Variability in prior expectations explains biases in confidence reports

Pablo Tano[1]*, Florent Meyniel[2], Mariano Sigman[3] and Alejo Salles[4]

Confidence in a decision is defined statistically as the probability of that decision being correct. Humans, however, tend both to under- and over-estimate their accuracy (and hence, their confidence), as has been exposed in various experiments. Here, we show that this apparent irrationality vanishes when taking into account prior participants' biases measured in a separate task. We use a wagering experiment to show that modeling subjects' choices allows for classifying individuals according to an optimism - pessimism bias that fully explains from first principles the differences in their later confidence reports. Our parameter-free confidence model predicts two counterintuitive patterns for individuals with different prior beliefs: pessimists should report higher confidence than optimists, and their confidences should depend differently on task difficulty. These findings show how apparently irrational confidence traits can be simply understood as differences in prior expectations. Furthermore, we show that reporting confidence actually impacts subsequent choices, increasing the tendency to explore when confidence is low, akin to a deconfirmation bias.

A level of confidence accompanies all of our decisions [1]. This sense of confidence can be reported explicitly, or implicitly through behavioral markers such as the amount of time willing to wait to obtain a response [2] or reaction times [3], the predisposition to wage [4] or opt-out to a lower but safe reward [5]. The use of such implicit measures has shown that a sense of confidence is present in rodents and nonhuman primates (see [6] for a review).

[1]Universidad de Buenos Aires. Facultad de Ciencias Exactas y Naturales. Instituto de Cálculo. Buenos Aires, Argentina. [2]Cognitive Neuroimaging Unit / NeuroSpin center / Institute for Life Sciences Frédéric Joliot / Fundamental Research Division / Commissariat à l'Energie Atomique et aux énergies alternatives ; INSERM, Université Paris-Sud; Université Paris-Saclay; 91191 Gif-sur-Yvette, France. [3]Universidad Torcuato Di Tella, Almirante Juan Saenz Valiente 1010, C1428BIJ Buenos Aires, Argentina. [4] CONICET-Universidad de Buenos Aires. Facultad de Ciencias Exactas y Naturales. Instituto de Cálculo. Buenos Aires, Argentina. *e-mail: pablotanoretamales@gmail.com

A quantitative approach to confidence helps formalize the concept and unify its different manifestations. In statistical decision theory, the normative definition of decision confidence is the probability of a certain choice being correct [7-10]. Many models of how confidence emerges in the brain have been proposed, such as accumulators [2, 11-13], drift diffusion [3], and attractor models [14, 15]. In these models, confidence is interpreted as an algorithmic construction from variables available in the decision-making process, rather than a readout from a Bayesian representation of knowledge [15], using confidence metrics such as the difference between decision variables, post-decision evidence and reaction time combined with evidence [3, 13, 15, 17, 18]. Depending on the parameter values, these models can be close approximations to confidence as the probability of the decision being correct or instead show systematic deviations from optimality.

The normative account of human decision confidence was recently put to test in [19], which shows that a statistical definition of confidence indeed predicts various aspects of human confidence reports, in both perceptual and knowledge-based tasks. This study is part of a resurgence of rationality as a paramount human trait, which came about with the realization that probabilities are the proper language in contexts of uncertainty such as those we encounter in everyday life [20, 21, 22]. The Bayesian rationality program came a long way in explaining human behavior in a wide range of higher cognitive domains, such as intuitive physics [23], intuitive psychology [24, 25], or causal inference [26]. Multimodal sensory integration remains a classic illustration of the flexibility and optimality of our inference mechanisms [27, 28].

The normative account and those supporting results [19, 29] may seem at odds with the fact that people often deviate from the statistical definition of confidence and are typically over- or underconfident in many tasks [8, 30-39]. For example, Griffin and Tversky showed cases of over- and underconfidence in intuitive judgements, proposing that those biases arise because people do not take into account the reliability (a.k.a. weight, credence) of the evidence at hand. Since taking into account the reliability of evidence is a landmark of statistical computations, Griffin, Tversky and several others therefore dismissed the statistical framework as suitable for the understanding of confidence.

However, there are many reasons why subjects may deviate from optimality. It may indeed be the case that subjects' computation does not adhere to statistical principles. But departure from optimality may arise even when one adheres to such principles, but, for instance, computes only approximately, or uses incorrect priors [40, 41]. Indeed, theory predicts that Bayesian rational

agents with different prior beliefs will be relatively overconfident about the accuracy of their estimators [42]. In this work, we build on this idea, performing and modeling a wagering experiment to ask whether the apparent irrationality of confidence reports can be reconciled with the statistical account of confidence if participants' biases on prior expectations are taken into account. In line with the heuristics and biases program, it has been thoroughly documented that humans have a bias for optimism [43-45], for which even neural mechanisms have been described [46]. Recently, quantitative modeling has shown that in some contexts this bias can lead to counterintuitive behaviors, such as optimists being sometimes more conservative in their choices [47]. Independently of the nature of this phenomenon, we encode the expectation bias (be it optimistic or pessimistic) as the prior entering the Bayesian inference mechanism, and postulate that it is necessary and sufficient to fully explain the otherwise irrational biases in confidence.

To avoid the inherent circularity of accounting for biases with priors [48], we designed a task that associates a popular multi-armed bandits gambling task [49-51] and confidence reports. We used the gameplay (not the confidence reports) to characterize the individual's prior bias about the machines' reward rates according to how much they explore, exploit and get rewarded in the task. Therefore, the prior bias of each participant is defined independently of the potential impact on confidence that we aim to explain.

Having a full task model not only allows us to study whether priors that dictate choice also condition the confidence reports. It also serves us to tackle the less usual question attaining to the *effect* of reporting confidence. We directly compare experimental situations with and without a confidence prompt and show an effect of asking for confidence akin to a "de-confirmation" bias: when confidence is low, producing a confidence response decreases subjects' commitment to their prior choice in the task.

Our analytic approach was the following. First, we assured that subjects were widely distributed in a pessimistic-optimistic scale, and that this trait was stable throughout the task. Second, using Bayesian probabilistic learning, we inferred subjects' beliefs about the machines' reward rates and their confidence about knowing which machine payed more at the moment of the confidence report. The parameter-free model predicts that normative statistical confidence should depend upon an interaction between the prior bias (fitted independently from confidence) and task factors such as difficulty of the task, and, non-trivially, that confidence should be overall *lower* in optimistic subjects. The actual human patterns of confidence reports indeed varied in the predicted way as a

function of their previously adjusted optimistic/pessimistic prior. In sum, our results resolve the tension between the normative rational account of behavior and the irrational trends seen in confidence reports.

Results

Predictions from our model.

In the popular bandit gambling game, a subject faces several machines (two here) associated with different reward rates. The subject repeatedly decides which machine to play, and observes whether this choice is rewarded or not at that specific trial. The machines' reward rates are unknown to the subject, but they can be learnt over the course of the game. Modeling actions in such an uncertain environment can be divided in two components: the learning component, through which the observer updates his representation of knowledge given the observations, and the decision component, through which the observer makes a decision towards his goal based on current knowledge. In bandit tasks, the goal is to maximize the total reward over a fixed number of trials. The optimal solution for the decision component can be computed using dynamic programming, but it requires an amount of calculations that grows exponentially with the amount of trials [52] and is thus unfeasible as a psychological mechanism. By contrast, the optimal solution for the learning component is much cheaper computationally. It is afforded by Bayesian inference, which represents beliefs about the machines' payoffs as probability distributions.

