 2011).

Recent studies investigated the neural implementation of the good news/bad news effect when analyzed in the context of probabilistic belief updating. Decreased belief updating after worst-than-expected information has been associated with a diminished correlation between negative prediction errors and BOLD signal in the right inferior prefrontal gyrus (IFG) (Sharot et al., 2011) and with a stronger white matter connectivity in a system including left IFG and left subcortical regions (Moutsiana et al., 2015). Congruent with those studies indicating prefrontal cortex implication in the good news/bad news effect, we found that the tendency to implement optimistic reinforcement learning was associated with a greater grey matter density in dorso-medial prefrontal cortex (DMPFC) and dorso-lateral prefrontal cortex (DLPFC). Our VBM results therefore suggest that inter-individual anatomical differences – either genetically programmed or induced by interaction with environment during development (Moutsiana et al., 2013)- in the aforementioned region could determine the way people learn from good and bad news. Importantly, as opposed to task-elicited fMRI activations, where inter-individual activation differences could be simply derived from differences in the behavioral variable, structural measures are not affected by performing the task and could therefore provide insight into stable neural markers of a trait of interest (Kanai and Rees, 2011).

Structural imaging leaves the question of the functional consequence of inter-individual differences unanswered. To fill this gap, we analyzed fMRI activity within the regions that have been shown to discriminate between optimists and realists at the anatomical level. We analyzed both decision (reaction time) and learning (outcome encoding) related activations. The results confirmed the implication of the DMPFC in the decision process (Shenhav et al., 2013). However, decision-related activity was not different between optimists and realists. On the other side, outcome-related activity in optimistic subjects was found significantly different from zero and significantly higher compared to realistic subjects. These results suggest that, firstly, anatomical differences have functional specific effects and, secondly, that optimistic update (i.e. the good news/bad news effect) is associated with increased outcome representation in the DMPFC (Ullsperger et al., 2014).

We defined subjects as optimists or pessimists, based on their computational phenotype that we extracted from an instrumental learning task analyzed in the framework of reinforcement learning models. To liken our model-based definition to the psychological concept of optimism, we confronted the computational and behavioral variables capturing optimistic reinforcement learning with a standard measure of dispositional optimism: the Life Orientation Test – revised. The LOT-R is the most commonly used self-report measure of dispositional optimism (Glaesmer et al., 2012) and has been accordingly found in previous studies to correlate with behavioral and neural signatures of the good news/bad news effect (Sharot et al., 2011). We found a positive and significant correlation between learning optimistically in our task and being optimistic in real life, meaning that answering to question about the expectation of good things to happen in one's life is connected to the way one incorporates positive and negative prediction errors in simple reinforcement situations.

An important question is unanswered by our study and remains to be addressed. Whereas our results clearly show an asymmetry in the learning process, we cannot decide whether the learning process itself involves the representational space of values or in that of probabilities. This question is related to the broader debate whether the reinforcement or the Bayesian learning framework better captures learning and decision-making: two views that has been hard to disentangle, because of largely overlapping predictions, both at the behavioral and neural level (Hampton et al., 2006; Lebreton et al., 2015; Mathys et al., 2011).

A legitimate question is why such learning bias survived in the course of evolution? An obvious answer to this question is that being (unrealistically) optimistic is and/or has been, at least in certain conditions, adaptive, meaning that it confers an advantage. Consistent with this idea, in everyday life dispositional optimism (Carver et al., 2010) has been linked for instance to better global emotional well-being, interpersonal relationship or physical health. Optimists are less likely to develop coronary heart disease (Tindle et al., 2009), have broader social network (Macleod and Conway, 2005) and are less subject to distress when facing adversity (Scheier et al., 1994). Such advantages of dispositional optimism could explain, at least in part, the pervasiveness of an optimistic bias in human. Concerning the specific context of optimistic reinforcement learning a recent paper Cazé and al. showed that in certain conditions (low rewarding environments), an agent learning asymmetrically in an optimistic manner (i.e. with a higher learning rate for positive than for negative feedback) objectively outperforms another "unbiased" agent in a simple probabilistic learning task. Thus, before any social, well-being or health consideration, it is normatively advantageous (in certain contingencies) to take more into account positive than negative feedback. Thus a possible explanation for an asymmetric learning system is that the conditions identified by Cazé et al. closely resemble to the statistics of the natural environment that shaped the evolution of our learning system.

Finally, when reasoning about the adaptive value of optimism, a crucial point to take into account is the significant inter-individual variability of unrealistic optimism (Garrett et al., 2014; Moutsiana et al., 2015, 2013; Sharot et al., 2011). As a social animal humans face both private and collective decision-making problems (Raafat et al., 2009). An intriguing possibility is that multiple "sub-optimal" reinforcement learning strategies are maintained in the natural population to ensure an "optimal" learning repertoire, flexible enough to solve at the group-level the value learning and exploration-exploitation tradeoff (Hills et al., 2014). This hypothesis needs to be formally addressed using evolutionary simulations.

To conclude, our findings shed new light on the nature of the good news/bad news effect and therefore on the mechanistic origins of unrealistic optimism. We found that the optimistic learning is "not specific" to high-level belief updating but a particular consequence of a more general "low-level" instrumental learning asymmetry, both anatomically and functionally linked to the dorsomedial prefrontal cortex.

Methods

Subjects.