For the learning component, observers begin each experimental block (wherein payoffs were fixed) with a prior distribution for the reward probability of each of the two machines in the experiment. We parameterized this distribution by using its mean b and a weight w, which quantifies how much does the agent trusts her prior beliefs (see *Methods* for details). These values (b and w) were fitted to subjects' choices. Low values of the prior mean b correspond to a pessimistic perspective, while high values of this parameter represent a more optimistic take. These distributions were then subjected to Bayesian updating after each machine choice and its corresponding outcome (see *Methods*). We posit (prediction #1) that subjects from the general population differ in their prior beliefs: there is a range of optimistic/pessimistic subjects (classified by their value of b), and those idiosyncratic priors should impact each subject's behavior consistently across different experimental conditions.

In the middle of a block, subjects were occasionally asked to report which machine payed more, and also to indicate their confidence in that answer. We formalized this confidence report as the probability of having identified correctly the best machine. The probability that the chosen machine has a higher reward rate than the other one depends on two factors: the difference in the distributions' means (*i.e.* one minus the *perceived difficulty*, denoted d), and the precision of those distributions (see *Methods* for model details). Importantly, those confidence levels are read out from the learnt distributions directly, without parameter fitting and independently from the decision-making process. In practice, for the parameters of this experiment, the optimal confidence level was mostly influenced by d, and comparatively little by the precision.

Decisions in gameplay, on the other hand, must face a tension between maximizing the immediate reward by selecting the machine that apparently pays more (a behavior known as exploitation) and exploring alternatives to more accurately learn the payoff associated to all machines and optimize future decisions. This exploration-exploitation tradeoff implies that in some situations the rational action (for long-term maximization of reward) would be to play the machine with the lowest payoff so far. Indeed, when comparing two decisions with the same machine history, humans chose the machine with the lowest payoff in (12±4)% of gameplay decisions, and only (4±3)% when they are asked which machine has the highest payoff so far (expressed as mean±s.d. , t=8.83, p<0.0001 (n=17)]. A heuristic solution to model this tradeoff is to more often try a specific machine the more it seems to pay than the other. We formalized this decision strategy with a sampling process characterized by the value of d and a parameter σ (adjusted and fixed to a single value for all subjects) which favours exploration by introducing uncertainty over the estimated value of the perceived difficulty d, increasing the probability of exploring the machine with the lowest payoff as d increases (see Model Details in *Methods*). We emphasize that the normative account of confidence in the 'which machine is better' decision depends only on the learning component and the decision made by the paricipant, that is, it is independent of the just mentioned process that may have given rise to the decision.

With the Bayesian solution to the learning component, one can expect two distinct features in the confidence reports when comparing high and low values of the prior bias b (*optimists* and *pessimists*, respectively). First, pessimists should in general have higher confidence than optimists, particularly when machines pay more (prediction #2). Although it may appear counter-intuitive that pessimistic people are more confident, this can be easily understood: expecting less, whenever they find a machine that pays somewhat well, they are very confident that this machine is indeed

the best one. This in turn is amplified in high generosity sceneraios because better machines are played more, so the beliefs of the unexplored machine remains dominated by the pessimistic prior. Optimists on the other hand expect more, so playing a good machine does not separate its reward distribution that much from the prior, and thus confidence is lower. This effect is illustrated in Fig. 1.

A second, subtler effect, is related to the change in confidence as a function of the difficulty of the task for agents with highly definite prior biases (w>2). We first introduce some useful nomenclature for this purpose. Selecting the best machine is difficult when both have paid similarly so far. We manipulated the *generative difficulty* of a block experimentally by systematically varying the difference between the real reward rates of machines. However, given that subjects experience those reward rates only through noisy observations, subjects may experience a difficulty level that departs from the generative difficulty level that we manipulated experimentally. This is particularly clear when a machine is left unexplored. To quantify objectively the difficulty of choices that subjects were exposed to, we computed the *unbiased difficulty*, namely, the difficulty that would be perceived by a subject who is unbiased (neither optimistic nor pessimistic). Specifically, it is computed as one minus the difference in mean reward rates at the moment of the report, as estimated by an agent with a non-informative prior distribution (i.e. with b and w fixed to 1 and 2). In the long run, if all options are repeatedly explored, prior biases (when any) will fade out and the *perceived difficulty*, which takes into account the prior bias of each subject, will be approximately equal to the unbiased difficulty.

In our model, if the machines pay little, both machines will be explored, optimistic and pessimistic prior biases will fade out, and all agents will experience a similar difficulty. Recall from above that in our model, confidence should normatively mostly reflect the perceived difficulty (d), so that when machines pay little, all subjects will agree to report confidence according to the same, unbiased difficulty. However, if at least one machine pays generously, a difference will arise between optimists and pessimists. Optimists, expecting more, will still eventually switch machines expecting a higher reward, making prior biases vanish and reporting confidence, as before, in line with the unbiased difficulty. Pessimists, on the other hand, will tend to stick to the higher paying machine, leaving the other one mostly unexplored, and therefore described essentially by its prior. This will in turn have the effect that for pessimists the perceived difficulty will be largely independent from the unbiased difficulty and therefore confidence in those subjects will be different from confidence

in optimistic subjects. This pattern of confidence changes across different difficulty levels is our prediction #3.

This last prediction can be summarized as follows: for every agent with highly definite prior beliefs (w>2), the perceived difficulty will be different from the unbiased difficulty of the task when some options are left unexplored (e.g. in high generosity situations). In this scenario, optimists are simply agents with a more explorative behavior than pessimists, and therefore are less influenced by this effect than pessimists.

Importantly, the two last predictions pertaining confidence are clearly falsifiable. This is because since the optimism levels for participants are estimated from independent data (i.e. the gameplay), the model for confidence has no adjustable parameters: its predictions are inescapable. Additionally, these predictions are caused by the Bayesian representation of knowledge, not by the specific confidence readout from beliefs' distributions. Therefore, similar effects of the prior bias can be predicted not only on the computation of statistical confidence, but also on other metrics like the distance between the means, the estimated payoff of the most generous machine, or the average payoff of both machines, to name a few. The critical aspect is that the estimation of the underlying distribution should take into account the subject's prior expectations.

Optimistic and pessimistic behaviors in gameplay

Participants played a two armed bandit game. Their task was to maximize the total reward. We varied the reward rates of machine only in distinct blocks, each comprising 16 trials (in total, 36,720 decisions and 765 confidence reports). There were three types of blocks: those with a question asked mid-block about which machine pays more, and the report of the continuous associated confidence level ('confidence' blocks), and as a control, blocks with only the 'which' question ('which' block) and blocks without any question ('no' blocks), see Fig. 2 and the *Methods* section for details.

We first model human gameplay behavior ignoring the confidence report. We fitted the model by varying b, w and σ across subjects. The simplest model, in which only b was varied, was the best one. Bayesian Model Comparison indicated that there was a probability xp>0.98 that this model was better than a model that varied only w or σ, and a probability xp>0.69 that it was better than a model that varied any combination of b, w and σ (see Analysis Details in *Methods*). In contrast with the other parameters, when varying b alone, the model showed a range of behaviors that covers the entire region displayed by human participants for various behavioral summaries (see Figs. 3 and 4);

and the adjusted value of b for each participant is consistent across different task conditions (see below). Therefore, from now on, we consider this simple model, varying only b across subjects, and fixing w and σ to their best fitting value at the group level (8 and 0.05 respectively). We also compare this model against two common strategies used in bandit problems [53]: Win-Stay-Lose-Shift (WSLS) and the optimal solution by dynamic programming. Neither of these is flexible enough to account for the behavioral palette observed in humans (see Fig 3).

We now move to our first hypothesis, namely, that a level of optimism can be consistently ascribed to each participant. The prior bias b, which captures the level of optimism, was fit individually for each participant. To test the idiosyncratic nature of this prior, we verified that it consistently impacts behavior across different conditions. More precisely, we split blocks in Generous and Easy, Generous and Hard, Avaricious and Easy, and Avaricious and Hard by median splitting the blocks according to their generative difficulty and their generative generosity (*i.e.* their average real reward rates). Then, we took two behavioral summaries: the average rewards vs. persistence (defined as the proportion of trials in which a machine was chosen immediately after a failure in that machine) and rewards vs. exploration obtained by the subjects in these four categories (Figs. 4 and S1, respectively), plus these same summaries on the aggregate blocks (Fig. 3). We fit the value of b for each participant so as to minimize the squared error between their behavior and the model's for all 10 summaries (the 10 panels in Figs. 3, 4 and S1). Figs. 3 and 4 show the values of b for the model and for each participant, which match those of the model and remain stable across the different categories. This amounts to a consistent assignment of prior bias for each participants. In other words, participants labeled as optimistic (pessimistic) in one category behave optimistically (pessimistically) in all other categories. We confirm the consistency of fitted b values by analyzing their variance across categories and across individuals. A permutation test yields $p < 10^{-5}$ (permutation test, see *Methods*), showing that the within-subject variance of fitted b values across different categories is much lower than the between-subject variance. We also performed a cross validation test by fitting each subject value of b for all categories except one, and compare this fitted value with the one fitted independently in the left-out category, yielding $R>0.71$ for all categories (n=17 subjects in each category, $p<0.01$ for all categories), reaffirming the intra-subject consistency of b across different environments.

Optimism inferred from gameplay behavior explains the bias in confidence

Our second prediction was that the level of optimism, fitted onto the gameplay, should predict the average confidence levels reported throughout the task. We examined participants' average

confidence reports for all payoff settings of the *'confidence'* type blocks after sorting participants into optimistic (*n=10*) and pessimistic (*n=7*), according to whether their prior bias *b* (fitted only from gameplay, without the use of the confidence reports) was greater or smaller than ½. The other parameters (w and σ) showed no significant difference between the two groups. As predicted by our model, optimistic participants were overall less confident in their answer to the question "which machine pays more?" than the pessimistic participants and the difference grows with the generosity of the task (Fig. 5 and Fig. S2a in Supplementary Information). The average reported confidence for optimists and pessimists in all blocks was 0.48±0.03 vs. 0.55±0.03 respectively, and 0.60±0.06 vs. 0.81±0.03 in high generosity (>0.65) blocks (expressed as mean ± s.e.m.). The average confidence reported by participants entered a two-way ANOVA with group (optimists and pessimists) and generosity condition (high and low) as between- and within-subjects factors, respectively. The ANOVA showed a main effect of the optimistic/pessimistic group ($F(1,15)= 5.445$, $p=<0.04$) and a significant interaction between group and generosity ($F(1,15)= 8.385$, $p<0.02$). Furthermore, although the average unbiased difficulty was approximately the same for all participants, the average reported confidence per participant was not. Indeed, it decreased with the *b* value adjusted for that participant in the same way as the increase in the average difficulty perceived by an optimal agent with that *b* value (see Fig. S2b in Supplementary Information).

We now turn to our third prediction, that confidence reports in pessimists and optimists should depend differently on the difficulty experienced in the task. The results, shown in Fig. 5, are striking: optimists and pessimists show very different confidence patterns, as predicted by the model.
The isoconfidence lines are mostly vertical for optimistic agents, signaling a direct dependence with task difficulty. For pessimists, however, the isoconfidence lines rotate from vertical to horizontal as machine generosity increases, indicating a weaker dependence with the unbiased difficulty in high generosity situations. To further emphasize the importance of taking into account participants' prior beliefs, we compare the predictions of our model with those made by a model with a non-informative prior for every participant: without accounting for participants' biases, the model would show approximately vertical isoconfidence lines for both groups in Fig. 5. Additionally, without separating optimists from pessimists, human confidence report would be overall judged as overconfident in high generosity situations and underconfident in low generosity situations when compared with the model with a non-informative prior (see Fig. S4 in Supplementary Information). However, this confidence bias is precisely what is expected by the normative model after including the *b* and *w* values fitted from the decision part of the task.

The Effect of Reporting Confidence

One corollary of the last section is that we have a good model to predict average confidence levels in different blocks accurately, as can be tested for the following purpose by a linear fit between human average block confidence and model prediction, which returns R^2=0.94 (p<0.001). This thus allows us to answer what the confidence report would be when it is not prompted for, which can in turn be used to compare behavior *after* the report and in the absence of such report, across similar levels of (un)reported confidence.

We estimated the probability to shift away from the previously chosen machine in high and low confidence blocks, splitting them at the median model confidence value (equal to 0.79). After subjects were asked for confidence, this shift probability increased by comparison with the no-report condition, and this shift was stronger in the low confidence blocks than in the high confidence block (see Fig. 6). The fact that the difference between producing a confidence report or not appears only in low confidence blocks tells us that this really amounts to an effect of asking for confidence, ruling out alternative explanations such as memory effects. When confidence is low, asking for it alters subsequent behavior strongly. We test this effect and the interaction through a two-way ANOVA, which showed a significant effect of the report type (which report vs. confidence report) on the shift probability after the report: F(1,15)=5.60 (p<0.05), and also significant for the interaction between confidence level and report type: F(1,15)=4.04 (p<0.05), indicating that the shift probability is different for different report types, but only when confidence is low.

Of course, our decision model as-is is incapable of explaining this effect, since it has no information regarding whether the confidence was reported or not. However, we can model the effect of the confidence report by adding a small *ad hoc* feature to the framework developed. When asked for confidence, the model increases its decision variance σ (see *Methods*) by a factor manually fitted to 0.05 times one minus the reported confidence. With this extra parameter, we are able to reproduce the post-report behavior (see Fig. 6).

Discussion

We presented a two-armed bandit experiment in which participants maximize their rewards by playing machines with unknown reward rates. Participants explored those machines and they were occasionally asked to report their confidence about knowing which machine pays more. We first

showed that each participant can be identified with a consistent and definite optimistic/pessimistic prior bias according to how much they explore, exploit and get rewarded in the game, independently of their confidence report. Statistical modeling of confidence predicts that subjects classified as pessimists should report higher confidence than subjects classified as optimists, specially in high generosity situations. It also predicts that confidence in those two types of players should be impacted differently by task difficulty. Participants' behavior conformed to these two predictions. Our results indicate the variability of confidence reports across participants are rational when taking into account their different prior expectations.

Taking into account participants' prior expectations can explain some apparently irrational behavior in various contexts [42]. For instance, we can revisit the result of Tversky and Griffin [33] in much the same way. Their experiment consisted of informing participants that a coin is biased, yielding one outcome 60% of time. Subjects were not informed if the bias was toward tails or heads. Outcomes of this mysterious coin were shown, and participants were asked to decide if the bias was toward tails or heads, and report their confidence in that decision. Subjects were typically overconfident when they observed a few tosses with a strong imbalance and underconfident when they observed many tosses with a moderate imbalance, by comparison with a mathematical model informed that the bias is exactly 60% (a delta function), as the authors reported. If instead we model this prior with some uncertainty (with a Beta distribution centered on 60%), then both humans and the normative model display the same pattern of overconfidence and underconfidence, and the magnitude of the confidence bias increases with the uncertainty of the prior (see Supplementary Information and Fig. S3). Although we did not reproduce Griffin and Tversky's experiment, and therefore cannot claim that participants' prior beliefs in that task are indeed better explained by a more permissive prior, a similar logic operates in our gambling task. If we ignore participants' prior bias from gameplay, and use the 'reasonable' non-informative prior instead, they would be overall misjudged as overconfident (underconfident) in high (low) generosity situations (see Fig. S4). We showed that participants' prior beliefs can be major determinants in the analysis of rationality in the human sense of confidence, and seemingly reasonable (but erroneous) assumptions, like a non-informative prior would be in our experiment, can lead to serious mistakes when judging rationality.