The first dataset (N=50) served as a cohort of healthy control subjects in a previous clinical neuroimaging study (Worbe et al., 2011). The second dataset involved the recruitment of new subjects (N=35). The local ethics committees approved both experiments. All subjects gave written informed consent before inclusion in the study and the study was carried out in accordance with the declaration of Helsinki (1964, revised 2013). In both studies the inclusion criteria were being older than 18 years and having no history of neurologic or psychiatric disorders. In experiments 1 and 2, men / women ratios were 27/23 and 20/15 respectively and the age means 27.1 ± 1.3 and 23.5 ± 0.7 respectively (expressed as mean ± S.E.M). In the first experiment subjects believed that they would be playing for real money, but to avoid discrimination between healthy subjects and

[b] Original French citation: "Qu'est-ce qu'optimisme? disait Cacambo. – Hélas! dit Candide, c'est la rage de soutenir que tout est bien quand on est mal." Voltaire (2014), *Candide ou l'optimisme*, Arvensa editions, p56, Ch. XIX. (Original work published in 1759).

patients, the final payoff was rounded up to a fixed amount of 80€ for every participant. In the second experiment subjects were paid the exact amount of money earned in the learning task, plus a fixed amount (average payoff 15.7±7.6€). **Table 3** reports the demographic features of the two cohorts.

Behavioral task and analyses

Subjects performed a probabilistic instrumental learning task described previously (Palminteri et al., 2009) (**Figure 1A**). Briefly, the task involved choosing between two cues that were associated with stationary reward probability (25% or 75%). There were 4 pairs of cues, randomly constituted and assigned to the 4 possible combinations of probabilities (25/25%, 25/75%, 75/25%, and 75/75%). Subjects were encouraged to accumulate as much money as possible and were informed that some cues would result in a win more often than others (the instructions have been published in appendix of the original study (Palminteri et al., 2009)). Subjects were given no explicit information regarding reward probabilities, which they had to learn through trial and error. The positive outcome reward was winning money (+0.50€); the negative outcome was getting nothing (0.0€) in the first experiment and losing money (-0.50€) in the second experiment. Subjects made their choice by pressing left or right response buttons with a left or right hand finger. Two given cues were always presented together, thus forming a fixed pair (choice context).

Regarding payoff, learning mattered only for pairs with unequal probabilities (75/25% and 25/75%). As dependent variable we extracted the correct response rate in asymmetric conditions (i.e. the left response rate for the 75/25% pair and right response rate in the 25/75% pair) (**Figure 1B**). In symmetrical reward probability conditions, we calculated the so-called "preferred response rate". The preferred response was defined as the most chosen option, i.e. chosen by the subject more than 50% of the trials. This quantity is therefore, by definition greater that 50%. The analyses focused on the preferred choice rate in the low reward condition (25/25%), where standard models predict greater frequency of negative prediction errors. Behavioral variables were compared within-subjects using paired two-tailed t-test and between-subjects using two-sample two-tailed t-test. Interactions were assessed using ANOVA.

Optimistic trait

In the second experiment, subjects also completed the French version of Life Orientation Test – Revised (LOT- R), which is a classical and the most frequently used (Glaesmer et al., 2012) psychological self report measure of dispositional optimism (Scheier et al., 1994; Trottier et al., 2008). Furthermore, recent studies showed that this scale might also explain part of the inter-individual variability associated with the prefrontal neural signature of the good news/bad news effect (Sharot et al., 2011). The scale includes ten sentences to be graded from 0 to 4 (from total disagreement to total agreement) and gives a score from pessimistic to optimistic. Subjects were then ranked according their scoring as previously described and their ranking was used to assess the correlation of individual scores with behavioral and computational measures of optimistic behavior (Sharot et al., 2007). We nonetheless note that the same analysis with raw scores lead to similar results.

Computational models

We fitted the data with reinforcement learning models. The model space included a standard Rescorla-Wager model (or Q-leanring) (Rescorla and Wagner, 1972; Sutton and Barto, 1998) (thereafter referred to as RW) and a modified version of the latter accounting differentially for learning from positive and negative prediction errors (thereafter referred to as RW±) (Frank et al., 2007; Niv et al., 2015). For each pair of cues, the model estimates the expected values of left and right options, Q_L and Q_R, on the basis of individual sequences of choices and outcomes. These Q values essentially represent the expected reward obtained by taking a particular option in a given context. In the first experiment, that involved only reward and reward omission, Q values were set at 0.25€ before learning, corresponding to the a priori expectation of 50% chance of winning 0.5€ plus a 50% chance of getting nothing. In the second experiment, which involved reward and punishment, Q values were set at 0.0€ before learning, corresponding to the a priori expectation of 50% chance of winning 0.5€ plus 50% chance of losing 0.5€. After every trial *t*, the value of the chosen option (e.g., L) was updated according to the following rule:

(1)

$$Q_L(t+1) = Q_L(t) + \alpha * \delta(t).$$

In the equation, $\delta(t)$ was the prediction error, calculated as:

(2)

$$\delta(t) = R(t) - Q_L(t),$$

and $R(t)$ was the reward obtained as an outcome of choosing L at trial t. In other words, the prediction error $\delta(t)$ is the difference between the expected reward $Q_L(t)$ and the actual reward $R(t)$. The reward magnitude R was +0.5 for winning 0.5€, 0 for getting nothing and -0.5 for losing 0.5€. The learning rate, α, is a scaling parameter that adjusts the amplitude of value changes from one trial to the next. Following this rule, option values are increased if the outcome is better than expected and decreased in the opposite case and the amplitude of the update is similar following positive and negative prediction errors.

The modified version of Q-Learning algorithm (RW±) differs from the original one (RW) by its Q values updating rule as follows:

(3)

$$Q_L(t+1) = Q_L(t) + \begin{matrix} \alpha^+.\delta(t) & if\ \delta(t) > 0 \\ \alpha^-.\delta(t) & if\ \delta(t) < 0 \end{matrix}$$

The learning rate α^+ adjusts the amplitude of value changes from one trial to the next when prediction error is positive (when the actual reward $R(t)$ is better than the expected reward $Q_L(t)$) and the second learning rate α^- does the same when prediction error is negative. Thus the RW± model allows for the amplitude of the update being different, following positive ("good news") and negative ("bad news") prediction errors and permits to account for individual differences in the way subjects learn from positive and negative experience. If both learning rates are equivalent, $\alpha^+ = \alpha^-$, RW± model equals the RW model. If $\alpha^+ > \alpha^-$, subjects learn more from positive than negative events. We refer to this case here as optimistic reinforcement learning. If $\alpha^+ < \alpha^-$, subjects learn more from negative than positive events. We refer to this case here as pessimistic reinforcement learning (**Figure 2B**).

Finally, given the Q values, the associated probability (or likelihood) of selecting each option was estimated by implementing the soft-max rule for choosing L, which is as follows:

(4)

$$P_L(t) = e^{(Q_{\square}(t)*\beta)}/(e^{(Q_L(t)*\beta)} + e^{(Q_L(t)*\beta)}).$$

This is a standard stochastic decision rule that calculates the probability of selecting one of a set of options according to their associated values. The temperature, β, is another scaling parameter that adjusts the stochasticity of decision-making and by doing so controls the exploration/exploitation trade-off.

Model comparison and subjects categorization

We optimized model parameters by minimizing the negative log-likelihood of the data given different parameters settings using Matlab's fmincon function, as previously described (Palminteri et al., 2015). Negative log-likelihoods (LLmax) were used to compute at the individual level (random effects) for each model the Bayesian information criterion as follows:

(5)

$$BIC = \log(ntrials)\, df + 2LLmax$$

We computed then the inter-individual average BIC in order to compare the quality of the fit of the two models, while accounting for their difference in complexity. The intra-individual difference in BIC (DBIC=BIC_{RW} - $BIC_{RW\pm}$) was also computed in order to categorize subjects in two groups and to quantitatively describe at the individual level the diverge from a realistic model (**Figure 2A**): RW± subjects, whose DBIC is positive, are better explained by RW± model. RW± subjects, whose DBIC is negative, are better explained by RW± model. We note that lower BIC indicated better fit.

The model parameters, (α^+, α^- and 1/β) were also compared between the two groups of subjects. Learning rates were compared using a mixed ANOVA with group (RW vs RW±) as a between-subject factor and learning rate type (+ or -) as a within-subject factor. The temperature was compared using a two-sample two-tailed t-test.

Model simulation

We also analyzed the models' generative performance by the mean of model simulations. For each participant we devised a virtual subject, represented by a set of individual best fitting parameters obtained. Each virtual subject dataset was obtained averaging 100

simulations, to avoid any local effect of the individual history of choice and outcome. The model simulations included all task conditions. The evaluation of generative performances involved the assessment of the "winning model's" ability to reproduce the key statistical effects of the data, as opposite to the "losing model". Unlike Bayesian model comparison, model simulation comparison is bounded to a particular behavioral effect of interest (in our case the preferred response rate. The model simulation analysis, which is focused on the evidence "against" a given model, it is complementary to the Bayesian model comparison analysis, which is focused in the evidence in favor of a model (Dienes, 2008; Popper, 1959).

Imaging data Acquisition & Analysis

Subject of the first experiment (N=50) performed the task magnetic resonance imaging (MRI) scanning. T1-weighted structural images and T2*-weighted echo planar images (EPIs) were acquired during the first experiment and analyzed with the Statistical Parametric Mapping software (SPM8; Wellcome Department of Imaging Neuroscience, London, England). Acquisition and preprocessing parameters were previously and extensively described (Palminteri et al., 2009; Worbe et al., 2011). We refer to these publications for details about image acquisition and preprocessing.

Voxel-based morphometry

The aim of the VBM analysis was to link computational phenotypes to neuroanatomical signatures, thus providing external validity to the computational constructs (optimistic as opposite to realistic) identified in the behavioral analysis. As in previous healthy subjects and patients studies, VBM analysis relied on the Diffeomorphic Anatomical Registration Through Exponentiated Lie algebra (DARTEL) toolbox implemented in SPM8 software and followed the standard procedure outlined in the VBM tutorial (Ashburner and Friston, 2000; Lebreton et al., 2013; Palminteri et al., 2012). The individual grey matter images images were entered in a multiple regression design analysis with one variable of interest (ΔBIC, as an integrate computational measure of optimistic reinforcement learning) and three regressors of no interest: gender, age and total intracranial volume used as covariates to control for demographic characteristics and confounding effects of brain size. All significant disclosed on glass brain and coronal slices (**Figure 3A** and **3B**), and reported in the text survived a threshold of $p<0.001$ (uncorrected) over the whole brain and contained a minimum of 30 contiguous voxels. The areas identified with this model-based VBM analysis (DMPFC and DLPFC) were then used as regions of interest (ROIs) for a supplementary VBM analysis, aimed to confirm the result with a model-free signature of optimistic learning (the preferred choice rate in the 25/25% condition) and for the subsequent fMRI analysis.