Taking into account participants' prior beliefs appears key to evaluating the rationality of their behavior in our task. Such an approach should nevertheless avoid two pitfalls. The first is circularity

(*assuming* a difference by appealing to priors rather than *explaining* it), which can be a limitation of Bayesian models. Indeed, in principle, any behavior can be accounted for by a particular set of prior beliefs [48, 54]. The second is overfitting (improving the goodness-of-fit by resorting to more free parameters). Our approach avoids these pitfalls by using a parsimonious, general Bayesian model that accounts for both choice and confidence, resulting in a parameter-free confidence model. Indeed, we assumed that these two aspects of behavior (choice and confidence) are exclusively affected by the same set of prior beliefs. Similarly to cross-validation methods, where data are divided into "training and test" sets [55] in order to limit the complexity of the model (i.e. the number of free parameters), our experimental design is divided into "training and test" tasks (choices and confidence report, respectively). Behavior in the training task is used to learn -and pin up- the prior parameters for each individual. If the resulting parameter-free model for the test task predicts human data accurately, it is likely to generalize well to future data, limiting the risk of overfitting [56]. A different way to learn the prior beliefs of participants is by iterated learning [57, 58], in which responses given in one trial affect the data shown in the next. It can be shown that, if certain conditions are met, responses are eventually sampled from their prior distribution. In principle, this method could be implemented in our environment by a repetition of various experiments in which a participant judges the payoff of a machine based on the observed outcomes in that trial, and the real payoff of the machine in the next trial would be equal to the estimated one by the participant in the previous trial (starting with a random payoff in the first trial). As in our approach, prior beliefs are then used to make parameter-free predictions in the test task and evaluate the generalization potential of the model.

We chose to display the confidence in a two dimensional plane instead of along a one-dimensional quantity (e.g. confidence vs. difficulty) following Aitchison *et. al.* [29], who argue that there is always the freedom of reparameterizing confidence reports by a monotonous function, which in one dimension would allow us to trivially explain any observed human confidence pattern by a suitable such transformation. When plotting the results along two independent variables this reparameterization is no longer possible, and the arising pattern of isoconfidence lines is now a robust indicator of the participants' behavior. Specifically, there are (at least) two kinds of optimality. The two dimensional representation only analyzes the transformation of incoming data into an internal representation from which confidence is read out as a continuous variable (first type of optimality), separating it from the mapping of this continuous variable onto some external scale in order to report it (second type of optimality). Since the first type of optimality is

independent of rescaling the confidence report by any monotonous function, the two-dimensional analysis of human rationality is independent of a direct matching between the numerical probability of being correct and the human confident report (i.e. the calibration of confidence), in which humans do not seem to be optimal [8, 35], and also independent of the different ways in which participants may use the confidence bar, as long as it is consistent for each participant. In particular, our two dimensional analysis studies how the knowledge of a rational agent should update in different situations (i.e. the 45 different blocks shown in Fig. 5), and what the *relative* values of confidence between these situations should be if it was a normative readout from the optimally updated knowledge. The predicted differences between optimists and pessimists strongly uphold human confidence as a readout from a probabilistic representation of knowledge that is optimally (or at least approximately optimally) updated from the prior [16].

Although confidence reports have traveled a winding road in the psychology and neuroscience literature, recent work is settling in on a statistically normative account of confidence [9, 19]. This study contributes to this view, showing how traits that have been traditionally seen as irrational can be in fact understood as differences in prior expectations. This further fuels the view of humans as rational animals, and signs off another success of the Bayesian rationality program [59].

The importance of principled, quantitative and robust behavioral models is not only theoretical, but also practical. Here, the availability of a successful model for confidence allowed us to study a further, 'higher order' phenomenon: how reporting confidence affects later behavior. This question, despite its simplicity, seems to have been overlooked in the literature so far. To our knowledge, we present the first contribution in this direction. We report that probing subjects for a confidence report increases explorative behavior in subsequent trials, as if subjects relied less on their prior experience. An accurate model of confidence is useful here to disprove alternative low-level explanations such as that the time spent answering the confidence question washes out the carried knowledge representation. Indeed, by telling low and high confidence regimes apart through the use of the model, we are able to show that the change in behavior is specific to low confidence situations, hence ruling out an explanation in term of forgetting.

However, our model for gameplay choices does not *explain* this increased explorative behavior. In the context studied, participants should have a fairly robust idea of the machine's payoffs by the time they get to the confidence question, such that increasing exploration at that stage proves

indeed suboptimal. An interesting possibility is that participants could interpret a confidence prompt as a hint that they are in the wrong track, and hence increase their exploration. Although this is partially responded by the separation between high and low confidence trials, it is nevertheless a matter for further inquiry. Similar interaction between behavior and the experimenter's question was reported in other experiments, including studies on causal inference in development. In such studies, children are asked repeatedly the same question. Their answers are typically modeled as samples from a distribution [60]. However, when the same person asks the same question twice, children tend to think that they have provided an incorrect answer, and thus change their answer the second time [61].

Our results also illustrate that principled quantitative models prove particularly informative when they predict non-trivial, maybe even counterintuitive behaviors [47]: a pessimistic agent should yields higher confidences than an optimistic one. This is particularly relevant in relation to the "irrational" optimism bias, by which we tend to expect more from the world than what it actually gives us [43]. This bias towards high expectations would thus mean that we should typically display *under*confidence with respect to an unbiased agent, according to the aforementioned relation. However, we found that a given individual can display both *under* and *over*confidence, with a tendency to the latter in most domains [62]. Here, we propose an explanation in a given task, showing that apparently irrational confidence judgements can be simply understood as varying prior biases. How general the form of this connection is, and, how optimism and overconfidence may coexist are interesting new avenues for research.

Methods

Experiment Details

A total of 18 adult participants played a two armed bandit game for which they were asked to maximize total reward. One participant was excluded from the analysis for obtaining a total reward consistent with random play. Each participant completed 135 blocks of 16 trials, giving a total of 36720 individual decisions and 765 confidence reports. Participants were informed they were going to play a series of unrelated blocks in each of which the payoff of the machines was unknown

but fixed, and their aim was to maximize the total reward in order to win a monetary prize. Blocks were clearly delineated from one another by pauses, and there was a message reminding participants that separate blocks were independent from one another.

In the 'which' type of block, participants were asked to choose which machine they thought has a higher nominal payoff, based on their limited experience at the moment of the report. In the 'confidence' type of block, they were also required to make a continuous judgement of confidence in their decision. As mentioned before, the statistical account for this measure is the probability that the decision made is correct. Finally, 'no' blocks included no question.

The nominal reward rates for the machines were chosen homogeneously between 0 and 1 and repeated for each type of block (45 blocks of each type). This choice was made in order to get the widest possible spectrum of payoffs, this being the reason we do not see the strong inverted U-shape in the plots of rewards vs. exploration characteristic of bandit experiments. The order of blocks was randomized for each participant, and each participant completed 6 demonstration trials before begining the task.

The task was designed and implemented in Python using the PyGame library [63], and lasted around one hour during which the participant was left alone in a quiet room. Average performance did not show a significant decay during the task.

Model Details

Bayesian knowledge update. Observers begin each block with a prior distribution *Beta(ps,pf)* for the reward probability of each of the two machines, where *ps* and *pf* encode fictitious prior successes and failures, respectively. A natural reparameterization of this distribution is by using its mean *b=ps/(ps+pf)*, which is a prior measure of the expected payoff, and *w=ps+pf*, which encodes the weight of prior evidence. Due to the conjugacy between the beta prior and the binomial likelihood assumed for the rewards, the posterior distribution after experiencing *s* successes and *f* failures results in a *Beta(ps+s,pf+f)*. Intuitively, low values of the prior mean *b* correspond to a pessimistic perspective, while high values of this parameter represent a more optimistic take.

Computation of statistical confidence. The normative statistical confidence that the agent should report after deciding (using any decision process) that machine B has a higher payoff that machine A is:

$$conf_B \propto \int_0^1 pdf_A(x)\left(1 - cdf_B(x)\right) dx$$

where $pdf_A(x)$ is the value of the Beta distribution of machine A (at the moment of the report) evaluated at x and $\left(1 - cdf_B(x)\right)$ is the proportion of the Beta distribution of machine B that lies in values higher than x . The confidence $conf_A$ that machine A is better than B is analogous.

Intuitively, the process of computing confidence in the decision that machine B pays more than A can be seen as taking an infinite number of samples from A, and for each sample calculating the proportion of the distribution of B that lies over it.

Decision-making model and motivation. The confidence report depends only on the learning component, whose optimal solution is presented above. By contrast, the optimal solution for deciding which machine to play given the experience so far in order to maximize future rewards is more difficult. However, this is a well studied problem that has been solved in finite-horizon bandit problems by dynamic programming -looking at all possible outcomes from the last trial to the current one-, an approach that requires an amount of calculations that grows exponentially with the number of remaining trials [52, 53].

Several heuristics have been developed in order to approximate the optimal solution, or mimic human judgements [53, 64]. We modeled the decision of which machine to play as follows. First, we define the variable d (perceived difficulty) as one minus the absolute value of the difference between the means of both machines' posterior distribution. A sample decision value is then taken from a normal distribution with mean (1-d) and standard deviation σ which we set equal to 0.05 throughout. If the sample is negative, then the machine with the lower estimated payoff is chosen (a decision to explore). If the percept is positive, the arm with the higher estimated payoff is chosen (a decision to exploit).

Several arguments support the choice of this heuristic as a model for the decision-making component. First, we compare different alternative decision models according to their mean per-trial likelihood in different conditions (see [64] for the strategy used to compare stochastic with deterministic models) and found that our model presented a high overall agreement with human data when compared to other alternatives (see Fig. S5 in Supplementary Information). Second, it captures the proportion of explorative decisions (choosing the machine with the lowest payoff so far) seen in humans as a function of the perceived and unbiased difficulty of the task, which is not the case for most other heuristics (like ε-greedy, which predicts a constant function) or the optimal model, which, for two-armed bandits, almost always choose the machine with the highest success ratio so far. Third, by only accounting for d, this heuristic ignores the uncertainty about the estimated reward rates (w) and assume a fixed randomness of choice across individuals (σ), so that

all the variation in the behavior of different participants can be accounted for only by a different prior mean b (Figs. 3, 4 and S1). This assumption agrees with the statistical analysis of the data, which shows that the model that varies only b was a better model when compared to any other model with one, two or three degrees of freedom (xp>0.69 -exceedance probability-, see below). Additionally, it is not possible to cover the entire spectrum of behavioral summaries seen in Figs. 3, 4 and S1 just by varying w or σ, and neither w nor σ are as consistent as b for each participant across different environmental conditions.

Analsis Details

Consistency of the prior bias. Participants' gameplay is shown with black triangles in Figures 3, 4 and S1, which use the following summary measures for the axes: average reward per block, average exploration per block (proportion of trials in which there is a change in machine choice), and average persistence per block (proportion of trials where a machine is chosen immediately after a no-reward trial in that machine). Each person is labeled with a unique color in all regimes, namely the value of b that minimizes the squared distance to the model predictions summed across the 10 summaries of Figures 3, 4 and S1. It is visually evident from these figures that the color label is consistent, in the sense that if a participant is best represented by a pessimistic (optimistic) value of the prior mean b in one condition, then this participant will likely be represented by a pessimistic (optimistic) value of b in all other conditions.

We can check the statistical validity of this assertion by performing a permutation test in the following manner. For each participant, we compute the variance of the values of b that best represent his or her behavior in each of the 10 different conditions *separately*. For example, if a participant is best represented by an optimistic behavior in some conditions, but by a pessimistic behavior in others, their value of the variance will be high. A within-subject variance that is smaller than the between-subject variance indicates that the optimistic/pessimistic difference pertains to the group level rather than for individual subjects. Therefore, the test consisted on performing 100,000 random permutations between the participants' labels across conditions, and measuring the variance for each surrogate across the 10 conditions. For every permutation, the calculated variance for all surrogates was higher than the variance of all participants without permuting, yielding $p<(1/100{,}000)$.

Two-dimensional presentation of the results. Since the iso confidence lines in Fig. 5 do not change if we reparameterize confidence by a monotonous function, and therefore neither does the prediction #3 of our model, we chose to display the data as was reported by the participant in the continuous confidence bar between 0 and 1, without any calibration function. If we, for example, assign quantiles to the answers of each subject, then prediction #2 of the model (namely, pessimists report higher confidence than optimists) is partially opaqued. However, since the strength of this asymmetry between optimists and pessimists increases when one option is played more frequently than the other, we still see, after reparametrization, that optimists and pessimists' confidence varies in a different way with the generosity of the task (which, in practice, is proportional to the difference between the exploitation of different options), both for humans and for the normative model. This is shown in Fig. S2a (see Supplementary Information).

To give the relevant information about the machines' reward history, four numbers are required: the successes and failures in each machine until that point. Therefore, it is possible that two points in the same location of the two-dimensional space constructed by displaying generosity vs. difficulty in Fig. 5 actually correspond to two different points in the four-dimensional space, i.e. to different machine histories. This effect is particularly important when comparing high generosity points between optimists and pessimists. For the latter, the points in this area correspond to a higher exploitation (choosing one machine more often than the other) than the points in this area for optimists. As explained before, this is part of the reason we find horizontal isoconfidence lines for pessimists but not for optimists.

We also note that the patterns we observe persist even when using the generative generosity and difficulty instead of the unbiased generosity and difficulty; they are simply more noisy. Finally, note that not all regions in this space are allowed, for instance, the block cannot be easy if both machines pay very little.

Model comparison. The exceedance probability (xp) of a model quantifies the probability that this model is more frequent than the others (within the tested set) in the general population of subjects. We computed exceedance probabilities from the model evidence using the software developed by [65]. The evidence for each model and each subject was calculated by integrating each model's mean likelihood over its parameters, under the i.i.d data assumption. The integral was approximated by a sum over a discrete grid. Grid points (10 points for b and w, 8 for σ) were spaced linearly for parameters b and w and exponentially for σ (which corresponds to a non-informative prior in log-space, a natural choice for variance parameters). The final result depends on the

integration limits chosen for each parameter. The parameter b is bounded between 0 and 1, but w and σ are unbounded. Limits were chosen so that the behavior of the model did not change significantly for parameter values beyond them (w between 2 and 120; σ between e^{-3} and e^{4}). However, the results are robust to the choice of these limits: similar exceedance probability values are obtained when the limit was moved plus or minus two points in the chosen scale for each parameter.

Data availability. The behavioural data are available here: https://figshare.com/articles/Behavioral_data/4788823

Acknowledgements

We would like to thank Falk Lieder, Tania Lombrozo and Tom Griffiths for insightful conversations, Melisa Bentivegna for help with early data collection, and the organizers of the Subjective Confidence Workshop at Les Treilles where some preliminary results were first presented. This work was partly funded by grants CONICET PIP 11220130100384CO and UBACYT 20020130200202BA. FM is funded by the French center for atomic energy (CEA). M.S. is sponsored by CONICET and the James McDonnell Foundation 21st Century Science Initiative in Understanding Human Cognition – Scholar Award.

Author contributions

P.T. performed the experiment, analysed the data, interpreted the results and wrote the manuscript.
F.M. provided analytical tools, interpreted the results and edited the manuscript.
M.S. interpreted the results and edited the manuscript.
A.S. designed the study, performed the experiment, analysed the data, interpreted the results, and wrote the manuscript.