Functional magnetic resonance imaging

The aim of the fMRI analysis was to explore the functional consequence of the neuroanatomical differences observed in the prefrontal cortex and associated with optimistic trait. The fMRI analysis was based on a single general linear model. Each trial was modeled as having two time points, stimuli and outcome onsets. Each time point was regressed with a parameter modulator. Stimuli onset was modulated by the reaction time at each trial (continuous variable); outcome onset was modulated by the outcome obtained in the trial (reward: 0.50€ and no reward: 0.0€). The parametric modulators were z-scored to ensure between subject scaling of regression coefficients (Palminteri and Lebreton, 2016). Given that different subjects have been shown to implement different computational strategies, to allow between-group commensurability we opted for taking as parametric modulators model-free quantities: reaction times, which reflect the choice difficulty (decision process) and outcome representation, which is used in the learning (update process). Linear contrasts of regression coefficients were computed at the subject level and compared against zero to assess the presence of variable encoding effect and between-group to assess functional correlates of the two computational phenotypes. The comparisons were made within ROIs (DMPFC and DLPFC) identified with the VBM analysis that served as hypotheses generator for the fMRI analysis.

Acknowledgments:

The authors acknowledge Yulia Worbe and Mathias Pessiglione for granting access to the first dataset. We thank Valentin Wyart and Bahador Bahrami for helpful comments. SP was supported by a Marie Sklodowska-Curie Individual European Fellowship (PIEF-GA-2012 Grant 328822). GL was supported by a PHD fellowship of the Ministère de l'enseignement supérieur et de la recherche. ML was supported by an EU Marie Sklodowska-Curie Individual Fellowship (IF-2015 Grant 657904) and acknowledges the support of the Bettencourt-Schueller Foundation. The second experiment was supported by the ANR-ORA, Nesshi 2010-2015 research project to SBG.

Bibliography

Ashburner, J., Friston, K.J., 2000. Voxel-Based Morphometry—The Methods. Neuroimage 11, 805–821. doi:10.1006/nimg.2000.0582

Carver, C.S., Scheier, M.F., Segerstrom, S.C., 2010. Optimism. Clin. Psychol. Rev. 30, 879–889. doi:10.1016/j.cpr.2010.01.006

Dienes, Z., 2008. Understanding Psychology as a Science: An Introduction to Scientific and Statistical Inference. Palgrave Macmillan.

Doll, B.B., Hutchison, K.E., Frank, M.J., 2011. Dopaminergic Genes Predict Individual Differences in Susceptibility to Confirmation Bias. J. Neurosci. 31, 6188–6198. doi:10.1523/JNEUROSCI.6486-10.2011

Eil, D., Rao, J.M., 2011. The good news-bad news effect: Asymmetric processing of objective information about yourself. Am. Econ. J. Microeconomics 3, 114–138. doi:10.1257/mic.3.2.114

Frank, M.J., Moustafa, A. a, Haughey, H.M., Curran, T., Hutchison, K.E., 2007. Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. Proc. Natl. Acad. Sci. U. S. A. 104, 16311–16316. doi:10.1073/pnas.0706111104

Frank, M.J., Seeberger, L.C., Reilly, R.C.O., O'Reilly, R.C., 2004. By carrot or by stick: cognitive reinforcement learning in parkinsonism. Science 306, 1940–3. doi:10.1126/science.1102941

Garrett, N., Sharot, T., Faulkner, P., Korn, C.W., Roiser, J.P., Dolan, R.J., 2014. Losing the rose tinted glasses: neural substrates of unbiased belief updating in depression. Front. Hum. Neurosci. 8, 639. doi:10.3389/fnhum.2014.00639

Gifford, R., 2011. The dragons of inaction: pychological barriers that limit climate change mitigation and adaptation. Am. Psychol. 66, 290–302. doi:10.1037/a0023566

Glaesmer, H., Rief, W., Martin, A., Mewes, R., Brähler, E., Zenger, M., Hinz, A., 2012. Psychometric properties and population-based norms of the Life Orientation Test Revised (LOT-R). Br. J. Health Psychol. 17, 432–445. doi:10.1111/j.2044-8287.2011.02046.x

Hampton, A.N., Bossaerts, P., 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. doi:10.1523/JNEUROSCI.1010-06.2006

Hills, T.T., Todd, P.M., Lazer, D., Redish, a. D., Couzin, I.D., 2014. Exploration versus exploitation in space, mind, and society. Trends Cogn. Sci. 19. doi:10.1016/j.tics.2014.10.004

Kahneman, D., Tversky, A., 1979. Prospect Theory: An Analysis of Decision under Risk Daniel Kahneman; Amos Tversky. Econometrica 47, 263–292. doi:10.1111/j.1536-7150.2011.00774.x

Kanai, R., Rees, G., 2011. The structural basis of inter-individual differences in human behaviour and cognition. Nat. Rev. Neurosci. 12, 231–242. doi:10.1038/nrn3000

Kolling, N., Behrens, T.E.J., Mars, R.B., Rushworth, M.F.S., 2012. Neural Mechanisms of Foraging. Science (80-.). 336, 95–98. doi:10.1126/science.1216930

Lebreton, M., Abitbol, R., Daunizeau, J., Pessiglione, M., 2015. Automatic integration of confidence in the brain valuation signal. Nat. Neurosci. doi:10.1038/nn.4064

Lebreton, M., Bertoux, M., Boutet, C., Lehericy, S., Dubois, B., Fossati, P., Pessiglione, M., 2013. A Critical Role for the Hippocampus in the Valuation of Imagined Outcomes. PLoS Biol. 11. doi:10.1371/journal.pbio.1001684

Macleod, A.K., Conway, C., 2005. Well-being and the anticipation of future positive experiences: The role of income, social networks, and planning ability. Cogn. Emot. 19, 357–74. doi:10.1080/02699930441000247

Mathys, C., Daunizeau, J., Friston, K.J., Stephan, K.E., 2011. A bayesian foundation for individual learning under uncertainty.

Front. Hum. Neurosci. 5, 39. doi:10.3389/fnhum.2011.00039
Moutsiana, C., Charpentier, C.J., Garrett, N., Cohen, M.X., Sharot, T., 2015. Human Frontal-Subcortical Circuit and Asymmetric Belief Updating. J. Neurosci. 35, 14077–14085. doi:10.1523/JNEUROSCI.1120-15.2015
Moutsiana, C., Garrett, N., Clarke, R.C., Lotto, R.B., Blakemore, S.-J., Sharot, T., 2013. Human development of the ability to learn from bad news. Proc. Natl. Acad. Sci. 110, 16396–16401. doi:10.1073/pnas.1305631110
Niv, Y., Daniel, R., Geana, A., Gershman, S.J., Leong, Y.C., Radulescu, A., Wilson, R.C., 2015. Reinforcement Learning in Multidimensional Environments Relies on Attention Mechanisms. J. Neurosci. 35, 8145–8157. doi:10.1523/JNEUROSCI.2978-14.2015
Palminteri, S., Boraud, T., Lafargue, G., Dubois, B., Pessiglione, M., 2009. Brain Hemispheres Selectively Track the Expected Value of Contralateral Options. J. Neurosci. 29, 13465–13472. doi:10.1523/JNEUROSCI.1500-09.2009
Palminteri, S., Justo, D., Jauffret, C., Pavlicek, B., Dauta, A., Delmaire, C., Czernecki, V., Karachi, C., Capelle, L., Durr, A., Pessiglione, M., 2012. Critical Roles for Anterior Insula and Dorsal Striatum in Punishment-Based Avoidance Learning. Neuron 76, 998–1009. doi:10.1016/j.neuron.2012.10.017
Palminteri, S., Khamassi, M., Joffily, M., Coricelli, G., 2015. Contextual modulation of value signals in reward and punishment learning. Nat. Commun. 6, 8096. doi:10.1038/ncomms9096
Palminteri, S., Lebreton, M., 2016. Assessing inter-individual variability in brain-behavior relationship with functional neuroimaging. bioRxiv 1–14. doi:http://dx.doi.org/10.1101/036772
Popper, K., 1959. The logic of scientific discovery, Journal of the Franklin Institute. doi:10.1016/S0016-0032(59)90407-7
Raafat, R.M., Chater, N., Frith, C., 2009. Herding in humans. Trends Cogn. Sci. 13, 420–428. doi:10.1016/j.tics.2009.08.002
Rescorla, R.A., Wagner, A.R., 1972. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement, in: Classical Conditioning: Current Research and Theory. p. 497.
Scheier, M.F., Carver, C.S., Bridges, M.W., 1994. Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A reevaluation of the Life Orientation Test. J. Pers. Soc. Psychol. doi:10.1037//0022-3514.67.6.1063
Schoenbaum, M., 1997. Do smokers understand the mortality effects of smoking? Evidence from the health and retirement survey. Am. J. Public Health 87, 755–759. doi:10.2105/AJPH.87.5.755
Sharot, T., Garrett, N., 2016. Forming Beliefs: Why Valence Matters. Trends Cogn. Sci. 20, 25–33. doi:10.1016/j.tics.2015.11.002
Sharot, T., Korn, C.W., Dolan, R.J., 2011. How unrealistic optimism is maintained in the face of reality. Nat. Neurosci. 14, 1475–1479. doi:10.1038/nn.2949
Sharot, T., Riccardi, A.M., Raio, C.M., Phelps, E. a, 2007. Neural mechanisms mediating optimism bias. Nature 450, 102–105. doi:10.1038/nature06280
Shenhav, A., Botvinick, M.M., Cohen, J.D., 2013. The expected value of control: An integrative theory of anterior cingulate cortex function. Neuron. doi:10.1016/j.neuron.2013.07.007
Shenhav, A., Straccia, M. a, Cohen, J.D., Botvinick, M.M., 2014. Anterior cingulate engagement in a foraging context reflects choice difficulty, not foraging value. Nat. Neurosci. 17, 1249–1254. doi:10.1038/nn.3771
Shepperd, J. a., Klein, W.M.P., Waters, E. a., Weinstein, N.D., 2013. Taking Stock of Unrealistic Optimism. Perspect. Psychol. Sci. 8, 395–411. doi:10.1177/1745691613485247
Shepperd, J. a., Ouellette, J. a., Fernandez, J.K., 1996. Abandoning unrealistic optimism: Performance estimates and the temporal proximity of self-relevant feedback. J. Pers. Soc. Psychol. 70, 844–855. doi:10.1037/0022-3514.70.4.844
Shepperd, J.A., Waters, E.A., Weinstein, N.D., Klein, W.M.P., 2015. A Primer on Unrealistic Optimism. Curr. Dir. Psychol. Sci. 24, 232–237. doi:10.1177/0963721414568341
Sutton, R.S., Barto, A.G., 1998. Introduction to Reinforcement Learning. Learning 4, 1–5. doi:10.1.1.32.7692
Tindle, H.A., Chang, Y.-F., Kuller, L.H., Manson, J.E., Robinson, J.G., Rosal, M.C., Siegle, G.J., Matthews, K.A., 2009. Optimism, Cynical Hostility, and Incident Coronary Heart Disease and Mortality in the Women's Health Initiative. Circulation 120, 656–662. doi:10.1161/CIRCULATIONAHA.108.827642
Trottier, C., Trudel, P., Mageau, G., Halliwell, W.R., 2008. Validation de la version canadienne-française du Life Orientation Test-Revised. Can. J. Behav. Sci. 40, 238–243. doi:10.1037/a0013244
Ullsperger, M., Fischer, A.G., Nigbur, R., Endrass, T., 2014. Neural mechanisms and temporal dynamics of performance monitoring. Trends Cogn. Sci. 1–9. doi:10.1016/j.tics.2014.02.009
Waters, E.A., Klein, W.M., Moser, R.P., Yu, M., Waldron, W.R., McNeel, T.S., Freedman, A.N., 2011. Correlates of unrealistic risk beliefs in a nationally representative sample. J Behav Med 34, 225–235. doi:10.1007/s10865-010-9303-7
Weinstein, N.D., 1980. Unrealistic Optimism About Future Life events. J. Pers. Soc. Psychol. 39, 806–820. doi:10.1037/a0020997
Worbe, Y., Palminteri, S., Hartmann, A., Vidailhet, M., Lehéricy, S., Pessiglione, M., 2011. Reinforcement Learning and Gilles de la Tourette Syndrome. Arch. Gen. Psychiatry 68, 1257–1266.