Additional information

Correspondance and requests for materials should be adressed to P.T.

References

1. Plous, Scott. The psychology of judgment and decision making. Mcgraw-Hill Book Company, 1993.

2. Kepecs, Adam, et al. "Neural correlates, computation and behavioural impact of decision confidence." Nature 455.7210 (2008): 227-231.

3. Kiani, Roozbeh, Leah Corthell, and Michael N. Shadlen. "Choice certainty is informed by both evidence and decision time." Neuron 84.6 (2014): 1329-1342.

4. Persaud, Navindra, Peter McLeod, and Alan Cowey. "Post-decision wagering objectively measures awareness." Nature neuroscience 10.2 (2007): 257-261.

5. Lak, A. et al. Orbitofrontal cortex is required for optimal waiting based on decision confidence. Neuron 84, 190–201 (2014).

6. Kepecs, Adam, and Zachary F. Mainen. "A computational framework for the study of confidence in humans and animals." Philosophical Transactions of the Royal Society of London B: Biological Sciences 367.1594 (2012): 1322-1337.

7. Berger, James O. Statistical decision theory and Bayesian analysis. Springer Science & Business Media, 2013.

8. Baranski, Joseph V., and William M. Petrusic. "The calibration and resolution of confidence in perceptual judgments." Perception & Psychophysics 55.4 (1994): 412-428.

9. Pouget, Alexandre, Jan Drugowitsch, and Adam Kepecs. "Confidence and certainty: distinct probabilistic quantities for different goals." Nature neuroscience 19.3 (2016): 366-374.

10. Peterson, Cameron R., and Lee R. Beach. "Man as an intuitive statistician." Psychological bulletin 68.1 (1967): 29.

11. Yu, Shuli, Timothy J. Pleskac, and Matthew D. Zeigenfuse. "Dynamics of postdecisional processing of confidence." Journal of Experimental Psychology: General 144.2 (2015): 489.

12. Navajas, Joaquin, Bahador Bahrami, and Peter E. Latham. "Post-decisional accounts of biases in confidence." Current Opinion in Behavioral Sciences 11 (2016): 55-60.

13. Paz, Luciano, et al. "Confidence through consensus: a neural mechanism for uncertainty monitoring." Scientific reports 6 (2016).

14. Wei, Ziqiang, and Xiao-Jing Wang. "Confidence estimation as a stochastic process in a neurodynamical system of decision making." Journal of neurophysiology 114.1 (2015): 99-113.

15. Insabato, Andrea, et al. "Confidence-related decision making." Journal of neurophysiology 104.1 (2010): 539-547.

16. Meyniel, Florent, Mariano Sigman, and Zachary F. Mainen. "Confidence as bayesian probability: From neural origins to behavior." Neuron 88.1 (2015): 78-92.

17. Pleskac, Timothy J., and Jerome R. Busemeyer. "Two-stage dynamic signal detection: a theory of choice, decision time, and confidence." Psychological review 117.3 (2010): 864.

18. Drugowitsch, Jan, Rubén Moreno-Bote, and Alexandre Pouget. "Relation between belief and performance in perceptual decision making." PloS one9.5 (2014): e96511.

19. Sanders, Joshua I., Balázs Hangya, and Adam Kepecs. "Signatures of a statistical computation in the human sense of confidence." Neuron 90.3 (2016): 499-506.

20. Jones, Matt, and Bradley C. Love. "Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition." Behavioral and Brain Sciences 34.04 (2011): 169-188.

21. Anderson, John Robert. The adaptive character of thought. Psychology Press, 1990.

22. Jaynes, Edwin T. Probability theory: The logic of science. Cambridge university press, 2003.

23. Sanborn, Adam N., Vikash K. Mansinghka, and Thomas L. Griffiths. "Reconciling intuitive physics and Newtonian mechanics for colliding objects." Psychological review 120.2 (2013): 411.

24. Goodman, Noah D., et al. "Intuitive theories of mind: A rational approach to false belief." Proceedings of the twenty-eighth annual conference of the cognitive science society. 2006.

25. Ullman, Tomer D., Noah D. Goodman, and Joshua B. Tenenbaum. "Theory learning as stochastic search in the language of thought." Cognitive Development 27.4 (2012): 455-480.

26. Tenenbaum, Joshua B., and Thomas L. Griffiths. "Theory-based causal inference." Advances in neural information processing systems (2003): 43-50.

27. Ernst, Marc O., and Martin S. Banks. "Humans integrate visual and haptic information in a statistically optimal fashion." Nature 415.6870 (2002): 429-433.

28. Deroy, Ophelia, Charles Spence, and Uta Noppeney. "Metacognition in Multisensory Perception." Trends in Cognitive Sciences 20.10 (2016): 736-747.

29. Aitchison, Laurence, et al. "Doubly bayesian analysis of confidence in perceptual decision-making." PLoS Comput Biol 11.10 (2015): e1004519.

30. Bar-Tal, Yoram, Anat Sarid, and Liat Kishon-Rabin. "A test of the overconfidence phenomenon using audio signals." The Journal of general psychology 128.1 (2001): 76-80.

31. Björkman, Mats, Peter Juslin, and Anders Winman. "Realism of confidence in sensory discrimination:

The underconfidence phenomenon." Attention, Perception, & Psychophysics 54.1 (1993): 75-81.

32. Camerer, Colin, and Dan Lovallo. "Overconfidence and excess entry: An experimental approach." The American Economic Review 89.1 (1999): 306-318.

33. Griffin, Dale, and Amos Tversky. "The weighing of evidence and the determinants of confidence." Cognitive psychology 24.3 (1992): 411-435.

34. Kvidera, Sara, and Wilma Koutstaal. "Confidence and decision type under matched stimulus conditions: Overconfidence in perceptual but not conceptual decisions." Journal of Behavioral Decision Making 21.3 (2008): 253-281.

35. Moore, Don A., and Paul J. Healy. "The trouble with overconfidence." Psychological review 115.2 (2008): 502.

36. Olsson, Henrik, and Anders Winman. "Underconfidence in sensory discrimination: The interaction between experimental setting and response strategies." Attention, Perception, & Psychophysics 58.3 (1996): 374-382.

37. Gigerenzer, Gerd, Ulrich Hoffrage, and Heinz Kleinbölting. "Probabilistic mental models: a Brunswikian theory of confidence." Psychological review 98.4 (1991): 506.

38. Mamassian, Pascal. "Overconfidence in an objective anticipatory motor task." Psychological Science 19.6 (2008): 601-606.

39. Graziano, Martin, and Mariano Sigman. "The spatial and temporal construction of confidence in the visual scene." PLoS One 4.3 (2009): e4909.

40. Ma, Wei Ji, and Mehrdad Jazayeri. "Neural coding of uncertainty and probability." Annual review of neuroscience 37 (2014): 205-220.

41. Griffiths, Thomas L., et al. "How the Bayesians got their beliefs (and what those beliefs actually are): comment on Bowers and Davis (2012)." (2012): 415.

42. Van den Steen, Eric. "Overconfidence by Bayesian-rational agents." Management Science 57.5 (2011): 884-896.

43. Sharot, Tali. "The optimism bias." Current Biology 21.23 (2011): R941-R945.

44. Sharot, Tali, Christoph W. Korn, and Raymond J. Dolan. "How unrealistic optimism is maintained in the face of reality." Nature neuroscience 14.11 (2011): 1475-1479.

45. Shepperd, James A., et al. "A primer on unrealistic optimism." Current directions in psychological science 24.3 (2015): 232-237.