Table 1: models fitting and parameters in the two experiments.

The table summarizes for each model its fitting performances and its average parameters: LLmax: maximal Log Likelihood; BIC:

Experiment / Model	LLmax	BIC	α	α^+	α^-	$1/\beta$
Experiment 1 (N=50)						
RW Model	45.1±2.2	99.4±4.4	0.32±0.05	-	-	0.16±0.03
RW± Model	40.0±2.4	93.6±4.7*	-	0.36±0.05#	0.22±0.05	0.13±0.03
Experiment 2 (N=35)						
RW Model	44.2±2.9	96.2±5.9	0.24±0.05	-	-	0.53±0.16
RW± Model	38.1±3.0	87.7±6.0*	-	0.45±0.06#	0.18±0.05	0.31±0.10

Bayesian Information Criterion (computed from LLmax); Alpha: learning rate for both positive and negative prediction errors (RW model); Alpha(+) : learning rate for positive prediction errors; Alpha(-) : average learning rate for negative prediction errors (RW± model); 1/Beta: average inverse of model temperature. Data are expressed as mean ± s.e.m. *P<0.01 comparing between the two models. #P<0.001 comparing between the two learning rates.

Table 2: Behavioral and simulated data.

Experiment / Model	Correct Response	Correct Response (RW model)	Correct Response (RW ± model)	Preferred Response	Preferred Response (RW model)	Preferred Response (RW± model)
Condition(s)	**Asymmetric**			**Symmetric (25/25%)**		
Experiment 1 (N=50)						
RW Group	74.25 ± 3.65	75.20 ± 2.55	75.35 ± 2.42	61.5 ± 1.94	58.14 ± 0.67	59.47 ± 0.81
RW± Group	77.83 ± 3.25	75.58 ± 1.94	77.75 ± 1.44	72.75 ± 2.63*	58.84 ± 0.55	69.36 ± 0.99
Experiment 2 (N=35)						
RW Group	72.01 ± 4.74	72.63 ± 3.64	72.51 ± 3.51	61.59 ± 2.23	57.10 ± 0.83	59.22 ± 0.89
RW± Group	76.21 ± 4.58	76.18 ± 2.62	78.55 ± 1.98	73.36 ± 3.17*	58.69 ± 0.74	70.72 ± 1.52

The table summarizes for each experiment and each group of subjects, behavioral and simulated dependent variables: Both real and simulated Correct Response in asymmetric conditions and both real and simulated Preferred Response in 25/25% condition. Data are expressed as mean ± s.e.m (in percentage). *P<0.01 two sample t-test.

Figures and legends

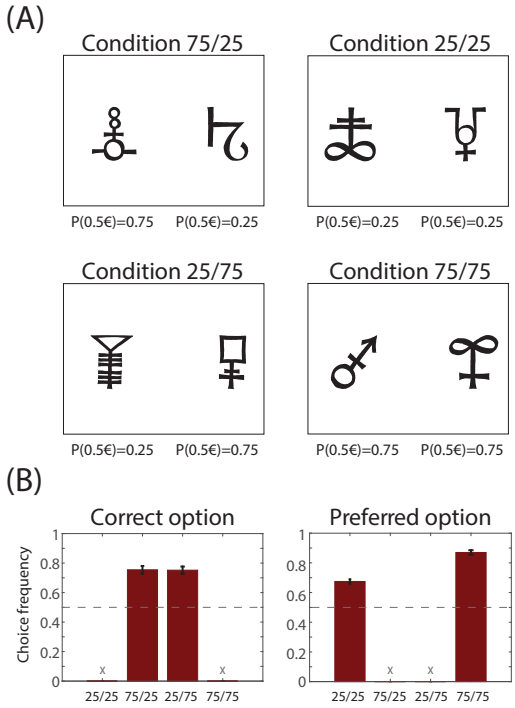

Figure 1: behavioral task and variables.
(**A)** Task's conditions and contingencies. Subjects selected between left and right symbols. Each symbol was associated with a stationary probability (p = 0.25 or 0.75) of winning 0.50€ and a reciprocal probability (1 – p) of getting nothing (first experiment) or losing 0.50€ (second experiment). In two conditions (rightmost column) the reward probability was the same in both symbols ("symmetric" conditions) and in two other conditions (leftmost column) the reward probability was different across symbols ("asymmetric" conditions). Note that the assignment between symbols conditions was randomized across subjects. (**B)** Dependent variables. In the leftmost panel, the histograms show the correct choice rate (i.e. choices directed toward the most rewarding stimulus in the asymmetric conditions). In the rightmost panel the histograms show the preferred option choice rate (i.e. the option chosen by subjects in more than 50% of the trials; this measure is relevant only in the symmetric conditions, where there is no intrinsic correct response). Bars indicate the mean and error bars indicate the SEM. Data are taken from the first experiment (N=85).

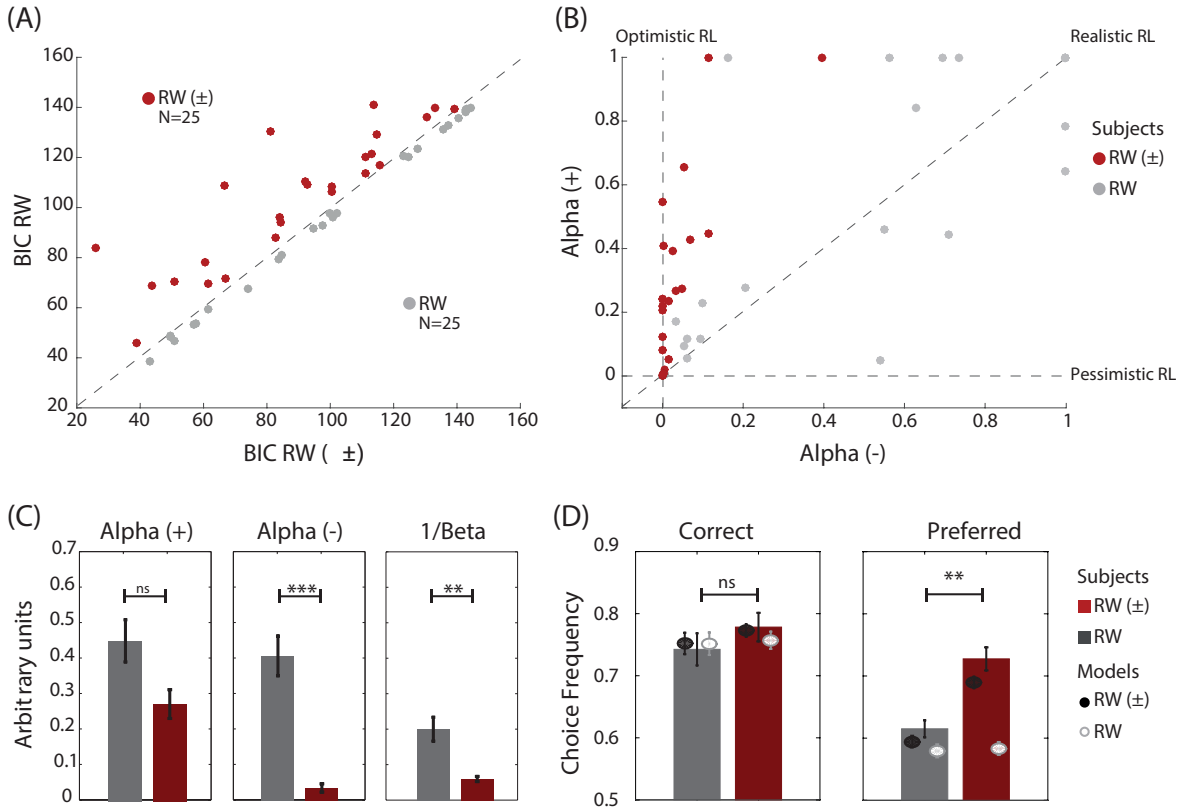

Figure 2: behavioral and computational identification of optimistic reinforcement learning
(A) Model comparison. The graphic displays the scatter plot of the BIC calculated for the RW model as a function of the BIC calculated for the RW± model. Smaller BIC values indicate better fits. Subjects are clustered in two populations according to the BIC difference ($\Delta BIC = BIC_{RW} - BIC_{RW\pm}$) between the two models. RW± subjects (displayed in red) are characterized by a positive ΔBIC, indicating that the RW± model better explains their behavior. RW subjects (displayed in grey) are characterized by a negative ΔBIC, indicating that the RW model better explains their behavior. (**B)** Model parameters. The graphic displays the scatter plot of the learning rate following positive prediction errors (α^+) as a function of the learning rate following negative prediction errors (α^-), obtained from the RW± model. "Standard" reinforcement learners are characterized by similar learning rates for both types of prediction errors. "Optimistic" learners are characterized by a bigger learning rate only for positive compared to negative prediction errors. "Pessimistic" learners are characterized by the opposite pattern. **C** The histograms the RW± model free parameters (the learning rates + and - and the inverse temperature 1/β) as function of the subjects' populations. **D** Actual and simulated choice rates. Histograms represent the observed and dots represent the model simulated of choices for both populations and both models, respectively for correct option (extracted from asymmetric condition), and from preferred option (extracted from the symmetrical condition 25/25%, see **Figure 1A**). Model simulations are obtained using the individual best fitting free parameters. *p<0.05, **, p<0.01, ***p<0.001, two-sample two-sided t-test.