46. Sharot, Tali, et al. "Neural mechanisms mediating optimism bias." Nature 450.7166 (2007): 102-105.

47. Lefebvre, Germain, et al. "Behavioural and neural characterization of optimistic reinforcement learning." Nature Human Behaviour 1 (2017): 0067.

48. Bowers, Jeffrey S., and Colin J. Davis. "Bayesian just-so stories in psychology and neuroscience." Psychological bulletin 138.3 (2012): 389.

49. Macready, William G., and David H. Wolpert. "Bandit problems and the exploration/exploitation tradeoff." IEEE Transactions on evolutionary computation 2.1 (1998): 2-22.

50. Cohen, Jonathan D., Samuel M. McClure, and J. Yu Angela. "Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration." Philosophical Transactions of the Royal Society of London B: Biological Sciences 362.1481 (2007): 933-942.

51. Daw, Nathaniel D., et al. "Cortical substrates for exploratory decisions in humans." Nature 441.7095 (2006): 876-879.

52. Gittins, John, Kevin Glazebrook, and Richard Weber. Multi-armed bandit allocation indices. John Wiley & Sons, 2011.

53. Steyvers, Mark, Michael D. Lee, and Eric-Jan Wagenmakers. "A Bayesian analysis of human decision-making on bandit problems." Journal of Mathematical Psychology 53.3 (2009): 168-179.

54. Diaconis, Persi, and David Freedman. "On the consistency of Bayes estimates." The Annals of Statistics (1986): 1-26.

55. Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. The elements of statistical learning. Vol. 1. Springer, Berlin: Springer series in statistics, 2001.

56. Myung, In Jae. "The importance of complexity in model selection." Journal of mathematical psychology 44.1 (2000): 190-204.

57. Lewandowsky, Stephan, Thomas L. Griffiths, and Michael L. Kalish. "The wisdom of individuals: Exploring people's knowledge about everyday events using iterated learning." Cognitive Science 33.6 (2009): 969-998.

58. Griffiths, Thomas L., and Michael L. Kalish. "Language evolution by iterated learning with Bayesian agents." Cognitive science 31.3 (2007): 441-480.

59. Oaksford and Chater. “Bayesian Rationality. The Probabilistic Approach to Human Reasoning.” Oxford University Press (2007).

60. Denison, Stephanie, et al. "Rational variability in children’s causal inferences: The sampling hypothesis." Cognition 126.2 (2013): 285-300.

61. Gonzalez, Aaron, et al. "Is that your final answer? The effects of neutral queries on children's choices." CogSci. 2012.

62. Salles, Alejo, et al. "The metacognitive abilities of children and adults." Cognitive Development 40 (2016): 101-110.

63. Shinners, Pete (2011). PyGame - Python Game Development. Retrieved from

http://www.pygame.org.

64. Zhang, Shunan, and J. Yu Angela. "Forgetful Bayes and myopic planning: Human learning and decision-making in a bandit setting." Advances in neural information processing systems. 2013.

65. Stephan, Klaas Enno, et al. "Bayesian model selection for group studies." Neuroimage 46.4 (2009): 1004-1017.

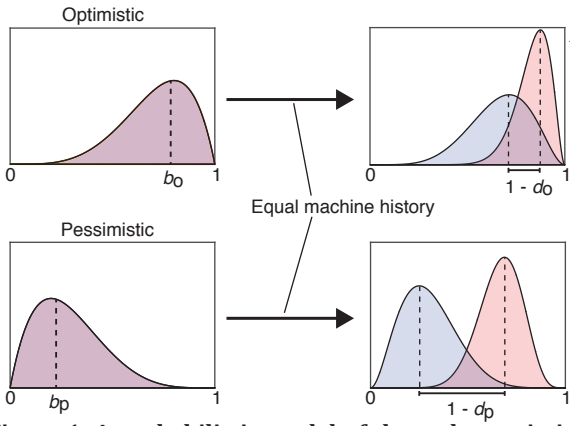

Figure 1. A probabilistic model of the task, pessimists should report more confidence than optimists. Initially (left), the prior belief distribution over the payoff of both machines lies in high values for optimists (high b_O) and low values for pessimists (low b_P). In situations in which one option is chosen more frequently than the other, optimists and pessimists are expected to differ largely in the confidence they report after receiving the exact same history of successes and failures in both machines (in this case, one failure in the blue machine and 6 successes in the red one). In practice, one necessary condition for choosing one option more frequently is to receive more reward from one of the options (represented by the distribution in red). In this situation, the distribution from the option left behind (blue) will not be far from the prior, yielding two distributions that overlap more in optimistic than pessimistic, and hence a lower perceived difficulty for pessimists than optimists ($d_P < d_O$), and a higher confidence report for pessimists than optimists.

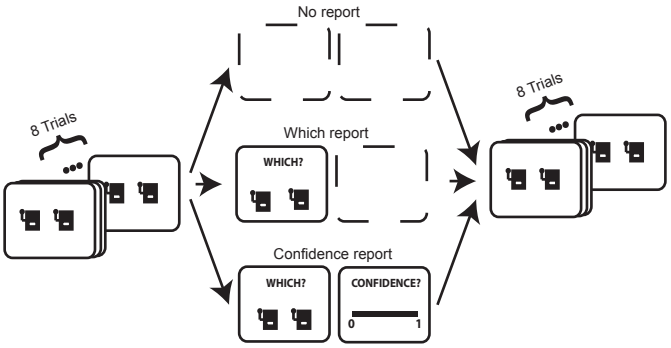

Figure 2. The bandit gambling task. Each block consisted of 16 trials. Depending on the type of block being played, the block was played without interruption (No report); the participant was required to choose which machine had the higher nominal payoff (Which report); or same as the Which report plus a continuous report between 0 and 1 for the confidence in that decision (Confidence report). Each participant played 45 of each type of blocks, each with different, fixed machine payoffs.

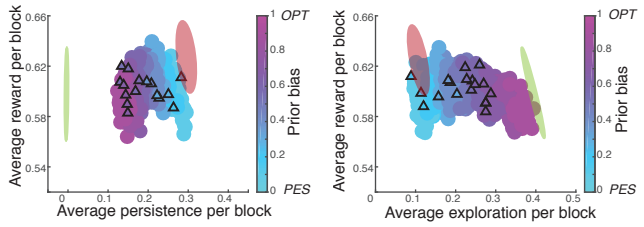

Figure 3. A range of behaviors accounted for by the optimism level. Average persistence per block corresponds to the proportion of trials in which a machine is chosen immediately after a no-reward trial in that machine. Average exploration per block is the proportion of trials in which there is a change in machine choice. Average reward per block is to the proportion of trials in which a reward was obtained. Participants are shown with blacked-edge triangles. The color within each triangle corresponds to the value of b fitted from these two panels together with the eight panels in Figs. 4 and S1. Each of these two panels shows behavior averaged over all 135 blocks. The prior bias corresponds to values of $b = ps/(ps+pf)$ between 0 (pessimistic) and 1 (optimistic). The coloured clouds correspond to 140 runs of the model. In each run, prior parameters ps and pf were sampled from a discrete uniform distribution from 1 to 7, with the constraint that $w=ps+pf=8$ in every run. Since the values of w and σ were fitted globally, the entire variation of human behavior can therefore be accounted for solely by the variation of the prior mean b. In comparison, the entire range of behaviors for the optimal (red) and Win-Stay-Lose-Shift (green) models are much more restricted and incompatible with the data.