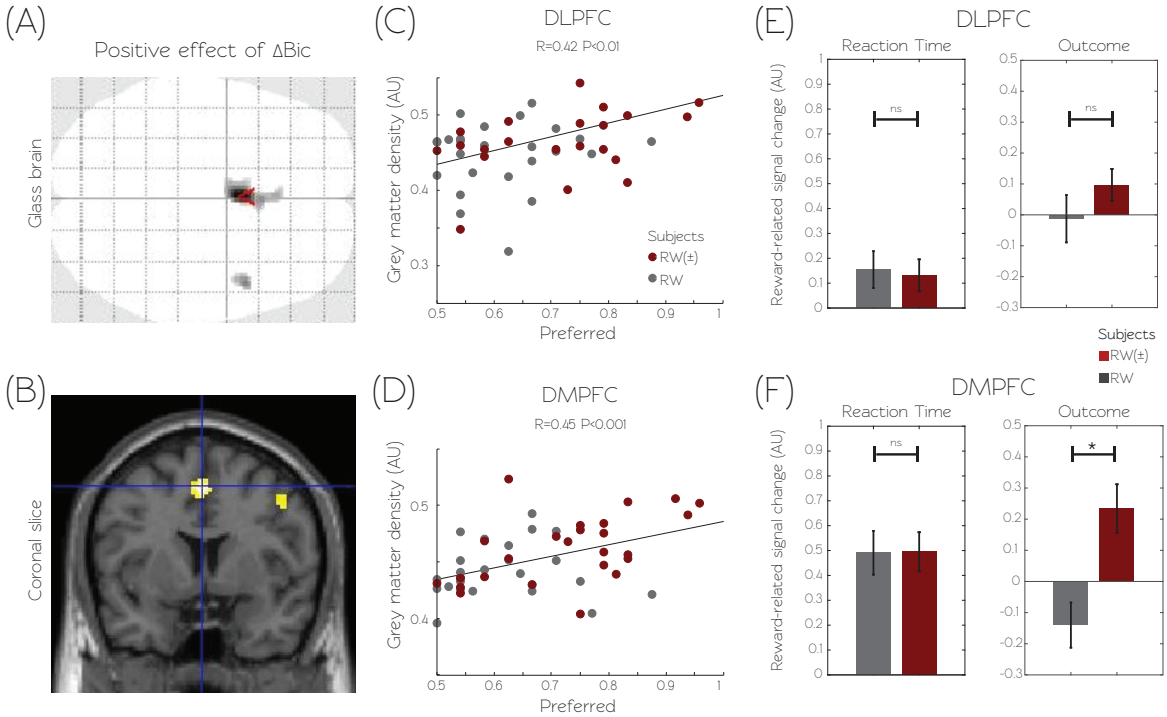

Figure 3: neural signatures of the optimistic reinforcement learning

(A) and **(B)** Computational correlation. Statistical parametric maps of grey matter density positively correlating with the ΔBIC (ΔBIC = BIC_{RW} - $BIC_{RW\pm}$). Areas colored in gray-to-black gradient on the axial glass brain and red-to-white gradient on the coronal slice showed a significant effect (*p<0.001 uncorrected with a cluster extent of minimum 30 contiguous voxels*). (**C**) and (**D**) Behavioral correlation. The scatter plots represent the Gray matter density in the dorsolateral and the dorsomedial prefrontal cortex (DLPFC and DMPFC) as a function of the preferred response choice rate (Figure 2C). RW± subjects (displayed in red) are characterized by a positive ΔBIC, indicating that their behavior is better explained by the RW± model. RW subjects (displayed in grey) are characterized by a negative ΔBIC, indicating that their behavior is better explained by the RW model. (**E**) and (**F**) Functional consequences. Histogram shows reward-related and time-related signals change in DLPFC and DMPFC at the time of reward onset for both populations. Bars indicate the mean and error bars indicate the SEM. *p<0.05, unpaired t-tests.

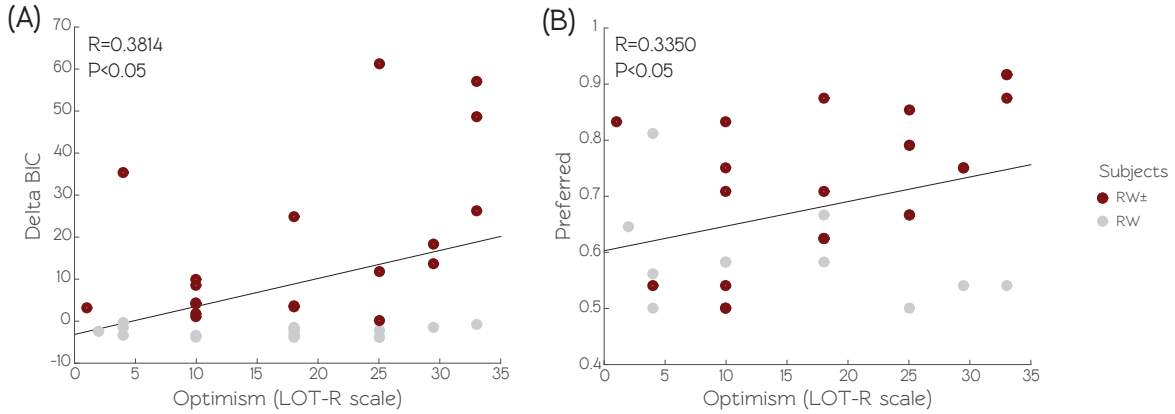

Figure 4: relation between optimistic reinforcement learning and optimistic life orientation trait.
(**A)** The scatter plot represents the ΔBIC variable (ΔBIC = BIC_{RW} - $BIC_{RW\pm}$) as a function of individual ranking in the optimistic life orientation trait (derived from the LOT-R scale). (**B)** The scatter plot represents the preferred option choice rate (calculated in the 25/25% condition) as a function of individual ranking in the optimistic life orientation trait (derived from the LOT-R scale).