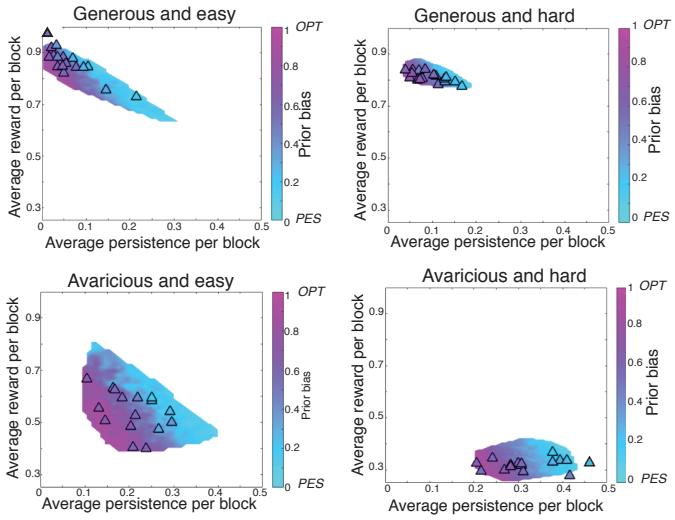

Figure 4. The optimism level is an idiosyncratic trait, stable across conditions. Average rewards vs. average persistence in different regimes. Legend as in previous figure. Each panel shows the averaged results from 10 to 15 different blocks, all belonging to a different regime (Generous and Easy, Generous and Hard, Avaricious and Easy, and Avaricious and Hard) according to the generative difficulty and generosity of the block. The coloured clouds correspond to 700 runs of the model, generated as in previous figure. The prior biases (colors) are assigned consistently to participants across regimes. This can be seen by noting that pessimist (optimist) participants tend to be represented by pessimist (optimist) values of the model in every regime. We validated this observation with a permutation test and a cross validation test (see the results in main text, details in *Methods*).

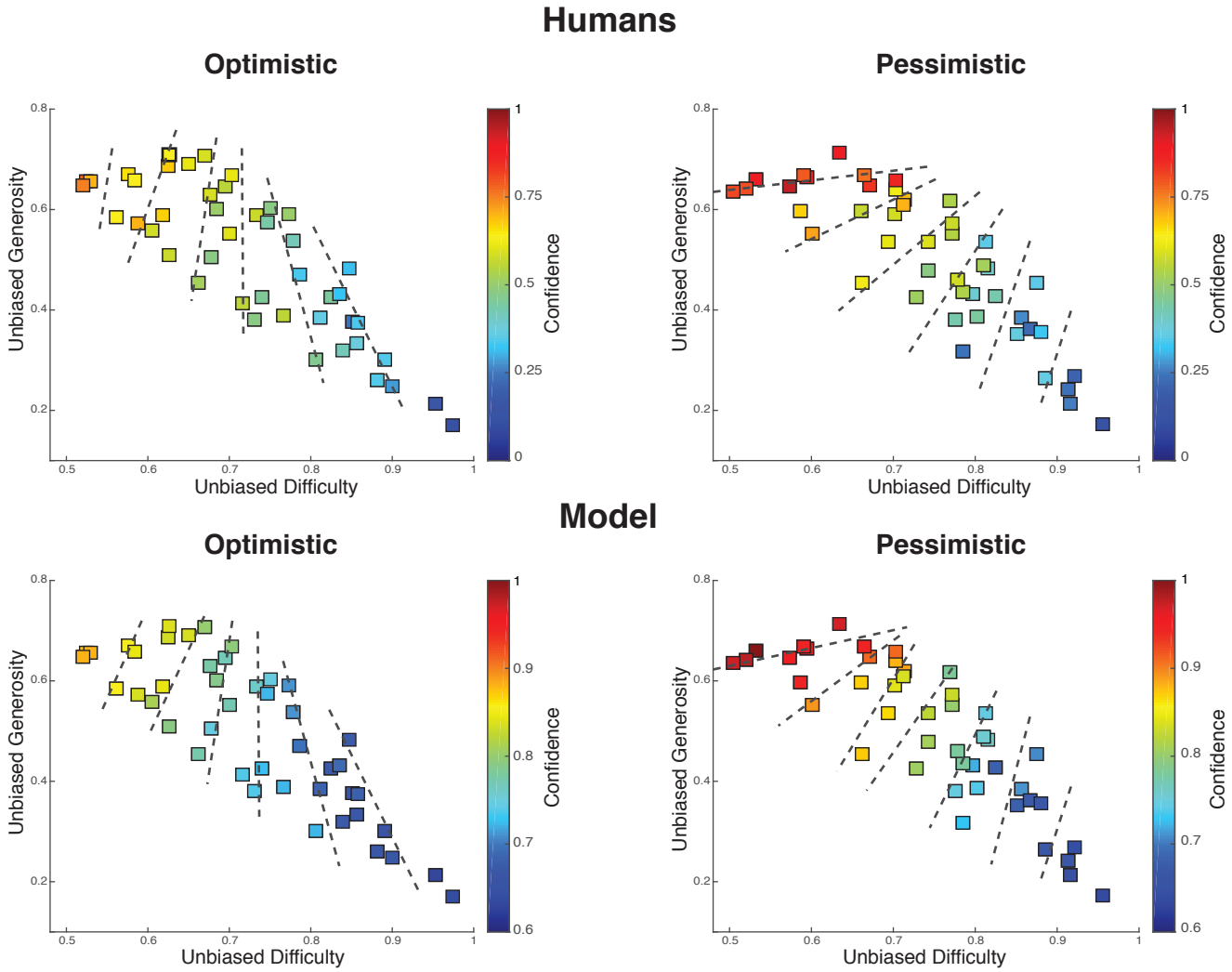

Figure 5. Over- and under-confidence are explained by the prior optimism level measured in gameplay. Confidence levels reported by humans and the normative model as a function of the unbiased difficulty and unbiased generosity on the block. Unbiased difficulty (generosity) is computed as the distance (average) between the means of machines' reward rates at the moment of the report, as estimated objectively given the exact observations received by subjects. Squares correspond to the 45 'confidence' type blocks, each with a different nominal payoff. The color of the squares represents the average reported confidence for all subjects in that block (10 optimists, 7 pessimists), and the position corresponds to the average uniform generosity and uniform difficulty in those blocks. The dotted isoconfidence lines were computed by first interpolating, then separating regions with polynomial isoconfidence curves and then performing linear fits over these curves. Humans are classified as pessimistic or optimistic based on their prior bias b, obtained separately from their decisions in gameplay. The differences between optimists and pessimists are accurately captured by the normative cofidence model. First, pessimists report higher confidence than optimists on average, and this is particularly salient in the region of high generosity (see Fig. 1 for an explanation of this effect). Second, isoconfidence lines are more horizontal for pessimists than for optimists in situations of high generosity. Note that a model with a non-informative prior for every participant would show approximately vertical isoconfidence lines for both groups.

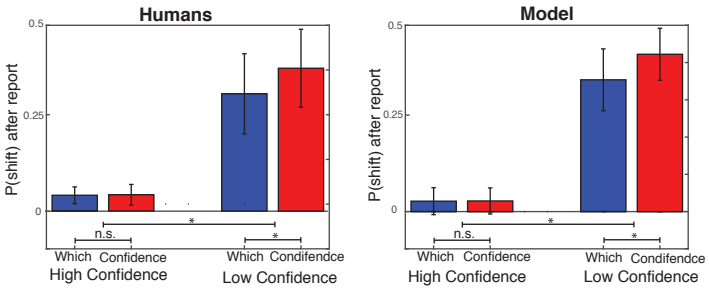

Figure 6. Probing confidence induces a deconfirmation bias in subsequent choices. In low confidence situations, the probability of shifting to the other option after the report is bigger in blocks in which confidence is reported ('confidence') than in blocks in which only the best machine is reported ('which'). The model results corresponds to the average results in 17 runs of the full task (corresponding to the 17 participants). The model was provided with a 'report mechanism' that distinguishes both types of block. Significant interactions (p<0.05) are indicated with "*" between the groups, non significant with "n.s.". Error bars indicate s.d. across participants (n=17).